Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 34 (2) APR. 2026

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 in improve quality on issues pertaining to science and engineering and its related fields.

Pertanika Journal of Science & Technology currently publishes 6 issues a year (*February, April*, *June, August, October, and December*). 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 focusses 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 programme such as Turn-It-In or any other software before submitting the manuscripts to the Editorial Office, *Pertanika* Journal Publication Section.

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 the 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 Editor-in-Chief examine the paper to determine whether it is relevance to the journal's needs in terms of novelty, impact, design, procedure, language as well as presentation and allow it to proceed to the reviewing process. If it is 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 to 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 to attend to the reviewers' suggestions and perhaps additional advice on how to revise 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 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 the 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 online.

Pertanika Journal of

SCIENCE & TECHNOLOGY

Vol. 34 (2) Apr. 2026

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
Mohamed Thariq Hameed Sultan
SEARCA Chair of Agriculture Technology and Innovation

JOURNAL SECTION
Head of Section:
Fatimah Abdul Samad

Journal Officers:
Ain Haziqah Ab Razak
Navaneetha Krishna Chandran
Nur Izni Shafiqah Mohd Hamidi
Saundarya Saran

Editorial Assistants:
Zulinaardawati Kamarudin

Pre-press Officers:
Aisyah Umairah Jasmin
Engku Erina Zahirah Engku Mohd Suhaimi
Ku Ida Mastura Ku Baharom

IT Officer:
Kiran Raj Kaneswaran

Journal Section Office
Putra Science Park
1st Floor, IDEA Tower II
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia

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

PUBLISHER
UPM Press Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia
Tel: +603 9769 3606
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 Utilisation, 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, Optimisation, 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 authorisation of the Publisher.
Copyright ©2026 Universiti Putra Malaysia Press. All Rights Reserved.

Pertanika Journal of Science & Technology
Vol. 34 (2) Apr. 2026

Contents

Foreword i
Mohamed Thariq Hameed Sultan

Multi-classification of Chest X-ray Images Using an Enhanced Deep Learning Approach 601
Revathi A and Savadam Balaji

Skin Cancer Classification from Dermoscopic Images based on Convolutional Neural Network and Vision Transformer 625
Gan Mengyun, Hazlina Hamdan, Mahmud Dwi Sulistiyo, and Abdulrahman Dira Khalaf

Design, Manufacturing, and Testing a Versatile Test Stand for MP100S Integration 647
Huu Danh Tran

Measuring the Performance of Blockchain Systems with an Improved Consensus Mechanism using Scalability Metrics 671
Kolli Lalitha Kumari and Lalitha Surya Kumari

Association Rule Mining to Detect Significant Factors of Autism Spectrum Disorder at Early Stage 695
Tania Akter, Farzana Tasnim, Shahriar Hassan, Mohammad Hanif Ali, and Mohammad Shorif Uddin

JaNarX: Smart Soil Moisture Prediction using Jackal Optimiser Enabled Machine Learning Model in IoT Coupled Agricultural Applications 721
Seema Jitendra Patil and B. Ankayarkanni

Interactive Location-based Augmented Reality Virtual Chatbot 743
Lai Yee Jen, Goh Eg Su, and Johanna Ahmad

Ensemble Docking and Molecular Dynamics Pattern Similarity Analysis for the Discovery of Dengue Virus RNA-Dependent RNA Polymerase Inhibitor 763
Amaal Mohammed Salih Nasr, Mohd Basyaruddin Abdul Rahman, Muhammad Alif Mohammad Latif, Faridah Abas, Siti Munirah Mohd Faudzi, and Bimo A. Tejo

Optimisation of Silver Nanoparticle Synthesis Using Aqueous Extract of *Passiflora foetida* L. via Response Surface Methodology 787
Silvia Handayani, Alyza A. Azmi, R. Rudiyanto, Wan Iryani Wan Ismail, Kustiariyah Tarman, Faridah Abas, Wan M. Khairul, and M. Maulidiani

Green Synthesis of Carbon Dots from Empty Fruit Bunch Using Hydrothermal for Tetracycline Removal 823
Wan Nuraishah Wan Ishak, Ying Pei Lim, Huey Ling Tan, Noor Fitrah Abu Bakar, Ng Law Yong, and Chun Yang Yin

Effect of β-cyclodextrin Concentration on the Antioxidant Stability in an Avocado Oil Emulsion-based System 845
Carissa Eilin Ardana, Anggun Puspitarini Siswanto, Mohamad Endy Yulianto, Arva Rafif Adi Prayoga, Elsan Febiyanti, and Hermawan Dwi Ariyanto

Response Surface Optimisation of a *Strobilanthes crispus*-Loaded Red Palm Oil Nanoemulsion with Enhanced Physicochemical Stability 861
Ee Lynn Fong, Chui Fung Loke, Chin Ping Tan, Tze Ven Poh, Bee Hiong Tan, and Oi Ming Lai

OD-SAM: Automated Zero-shot Segment Anything Model for Optic Disc Segmentation 889
Bhuvaneswari S and Subashini P

Review Article 913
A Critical Review of User-Centric QoS Provisioning Frameworks in 5G Networks
Wai Leong Pang, Wen Hao Anselm Chow, Hui Hwang Goh, Swee King Phang, and Kah Yoong Chan

Improved Speech Enhancement using Parallel MVDR Beamforming and Coherent-to-Diffuse Power Ratio Post-filtering 951
Sureshkumar Natarajan, Mohd Khair Hassan, Raja Kamil, Faisul Arif Ahmad, Syaril Azrad, June Francis Macleans, Hussna Elnoor M. Abdalla, Nurbek Saparkhojayev, and Syed Abdul Rahman Al-Haddad

A Genetic Algorithm-optimised Hybrid Framework Integrating Statistical Forecasting Models and Extreme Learning Machine for Dengue Case Forecasting in the United States of America 975
Dipankar Das, Arijit Chakraborty, Sajal Mitra, and Soumyabrata Nag

Review Article 1005
Visualising Global Trends in Soil Infiltration Study by using Bibliometric Analysis
Nor Farah Atiqah Ahmad, Siti Nooraiin Mohd Razali, Suhaila Sahat, and Muhammad Azraie Abdul Kadir

Integrating Time Lag Effects in Predictive Modelling Using Random Forest for Early Detection of Bagworm Outbreaks 1025
Yi Peng Wang, Nurul Hawani Idris, Norhayu Asib, Mohamad Hafis Izran Ishak, and Alvin Lau Meng Shin

Design of an Integrated Entropy-based Framework for Dynamic Risk Weight Allocation and Mitigation Strategy Prioritisation in High-speed Rail Projects 1053
Yogesh P. Kherde, Uday P. Waghe, Radhika S. Thakre, Rajesh M. Bhagat, Sanskruti M. Padmawar, Yogeshwari D. Waje, Rishi S. Lohia, Anup K. Chitkeshwar, and Vaibhav Dhawale

TF-IDF Versus Linguistic Features: Evaluating Feature Sets for MBTI Personality Type Classification 1077
Janice Allison Anak Sabang, Stephanie Chua, and Puteri Nor Ellyza Nohuddin

Comparative Air Gasification of Agricultural Residues in an Indigenously Designed Fluidised Bed Gasifier 1107
Shahbaz Hussain, Abdul Nasir, Ch. Arslan, Shakeel Ahmad Anjum, and Sibel Irmak

Spectral Acoustic Characterisation for Bearing Fault Detection in Marine Propulsion Motors 1139
Iradiratu Diah Prahmana Karyatanti, Istiyo Winarno, Muhammad Taufiqurrohman, Dwisetiono, and Ardik Wijayanto

Quantification of Some Physical Properties of Matoa (*Pometia pinnata*) Fruit: Mass Modelling for Machine Designing and Postharvest Processing 1163
Ubaidillah Ubaidillah, Risty Dwi Wulandari Silalahi, Dimas Firmanda Al Riza, and Bambang Dwi Argo

FCA_Unet: An Enhanced UNet with Feature Coordination Attention for Wheat Leaf Disease Severity Assessment 1181
Hongyan Zang, Annie Anak Joseph, Kuryati binti Kipli, David Bong Boon Liang, Wanzhen Wang, and Rong Liu

An Interpretable Random Forest with SHAP Explanations for Multiclass Skill Level Classification Model in Malaysia's Labour Market 1203
Rabi'atul'adawiah Shabli, Ahmad Zia Ul-Saufie, Arief Gusnanto, and Nurain Ibrahim

How Gamification Motivates in Quranic Memorisation: A Method for Design and Suggestions for Future Research 1231
Siti Hasrinafasya Che Hassan, Syadiah Nor Wan Shamsuddin, and Nor Hafizi Yusof

Development of an Indoor Testing Facility to Standardise the Analysis of UAV Spraying Efficacy 1265
Siti Amni Ismail, Azmi Yahya, Ahmad Suhaizi Mat Su, Anas Mustafah, Norhayu Asib, and Emmanuel Otache Adanu

Foreword

Welcome to the second issue of 2026 of 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 27 articles: two review articles and the rest are regular articles. The authors of these articles come from different countries namely Bangladesh, China, India, Indonesia, Iraq, Kazakhstan, Malaysia, Nigeria, United Arab Emirates, United Kingdom, USA, and Vietnam.

The article entitled "FCA_UNet: An Enhanced UNet with Feature Coordination Attention for Wheat Leaf Disease Severity Assessment" presents a significant advancement in precision agriculture through the development of a robust deep learning framework for disease detection. Addressing the limitations of conventional segmentation models, the study integrates attention mechanisms and multi-scale feature fusion to enhance the detection of small and complex lesion patterns under challenging environmental conditions. The incorporation of lightweight backbone architecture further ensures computational efficiency, making the model suitable for deployment on resource-constrained agricultural devices such as drones and portable monitoring systems. The model achieved an impressive 89.85% mIoU, representing a substantial improvement over traditional UNet architectures, while also reducing parameter complexity. This contribution is particularly relevant in supporting early disease detection, minimising crop losses, and promoting sustainable agricultural practices through intelligent monitoring systems. Further details of this study can be found on page 1181.

The issue also features the article titled "An Interpretable Random Forest with SHAP Explanations for Multiclass Skill Level Classification Model in Malaysia's Labour Market," which underscores the growing importance of explainable artificial intelligence in socio-economic analysis and policymaking. While machine learning models are widely recognised for their predictive capabilities, their lack of interpretability often limits their practical adoption in policy contexts. This study addresses this challenge by integrating Random Forest with SHapley Additive exPlanations (SHAP), enabling both high predictive performance and transparent interpretation of key influencing factors. Using a large-scale dataset from Malaysia's labour market, the model achieves strong classification performance while identifying critical determinants such as education level, wages, and employment activity. The incorporation of interpretability not only enhances trust in the model but also provides actionable insights for policymakers in designing targeted interventions

for workforce development and human capital planning. This research represents a meaningful step towards bridging the gap between advanced analytics and practical decision-making in national policy frameworks. Further detailed information can be found on page 1203.

Another noteworthy contribution is the article entitled "Development of an Indoor Testing Facility to Standardise the Analysis of UAV Spraying Efficacy," which addresses critical gaps in agricultural mechanisation and pesticide application technologies. The study introduces an innovative indoor experimental facility specifically designed to simulate UAV spraying conditions in a controlled and repeatable environment. By enabling precise monitoring of rotor-induced airflow, droplet formation, and spray deposition patterns, the facility provides a reliable platform for evaluating the performance of UAV systems under varying operational parameters such as flight speed, payload, and nozzle configuration. This standardisation is particularly important given the inconsistencies associated with traditional field-testing methods. The proposed facility not only enhances the accuracy and reproducibility of experimental results but also contributes towards safer pesticide application, reduced environmental contamination, and improved efficiency in large-scale agricultural operations. As UAV technology continues to gain prominence in precision farming, advancements play a crucial role in bridging the gap between experimental research and real-world implementation. Full details of this study are available on page 1265.

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
Mohamed Thariq Hameed Sultan

Pertanika J. Sci. & Technol. 34 (2): 601 - 624 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Multi-classification of Chest X-ray Images Using an Enhanced Deep Learning Approach

Revathi A[1, 2]* and Savadam Balaji[1]

[1]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, 500075 Hyderabad, Telangana, India*
[2]*Department of Information Technology, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, 500090 Hyderabad, Telangana, India*

ABSTRACT

Healthcare practitioners can rapidly recognise lung problems due to the invaluable and essential role that chest X-ray imaging plays in lung diagnosis. Recently, deep learning techniques have gained popularity and demonstrated promising outcomes in automatically interpreting medical images, especially in the specialty of chest radiology. For classifying chest X-ray images into multiple classes, we propose an enhanced deep learning approach in this research. Initially, the input images are resized, normalised, and data augmented to improve the network's generalisation ability. The local and global features are extracted using the MobileNetV2 model. An Intelligent Prairie Dog Optimisation (IPDO) algorithm is employed to select the optimal features, thereby minimising the dimensionality of the features. This algorithm finds the optimal solution with a higher convergence rate. Finally, the multi-class chest x-ray images are classified using the Enhanced GoogLeNet (EGoogLeNet) model. The model parameters are optimised using the Adam optimiser. The proposed approach is evaluated using a large-scale C19RD dataset, and the findings indicate improved performance in lung multi-class diagnosis. Extensive experiments are conducted against existing state-of-the-art approaches to assess the proposed model's performance, yielding significant improvements in F1-score, recall, precision, and accuracy. Promising findings are found in the assessment of the proposed method's performance (EGoogLeNet), which achieves a multi-class classification accuracy of 99.41%. The proposed model performed better, enabling medical professionals to diagnose and treat patients more rapidly and effectively.

Keywords: Adam optimiser, chest x-ray images, deep learning, enhanced GoogLeNet, MobileNetV2 model

ARTICLE INFO

Article history:
Received: 18 September 2024
Accepted: 09 December 2025
Published: 01 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.01

E-mail addresses:
revathi_a@vnrvjiet.in (Revathi A)
balajis@klh.edu.in (Savadam Balaji)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Lung disorders, including opacity, fibrosis, viral pneumonia, and COVID-19 pneumonia, as well as tuberculosis, are significant global health problems that affect many people worldwide (Singh et al., 2025). These diseases are characterised by their detrimental impact on pulmonary function, most notably reducing lung elasticity (Prasath et al., 2025). Due to this decreased flexibility, the lungs can store less air overall, thereby compromising respiratory performance (Slimi et al., 2025). Certain lung diseases can spread quickly, particularly infectious diseases like pneumonia and opacity. This highlights the vital importance of early and precise diagnosis (Sharmili et al., 2026). Early diagnosis of these conditions is crucial because it enables the prompt initiation of proper care, which is necessary to slow the disease's progression and improve patient outcomes (Slika et al., 2024). Individual patients benefit from timely and accurate lung disease diagnosis, as it enables them to receive proper treatment. Maintaining public health is essential because it prevents the transmission of contagious respiratory diseases (Bhardwaj et al., 2025).

Using cutting-edge technology has become crucial in the dynamic and constantly evolving field of healthcare diagnosis, particularly in pulmonary health (Kaushik et al., 2025). In this regard, chest X-ray imaging has become a vital diagnostic tool. It provides an effective and non-invasive method for identifying and evaluating lung abnormalities (Odeh et al., 2024). For healthcare practitioners, this imaging technology is crucial. It is essential for the timely and precise detection of various lung diseases, making it a crucial component of lung disease management and treatment (Kumar et al., 2025). Chest X-ray (CXR) imaging remains one of the most widely used and accessible imaging modalities for diagnosing lung diseases due to its low cost, speed, and ability to provide sufficient anatomical information for detecting thoracic abnormalities (Satapathy et al., 2025). Although CXR may not match the resolution of CT scans, it still enables the visualisation of critical features such as infiltrates, consolidations, and nodules that are relevant for clinical assessment (Tekin et al., 2025). Clinicians need these intricate visualisations to correctly detect, monitor, and treat lung diseases (Okolo et al., 2025).

The importance of chest radiography in diagnosing and treating lung disorders is increasing as medical science advances, underscoring the need for ongoing research and improvement in this vital field of medicine (Gonçalves et al., 2025). However, CXR images are often affected by issues such as overlapping anatomical structures, low contrast for subtle lesions, and variations in image quality, which can obscure critical diagnostic details (Chen et al., 2026). Radiologists also face challenges, including inter-observer variability and difficulty in distinguishing visually similar abnormalities, underscoring the need for advanced deep learning approaches to enhance accuracy and consistency (Yadav et al., 2024).

Deep learning methods utilise extensive medical image datasets to identify complex patterns and anomalies that may be overlooked by more conventional techniques, thereby leading to a more comprehensive understanding of pulmonary disorders

(Maniruzzaman et al., 2024). As a result, deep learning in chest radiology signifies not only a breakthrough in technology but also a substantial improvement in the capacity of healthcare providers to accurately and successfully identify and treat lung conditions (Zhao et al., 2025). Incorporating deep learning into chest imaging improves diagnostic precision and encourages the creation of individualised treatment plans (Egala et al., 2025a). These developments represent a breakthrough in managing and treating pulmonary health junctures and are essential to improving patient outcomes (Varadharajan et al., 2025).

To identify different lung-related disorders, this research proposes a robust and enhanced multi-class classification framework utilising a deep learning model. Our method utilises CNNs due to their demonstrated ability to automatically learn hierarchical and spatially invariant characteristics from medical images, despite many previous studies having used CNNs for comparable diagnostic tasks. CNNs are especially effective at gathering local and global contextual information, which is crucial for identifying subtle abnormalities in chest X-rays, such as nodules, consolidations, or opacities. Furthermore, the proposed model integrates an optimised feature selection mechanism and an improved classification module, which together enhance the discriminative power of CNN-based representations. In contrast to conventional techniques and previous CNN variations, our improved CNN-based method ensures consistent performance across diverse imaging conditions, reduces false diagnoses, and enhances classification accuracy. The framework exhibits strong generalisation after being trained and validated on a sizable, well-annotated dataset, providing physicians with a diagnostic tool that is more precise, effective, and interpretable.

The significant contributions of the proposed research are:

- A novel hybrid framework is proposed that combines an efficient feature extraction mechanism using MobileNetV2 with an advanced classification model (Enhanced GoogLeNet) tailored for accurate multi-class classification of chest X-ray images.
- A new meta-heuristic feature selection algorithm, named IPDO, has been developed to optimise the high-dimensional feature space extracted from images. By selecting the most pertinent features with the least computational overhead, the IPDO algorithm significantly reduces redundancy and improves classification performance.
- An enhanced variant of the GoogLeNet model (EGoogLeNet) is designed and integrated into the framework, providing enhanced discriminative power and deeper feature representation for fine-grained classification across multiple lung disease classes, including COVID-19, normal, lung opacity, and viral pneumonia cases.
- Using the extensive publicly accessible C19RD dataset, extensive experimental validation is carried out. The proposed model yields higher performance results, demonstrating its superiority in both scientific research and practical applications.

The remaining part of this research is structured as follows: recent research-related works on chest X-ray image-based deep learning for disease detection and classification are presented in Section 2. Section 3 describes the methodology. The experimental results and comparison details are included in Section 4. Section 5 concludes the proposed research.

Problem Statement

Chest X-ray (CXR) imaging is one of the most widely used diagnostic tools for detecting a variety of lung diseases due to its cost-effectiveness, speed, and non-invasiveness. However, there are several significant challenges to automatically interpreting CXR images. These include overlapping tissues, low contrast across anatomical structures, slight variations in visual patterns among various lung diseases, and varied image capture settings. Even experienced radiologists find it challenging to reliably distinguish between diseases such as COVID-19, lung opacities, pneumonia, and normal lung conditions due to these complications.

Standard Convolutional Neural Networks (CNNs), while successful in binary classification tasks, often struggle with multi-class classification of lung diseases from CXR images. For number of classes, their accuracy decreases due to their limited capacity to capture contextual connections and fine-grained spatial data. Furthermore, issues such as uneven data distribution, intra-class similarity, and inter-class variation, all of which are prevalent in real-world CXR datasets, may not be adequately handled by traditional CNNs.

An improved deep learning framework that can reliably and accurately diagnose several types of lung disease from CXR images is urgently needed for advancement, in addition to addressing these limitations. The proposed research introduces a novel enhanced CNN-based architecture, specifically designed to handle the complex nature of CXR images and the challenges of multi-class disease classification. The primary objective is to improve early diagnosis, treatment planning, and patient outcomes by providing a diagnostic assistance system that is more reliable and practical.

Related Prior Works

To contextualise the novelty of our proposed framework, this section thoroughly examines previous contributions in the field, with an emphasis on architectural improvements, optimisation techniques, attention mechanisms, feature fusion, and lightweight deployment.

Convolutional Neural Networks with Architectural Enhancements

Ali et al. (2024) introduced CX-RaysNet, combining standard and group convolutional layers with small filters and dropout regularisation. This design enhanced subtle pattern recognition in CXR images, achieving 97.25% multi-class accuracy. However, it focussed

primarily on low-power device deployment, lacking a mechanism for dimensionality reduction of high-level features, which limits scalability in high-dimensional feature scenarios.

Similarly, XcepCovidNet, as described by Juneja et al. (2024), utilised transfer learning with hyperparameter tuning and outperformed multiple pre-trained models. While effective for COVID-19 detection, it did not introduce a novel feature selection or optimisation mechanism, limiting model interpretability and customisation.

By combining an improved feature extraction model (MobileNetV2) with a unique IPDO technique, which not only selects the most pertinent features but also reduces computational cost, our proposed model builds upon these studies.

Hybrid Models using Wavelet Transforms and MBConv Blocks

According to Bhosale and Yadav (2024), the wavelet features improved frequency-based analysis for five-class classification. Although their model achieved 95.5% accuracy, it required substantial computational resources and lacked effective mechanisms for model interpretability.

For optimal feature dimensionality reduction, our model employs IPDO, a metaheuristic feature selector, which reduces computational load while maintaining performance.

Attention Mechanisms and Ensemble Learning

Sayeed et al. (2024) utilised attention-based ensembles of Xception, EfficientNetB7, and DenseNet201, integrated with a Residual U-Net for segmentation. Although this approach achieved high accuracy (98.21% for binary), its complexity and ensemble design increased inference time and model overhead.

Our proposed method achieves better accuracy with significantly less complexity and training time by utilising a single upgraded classification model (EGoogLeNet) with Adam optimization and feature optimisation using IPDO.

Optimisation and Explainability in Deep Models

Abhishek et al. (2025) proposed DICA-Net, using Attention U-Net for segmentation and Pigeon Local Search Optimisation (PLSO) for feature selection. With an accuracy of 99.21%, it is one of the best-performing models. However, the optimisation algorithm is tailored to a specific task and may not generalise across diverse datasets.

Inspired by the behaviour of prairie dogs, our IPDO algorithm offers greater adaptability, enabling faster convergence and wider applicability across various lung diseases.

Balaha et al. (2025) introduced LDD-VTS, a ViT-based model with SHAP for explainability. Although interpretability is improved, the model relies on transformer architectures, which can be computationally intensive and less efficient for smaller datasets.

With a CNN-based architecture and lightweight feature selection, we prioritise efficiency and performance balance to enable real-time applications without compromising accuracy in our proposed research.

Feature Enhancement and Transformer-based Models

Egala et al. (2025b) employed a Stacked Residual Coordinate Network and innovative techniques, including the Gaussian Fourier Pyramid, Termite Alate Optimisation, and a Multi-GAN Transformer. While it achieved 99% accuracy, the multi-layer and transformer-based approach makes deployment on standard hardware more difficult.

In contrast, our approach utilises a modified MobileNetV2 + EGoogLeNet architecture, which is optimised for both training efficiency and deployment scalability.

CNNs with SE Blocks, Capsule Networks, and Hybrid Pipelines

Shanthi et al. (2025) presented a model combining SENet attention and capsule networks with CBSO optimisation. The system achieves 99% accuracy on the COVIDx CXR-4 dataset, primarily optimised for COVID-related data. Similarly, Zhu et al. (2025) combined VGG16 with self-attention and SE blocks, but their model was limited to six disease categories with an accuracy of 93%.

In contrast, our model is verified on the larger C19RD dataset and achieves 99.41% accuracy, outperforming many of the research studies above. It handles multi-class classification of lung diseases.

It is clear from the review of the literature that recent developments in deep learning for the classification of chest X-ray images have investigated a range of strategies, such as attention mechanisms, hybrid frameworks (e.g., CNN-transformers, CNN-wavelet), architectural modifications (e.g., group convolutions, MBConv blocks), and optimisation techniques. Although many of these models have produced encouraging outcomes, several drawbacks remain. Notably, most methods either lack an efficient and scalable feature selection mechanism or rely heavily on computationally expensive architectures, which can hinder real-time deployment and generalisability across diverse datasets.

These findings underscore the need for a classification framework that can handle high-dimensional image features and is not only accurate but also interpretable, computationally efficient, and effective. By combining MobileNetV2 for efficient and lightweight feature extraction, an effective IPDO algorithm for feature selection and dimensionality reduction that enhances model robustness and efficiency, an EGoogLeNet model designed for high-precision multi-class classification, and performance validation on a large-scale dataset (C19RD), our proposed approach directly solves these gaps. It also shows superior accuracy and generalisation when compared to other methods.

As a result, the literature analysis highlights the importance and necessity of our proposed method, establishing it as a significant advance in automated lung disease diagnosis utilising CXR imaging from both scientific and practical perspectives.

MATERIALS AND METHODS

The overall architecture of the proposed enhanced deep learning framework is illustrated in Figure 1. The following processes are included in a traditional framework for predicting and classifying lung disease using chest X-ray images: data collection, pre-processing and augmentation, feature extraction, feature selection, and classification. The spatial and channel attention mechanisms are combined with EGoogLeNet as the primary method for classifying diseases. For experimental analysis, a large multi-class chest X-ray dataset for lung disease is used.

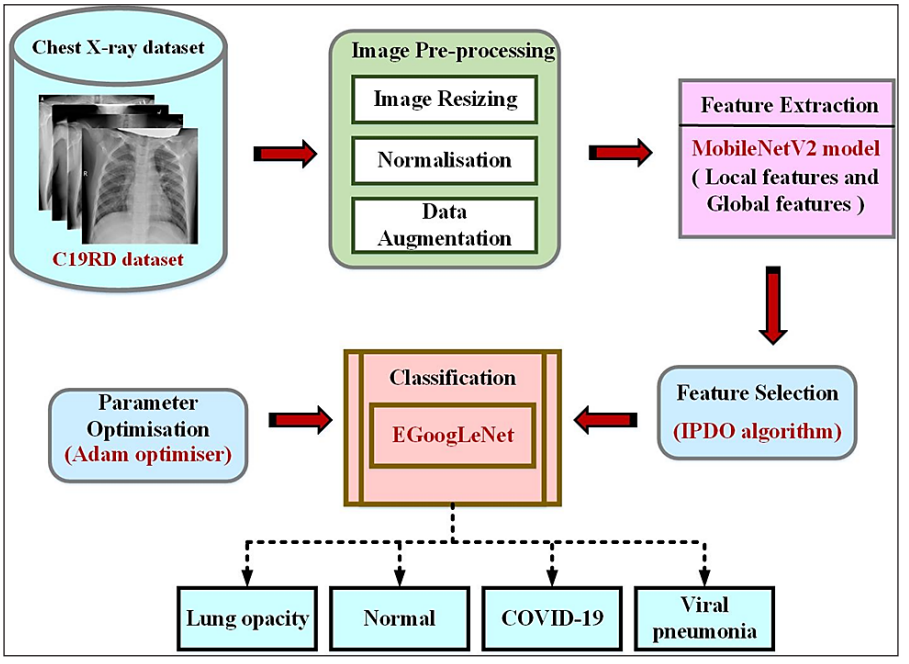


Figure 1. Overall architecture of the proposed enhanced deep learning framework

Database Description

The COVID-19 Radiography Database (C19RD) served as the main experimental collection in our research (Patel et al., 2024). This extensive collection comprises 21,165 chest X-rays, covering a variety of cases. The images of COVID-19-positive cases total 3,616, images of normal lung conditions indicate 10,192, and images of viral pneumonia

show 1,345. Representative sample chest X-ray images from the C19RD dataset used in this research are shown in Figure 2. The C19RD dataset is publicly available on Kaggle and was used in this study under its permissible open-access research license (Kaggle, 2024).

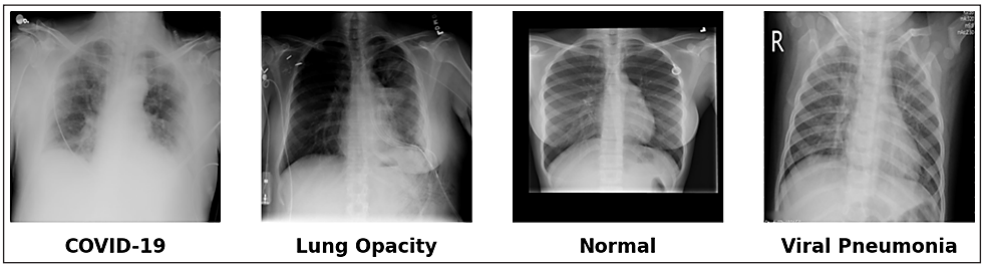


Figure 2. Representative sample images from the C19RD dataset showing (a) COVID-19; (b) lung opacity; (c) normal; and (d) viral pneumonia classes

Data Splitting

According to our research, the dataset was split into three sections: 60% was used for training, 20% for validation, and the remaining 20% for testing. Table 1 presents the distribution of the dataset before data augmentation across the training, validation, and testing partitions.

Table 1
Distribution of the dataset before data augmentation

Class Name	Training	Validation	Testing
Normal	6516	1629	2047
COVID-19	2291	572	753
Viral Pneumonia	867	216	262
Opacity	3833	958	1221
Total	13507	3375	4283

Image Pre-processing and Augmentation

To guarantee that the CNN architecture is compatible, image pre-processing is essential. Each image in the collection is initially resized to a consistent 224 × 224 pixels size. Additionally, image augmentation methods are employed to address the issue of the training set's limited number of images and enhance training effectiveness. Standard data augmentation techniques are used to rectify the training dataset's class imbalance. Even though the entire dataset contains 21,165 chest X-ray images, there is a notable disparity in the distribution of classes, particularly in the training set. For example, classifications such as "Viral Pneumonia" and "COVID-19" have far fewer samples (867 and 2,291, respectively), whereas the "Normal" class comprises 6,516 images. This imbalance can lead

to biased model learning, where the network favours the overrepresented class, reducing its generalisation ability and performance on minority classes. To artificially enhance the number of minority class samples, data augmentation was applied to the training set. This made each class roughly comparable to the "Normal" class, which consisted of about 6,500 images. Moreover, although the overall dataset size is substantial, data augmentation ensures robustness by simulating variability and preventing overfitting, especially since the proposed model is designed for high sensitivity across multiple lung disease categories.

Table 2 summarises the data augmentation techniques applied to the training dataset, along with their parameter ranges for balancing classes and improving model generalization. The augmentation, normalisation, and resizing procedures work together to maximise training efficiency and enhance the overall performance and reliability of the CNN model for lung disease diagnosis. Examples of the applied data augmentation techniques used to enhance dataset diversity are presented in Figure 3.

Table 2
Data augmentation techniques and parameter ranges used in the training dataset

Parameter	Value
Shear	0.35
Height shift	0.25
Zoom	(0.4, 0.9)
Rotation	(+30, −30)
Width shift	(0.7, 1.25)
Brightness Range	(0.5, 1.30)
Fill mode	Nearest
Horizontal flip	TRUE

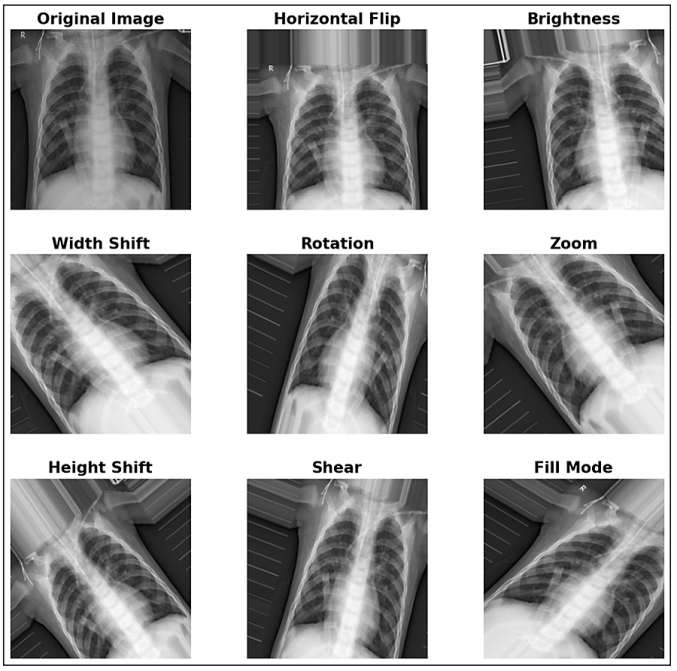


Figure 3. Example augmented images illustrating the application of data augmentation techniques to increase dataset variability

Feature Extraction

Due to its effective depth-wise separable convolutions, lightweight design, and robust performance on embedded and mobile vision tasks, MobileNetV2 is utilised for feature extraction in the proposed research. The MobileNetV2 model utilises Inverted Residuals with Linear Bottlenecks, which enables quicker computation and reduced memory consumption without sacrificing accuracy. When working with large amounts of medical imaging data, such as MRI, CT, and X-ray images, this is especially helpful. Additionally, MobileNetV2 achieves a balance between model accuracy and complexity, preventing overfitting and facilitating efficient feature learning from small datasets. It is a good fit for the proposed classification challenge due to its demonstrated ability to extract reliable and discriminative features from medical images.

In this research, the MobileNetV2 model is utilised to extract both local and global features of chest X-ray images, aiming to enhance classification performance. The MobileNetV2 model improves performance across a range of model sizes on various tasks and benchmarks. MobileNet models employ an excellent concept, utilising depth-wise separable convolutional blocks in place of expensive convolutional layers.

Feature Selection

In this research, the IPDO is employed to reduce the dimensionality of features. A metaheuristic algorithm that simulates prairie dog foraging is called the IPDO algorithm. Every day, prairie dogs participate in social activities such as foraging, cavern construction, cave maintenance, and predator defence. Thus, four-time intervals comprise the IPDO algorithm, based on the routines of prairie dogs. Based on a fixed mirror lifestyle, we then divide exploration and exploitation. In the spatial dimension, each prairie dog's foraging activity is represented by a $1 \times dim$ vector. Prairie dogs' mobility range is limited by the upper-bound UB and lower-bound LB, which prevents them from straying from their trajectory when foraging. A difficulty can be solved by placing each prairie dog in its own set at various places.

Classification

This research presents supervised deep learning for multi-class Classification of the most prevalent chest diseases. The multi-class chest X-ray images are classified using an enhanced GoogLeNet (EGoogLeNet) model.

Among the most widely used CNN models for image detection and classification is the GoogLeNet model, which has the advantages of low parameters and high classification and recognition accuracy. GoogLeNet successfully captures the features of different scales by using pooling layers of varying widths and parallel multiple convolution kernels. It presents techniques that facilitate network training and optimisation, including the

structure of Inception, design of depth and width, parameter efficiency, average pooling, and auxiliary classifiers. Figure 4 represents the standard Inception module introduced in the GoogLeNet (Inception v1) architecture. It is a powerful CNN model that successfully increases the network's capacity for classification, accuracy, and computational efficiency.

As per the attention mechanism, an enhanced GoogLeNet network model is proposed, building upon the GoogLeNet model to classify chest X-ray images precisely. By discovering the weights of each channel and space, the network can concentrate more attention on the information. It is of greater importance for the process of classifying images due to the attention mechanism, as different disease types correlate to separate locations in the X-ray image feature map. As a result, the attention mechanism selection enhances the network's capacity to adjust to variations in different images, allowing for more emphasis on the features that are most suitable for the classification process within the network. This enhances the model's capacity for generalisation and recognition accuracy.

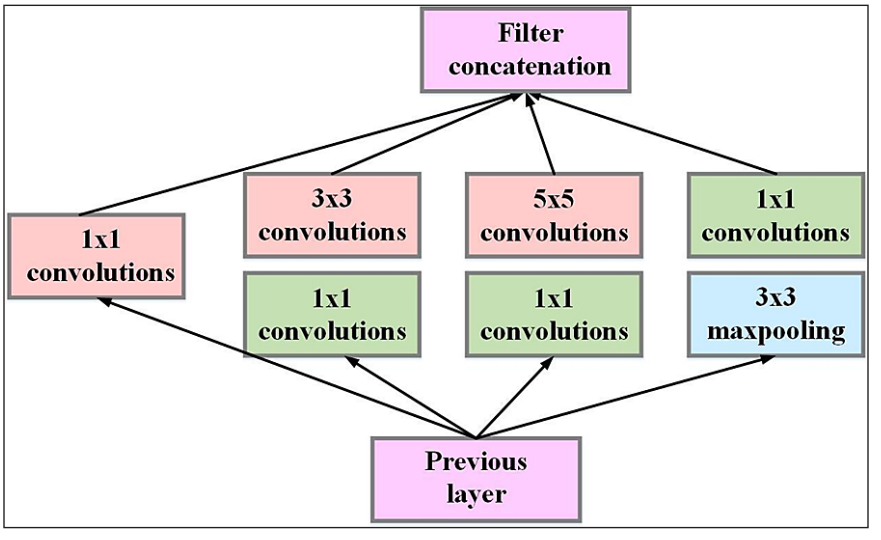


Figure 4. Inception module structure used in the GoogLeNet architecture

GoogLeNet is used by the model as the primary network. Prior to the input image reaching GoogLeNet's first convolutional layer, a weighted feature map is created by introducing a spatial attention mechanism. Following the final GoogLeNet stage, the weighted feature representations are routed to the GoogLeNet model's first convolutional layer and processed. The Channel Attention Mechanism module follows the Inception module. By combining convolution processes of various kernel sizes (1×1, 3×3, and 5×5) in parallel, the Inception module significantly improves multi-scale feature extraction in this situation. Because disease patterns vary in size and shape, a model is needed that can simultaneously capture both local and global features, which is especially useful in

medical imaging. The Inception module ensures that a rich collection of spatial features is first collected when combined with the Channel Attention Mechanism. Then, by emphasising the most relevant channels, the attention module can refine these features. The model's capacity to identify intricate and delicate lung disease patterns in chest X-rays is enhanced by this sequential structure, which improves classification accuracy. On the feature maps processed, the different channels' weights are modified adaptively, which is the main purpose of the GoogleNet model and the spatial attention mechanism in the network, prioritising the characteristics pertinent to the categorisation task. Additionally, this reduces the impact of redundant features, thereby enhancing the network's ability to adapt. Additionally, by reducing the network parameters and the model's complexity, this attention mechanism is used.

Figure 5 illustrates our proposed EGoogLeNet architecture, which incorporates spatial and channel attention mechanisms. In the proposed EGoogLeNet model, both the spatial and channel attention mechanisms are applied sequentially to enhance the feature map before it is passed into the main classification layers. Immediately following the processing of the input features by the first convolutional and pooling layers, the spatial attention mechanism is implemented. By highlighting the locations of the pertinent features in the image, this method calculates spatial weights that enable the model to focus on significant areas within the X-ray image.

The output of the spatial attention module is then subjected to the channel attention mechanism. To help the model concentrate on the most essential pathological features for classification, this technique adaptively recalibrates the relevance of each feature channel.

These two attention mechanisms are processed in a cascaded manner, with spatial attention followed by channel attention, which allows the model to learn the most essential disease feature maps. The spatial and channel weighting are combined into a single refined feature representation by fusing the feature maps produced by the two methods using element-wise multiplication as illustrated in Figure 5. For final classification, this fused attention-enhanced feature map is subsequently sent via the GoogLeNet architecture's final layers.

Attention Mechanism

The poor performance of direct feature map recognition can be attributed to several issues, including the image's low resolution and the slight variation in the tumour region across various types of chest X-ray images. Since different types of lung diseases have distinct characteristics, the classification model must determine the importance of image information and where it should be placed in the image to create a more useful feature map by assigning distinct weights to each component of the input feature map in the GoogLeNet model.

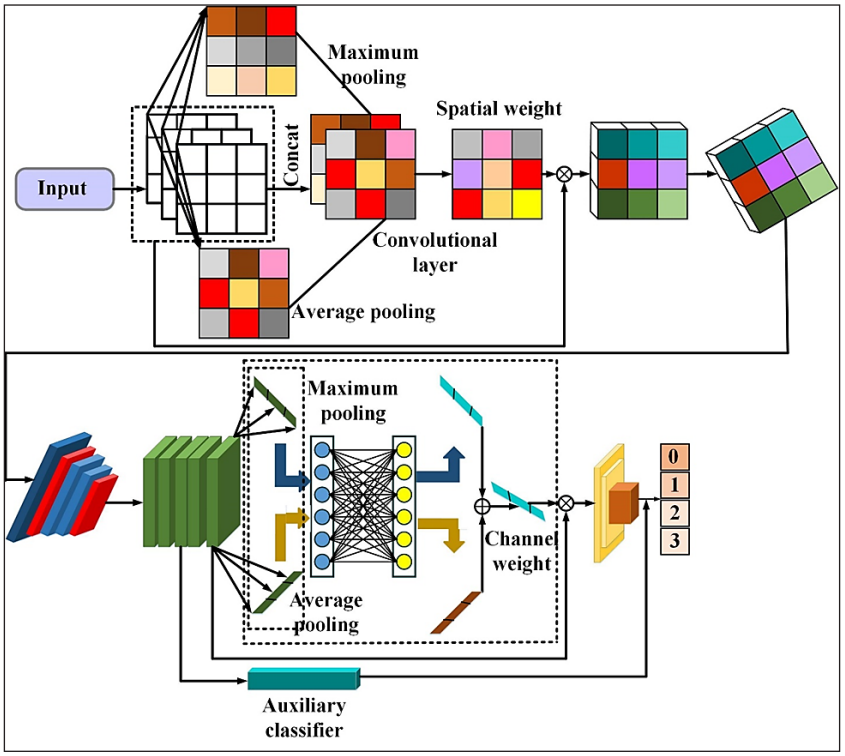


Figure 5. The attention mechanism-based general architecture of the EGoogLeNet model

The proposed EGoogLeNet design enhances the model's focus on the most important features, both spatially and across channels, by integrating the spatial and channel attention mechanisms sequentially to create a single attention module. There are two steps in the fusion process. The spatial attention mechanism is first applied to the input feature map, which highlights the locations of the image's most informative parts. To accomplish this, global max pooling and average pooling are applied along the channel dimension. The spatial attention map is then created by concatenating the data and feeding them into a convolutional layer, which is followed by a sigmoid activation. The original feature map and this spatial attention map are then multiplied element by element to highlight significant spatial regions.

Then, the output from the spatial attention is fed into the channel attention mechanism, which learns what features (channels) are most important for classification. Across spatial dimensions, it applies global max pooling and average pooling, and feeds a shared MLP with the combined features, using a ReLU activation. A sigmoid function is applied to the output to compute channel-wise weights. These weights are multiplied by the input feature map to enhance the discriminative features.

Experimental Settings

Table 3 presents the detailed computational environment used to implement and evaluate the proposed EGoogLeNet-based chest disease classification framework. We used the Python programming language.

Table 3
Summary of the computational environment and system configuration used to implement and evaluate the proposed EGoogLeNet framework

Name	Values/Types
Language	Python
GPU	NVIDIA RTX 3050 8 GB
RAM	16 GB
Operating System	Windows 10 Pro

To ensure appropriate training without excessive overfitting, our model was trained for over 100 epochs. The adaptability of the Adam optimiser in modifying network weights led to its selection. The model was able to converge gradually by using a 0.0001 learning rate and a 32 mini-batch size, which were chosen for their superior convergence qualities. Cross-entropy was the loss function employed, a common choice for classification problems.

RESULTS AND DISCUSSION

The experimental results and discussion on the multi-disease classification model based on the EGoogLeNet model are presented in this section.

Evaluation Metrics

Within the field of deep learning, model evaluation for a specific task depends on specific quality metrics. Recall, F1-score, precision, and accuracy provide data on the model's overall performance.

Equation 1 defines accuracy as the proportion of effectively classified images in the model to the total number of evaluated images. Equation 2 defines precision as the percentage of real positives in the subset of data the model has classified as positive. Equation 3 provides recall, also known as sensitivity, which measures the model's precision in identifying all pertinent samples within the true positive category. Furthermore, as demonstrated by Equation 4, the F1-score is a composite metric that balances recall and accuracy to provide a single measure of the classifier's accuracy.

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

$$Pr\,e\,cision = \frac{TP}{(FP+TP)} \quad [2]$$

$$Re\,c\,all = \frac{TP}{(FN+TP)} \quad [3]$$

$$F_1 - Score = 2 \times \frac{(Pr\,ecision \times Re\,call)}{(Pr\,ecision + Re\,call)} \quad [4]$$

where TN: True Negative, TP: True Positive, FN: False Negative, and FP: False Positive.

Training and Validation Evaluation

The training and validation accuracy and loss curves of the proposed EGoogLeNet model are depicted in Figure 6. The accuracy curve (Figure 6a) shows a sharp improvement over the first 15 epochs, with training accuracy rising from about 70% to over 90%. The model's great generalisation capacity and low overfitting are demonstrated by the training and validation accuracies, which progressively converge after 20 epochs and reach a peak of approximately 99%. The model appears to maintain constant performance throughout unknown data, as seen by the close relationship between the validation accuracy and the training accuracy.

During the early training phase, the training loss drops sharply from above 0.50 to below 0.15 in the first 20 epochs, as shown by the loss curve (Figure 6b). This pattern shows effective model parameter optimisation using the Adam optimiser and is consistent with the fast accuracy gains. After 60 epochs, the validation loss stabilises at a level below 0.10, closely tracking the training loss throughout the process. Even after extensive training over 100 epochs, the EGoogLeNet model does not experience overfitting, as evidenced by the lack of discernible divergence between the two loss curves.

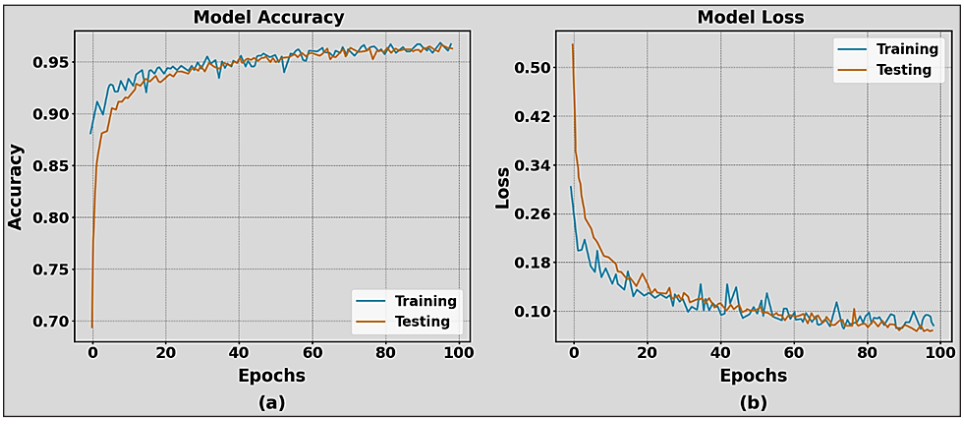


Figure 6. Training and validation accuracy and loss curves of the proposed EGoogLeNet model

These findings support the efficiency of the proposed hybrid framework, which combines the IPDO algorithm with MobileNetV2-based feature extraction to provide an optimal feature set that enhances classification accuracy and accelerates convergence. The proposed method's resilience for accurate and dependable multi-class lung disease diagnosis is further supported by the final stabilised high accuracy (99.41%) and low loss values across both training and validation stages.

Evaluation Results

The test set results are used to evaluate the proposed research's performance and compare the findings with those of recent existing deep learning models.

The classification results of the proposed EGoogLeNet model are summarised in Table 4. On the C19RD test set, the proposed EGoogLeNet model demonstrated a remarkable capacity to distinguish between multiple lung disease stages, achieving an overall classification accuracy of 99.41%. Precision and recall values are consistently high across all classes, as shown in the classification report, indicating that there are few false positives and false negatives. COVID-19 cases were specifically detected with 0.98 precision and 0.99 recall, guaranteeing dependable detection with minimal misclassification. Additionally, cases of lung opacity and viral pneumonia had a 0.99 recall, demonstrating the model's ability to identify subtle radiographic abnormalities that are frequently difficult for human interpretation.

The model can accurately identify healthy chest X-rays while minimising unnecessary false positives, as demonstrated by the Normal class's high precision of 0.99 and recall of 0.98. Even with differences in class size, the weighted-average precision of 0.99 and the macro-average precision of 1.00 show equal performance across all categories. The stability and dependability of the classification findings are confirmed by the consistently high F1-scores (0.98–0.99) for each class. The proposed improved deep learning model for multi-class chest disease classification yields an overall precision of 99.50%, recall of 99.50%, and F1-score of 99.25%. The model's promise as a reliable decision-support tool for fast and precise multi-class chest disease diagnosis in clinical settings is highlighted by its high precision and recall across all disease categories.

We conducted a series of controlled experiments by varying each hyperparameter within its tested range and monitoring the corresponding classification performance of the proposed EGoogLeNet model, thereby supporting the justification summary in Table 4. The outcomes demonstrated that, when compared to other settings, the ideal values indicated in Table 4 consistently yielded higher accuracy, F1-score, and convergence stability. For example, increasing the learning rate beyond 0.0001 resulted in unstable training with noticeable fluctuations in validation accuracy, while lower values slowed convergence without improving final performance. Similarly, batch sizes above 32 led to suboptimal

F1-scores due to the underrepresentation of minority classes within each batch, whereas sizes below 16 produced noisy gradients and slower optimisation. The optimal epoch value (100) was chosen because the validation accuracy reached saturation at this point, and further training did not yield additional improvement. These experimental observations validate the selection of the optimal hyperparameters presented in Table 4.

The class-wise prediction performance of the proposed EGoogLeNet model is illustrated through the confusion matrix in Figure 7. The matrix's diagonal dominance indicates that the great majority of samples were appropriately categorised. In 745 out of 753 cases,

Table 4
Classification results of the proposed EGoogLeNet model showing F1-score, recall, precision, and overall accuracy

Class	Precision	Recall	F1-score	Support
COVID-19	0.98	0.99	0.98	753
Lung Opacity	0.99	0.99	0.99	1221
Normal	0.99	0.98	0.99	2047
Viral Pneumonia	0.98	0.99	0.99	262
Accuracy	-	-	0.9941	4232
Macro Avg	1	0.99	0.99	4232
Weighted Avg	0.99	0.99	0.99	4232

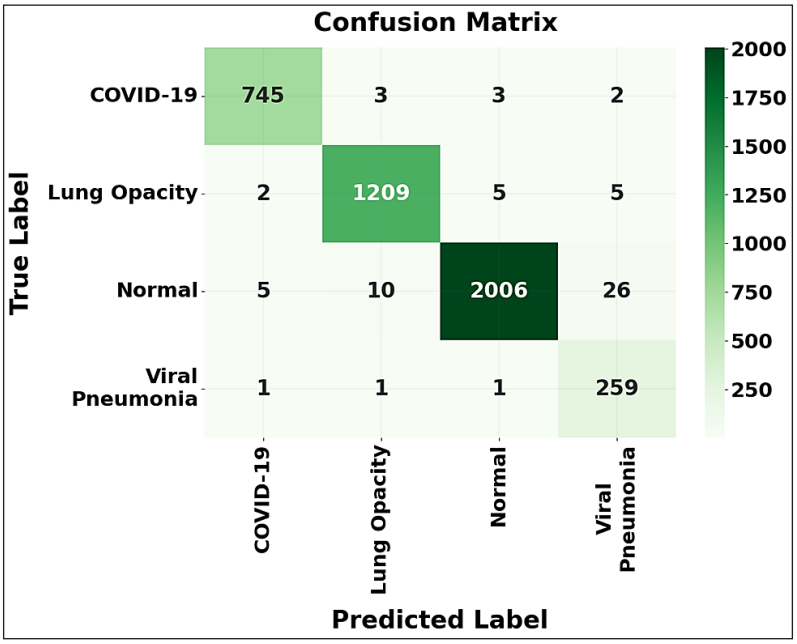


Figure 7. Confusion matrix illustrating the class-wise prediction performance of the proposed EGoogLeNet model

COVID-19 patients were accurately recognised, with only slight misclassifications into the categories of lung opacity, normal, and viral pneumonia. Lung Opacity achieved similarly high accuracy, with 1,209 correct predictions out of 1,221 cases, and only a few samples were misclassified into other classes.

For the Normal category, 2,006 out of 2,047 instances were correctly recognised, with the small number of errors primarily involving misclassification into Viral Pneumonia, likely due to overlapping radiographic features. Viral pneumonia exhibited the smallest sample size but still achieved 259 correct predictions out of 262 cases, showing the model's robustness even with limited training data.

These results, combined with the high recall, precision, and F1-scores, confirm the proposed hybrid framework's efficiency in learning highly discriminative features for each category. The minimal off-diagonal values indicate that integrating MobileNetV2 feature extraction, IPDO feature selection, and the EGoogLeNet classifier enables precise differentiation between visually similar lung disease patterns, making it well-suited for practical diagnostic applications.

The Receiver Operating Characteristic (ROC) curves and AUC values for each disease category obtained using the proposed EGoogLeNet model are shown in Figure 8. Each class has exceptionally high Area Under the Curve (AUC) values; COVID-19 achieved an AUC of 0.9973, indicating a nearly perfect trade-off between sensitivity and specificity. Pneumonia and Normal classes also exhibit strong separability, with AUC scores of 0.9959 and 0.9906, respectively. Despite being a challenging category due to its visual resemblance to other diseases, lung opacity achieved an AUC of 0.9863, confirming the robustness of the model's learned features.

For all classes, the ROC curves' sharp rise toward the upper-left corner indicates high true favourable rates and few false positives. The precision and recall scores of the classification report are consistent with these high AUC values, demonstrating the effectiveness of the MobileNetV2 feature extraction and IPDO feature selection process in producing highly separable feature representations. The proposed framework is a dependable tool for multi-class chest disease identification in actual clinical settings, as the ROC analysis demonstrates that it consistently maintains robust decision boundaries across various disease types.

Sample prediction outputs generated by the proposed EGoogLeNet model on test images are displayed in Figure 9. The green labels indicate correctly classified images, while red labels show misclassified samples. Several chest X-ray samples from each class are displayed to help visualise the proposed model's classification performance. The model's capacity to identify pertinent patterns for each disease category is demonstrated by the fact that most images are correctly labelled.

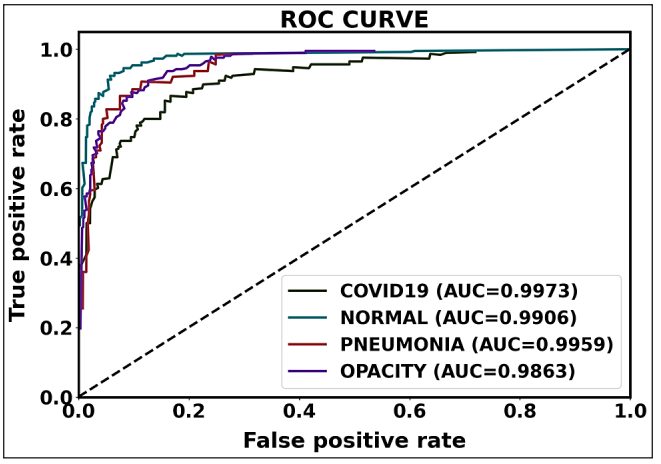


Figure 8. ROC curves with AUC values for each class obtained using the proposed EGoogLeNet model

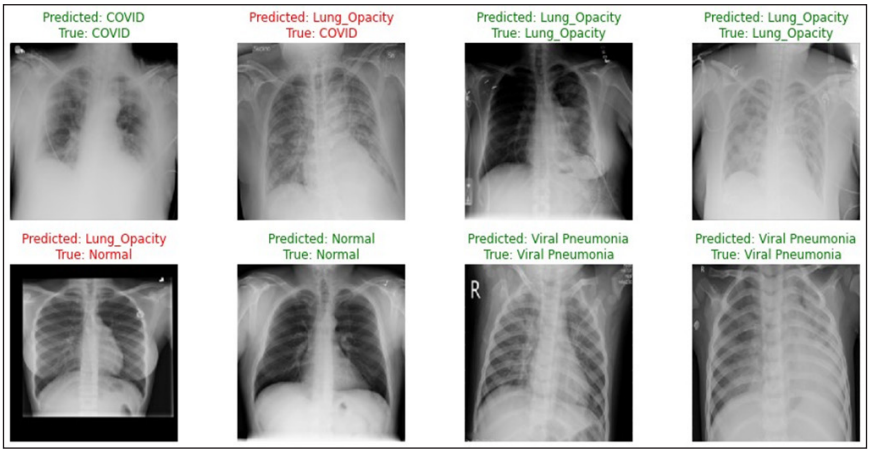


Figure 9. Sample predictions generated by the proposed EGoogLeNet model on the test images

Performance Comparison

Table 5 summarises a comprehensive performance comparison between the proposed EGoogLeNet framework and several recent state-of-the-art deep learning models for multi-class lung disease classification. The results clearly demonstrate that the proposed model outperforms the baseline on each evaluation parameter. In particular, the proposed model achieves an overall accuracy of 99.41%, significantly surpassing popular architectures such as CX-RaysNet (97.20%) and EfficientNet-B4 (96%). With F1-score, recall, and precision of 99.25%, 99.50%, and 99.50%, respectively, our proposed framework outperforms the nearest competitor (DenseNet-201) by significant margins.

Table 5
Comparative performance analysis of the proposed EGoogLeNet model against recent state-of-the-art approaches for multi-class lung disease classification

Reference	Classes	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
(Patel et al., 2024)	5	EfficientNet-B4	96	93	95.6	94.6
(Ali et al., 2024)	5	CX-RaysNet	97.20	97.02	96.43	97.25
(Thavasimuthu et al., 2024)	5	Multi-layer Perceptron Neural Network (MPNN)	92	95	97	98
(Karaddi et al., 2024)	6	Modified ResNet50	-	-	-	98.03
(Alshmrani et al., 2023)	5	CNN	96.48	97.56	93.75	-
Proposed Approach	4	EGoogLeNet	99.41	99.50	99.50	99.25

The IPDO algorithm, which minimises redundancy and improves class separability, and MobileNetV2, which enables effective and discriminative feature extraction, are responsible for the improvement. Furthermore, the EGoogLeNet classifier produces high-confidence predictions in all four categories by efficiently utilising the revised feature space.

Our framework achieves superior performance while maintaining computational efficiency when compared to models that use deeper or more sophisticated backbones (e.g., EfficientNet-B4, XceptionNet). This indicates that the proposed hybrid model not only performs better in terms of prediction accuracy than existing models, but it also holds great promise for real-world clinical use, where accuracy and inference speed are crucial.

Discussion

With an overall accuracy of 99.41%, the proposed EGoogLeNet-based multi-class classification framework shows very promising performance in recognising COVID-19, Lung Opacity, Normal, and Viral Pneumonia cases from chest X-ray images. The improved resilience and generalisation capability of the model are mainly due to the better pre-processing pipeline, multi-level feature extraction, and effective classification of abnormal patterns.

Even if the performance was excellent, some restrictions must be noted. The evaluation is restricted to the C19RD dataset, which may limit generalisability to other populations and imaging situations, as it may not accurately reflect the variances among multi-institutional clinical environments. Furthermore, despite the model's excellent discriminative capacity, its real-world interpretability remains limited. To boost clinical trust, explainability approaches, such as Grad-CAM or SHAP, must be further integrated. The diagnostic depth of the model may be limited by the lack of multi-modal imaging and the exclusive use of chest X-ray data, particularly in complex instances where CT or clinical metadata could provide complementary insights.

However, the results of this research have significant implications for both clinical practice and future research. The proposed model demonstrates potential for deployment as a computer-aided diagnosis tool to support radiologists, particularly in resource-limited settings, and can enhance triage efficiency during large-scale screening scenarios such as respiratory outbreaks. Expanding validation across various cohorts, integrating multi-modal data fusion for improved diagnosis accuracy, and incorporating explainable AI frameworks to enhance model transparency and clinical acceptance should be the primary goals of future research.

CONCLUSION

Using chest X-ray images, this research classified multiple lung diseases. An enhanced deep learning architecture was the foundation for the system's development and evaluation. The proposed research consists of four primary phases. To improve the dataset's quality, the initial pre-processing step handles image scaling, normalisation, and data augmentation techniques. Using the MobileNetV2 model, the second phase extracts the effective local and global features relevant to the disease. To reduce network complexity, important disease features are selected in the third phase using the IPDO algorithm. Finally, the EGoogLeNet model is used to classify the multi-class chest X-ray images. Our evaluation process on a large-scale C19RD dataset demonstrates the framework's outstanding potential, outperforming other recent state-of-the-art techniques for lung disease classification. The 99.41% accuracy achieved is evidence of the framework's accuracy and dependability. The framework performed exceptionally well, showing superior accuracy on key metrics such as AUC, F1-score, recall, and precision, with average values of 0.9925, 0.9925, 0.9950, and 0.9950, respectively.

We plan to expand this research in multiple ways in the future by utilising GANs to generate a more representative and extensive range of artificial medical images, improving the approaches' ability to be applied to several patient populations, combining it with electronic health records and additional diagnostic instruments to enable more comprehensive patient evaluations, and extending its use to cover further respiratory and thoracic diseases.

ACKNOWLEDGEMENT

We are grateful to everyone who has supported and encouraged us during this research. Your invaluable advice and assistance have been instrumental in completing this work.

REFERENCES

Abhishek, G. S. V. R., Singla, A., & Eshwar, D. (2025). DICA-Net: Optimising chest X-ray classification with attention U-Net and pigeon local search. *Network Modeling Analysis in Health Informatics and Bioinformatics, 14*(1), Article 20. https://doi.org/10.1007/s13721-025-00513-5

Ali, A., Wang, Y., & Shi, X. (2024). Detection of multi-class lung diseases based on customised neural network. *Computational Intelligence, 40*(2), e12649. https://doi.org/10.1111/coin.12649

Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal, 64*, 923-935.https://doi.org/10.1016/j.aej.2022.10.053

Balaha, H. M., Ahmed, R. A., & Balaha, M. H. (2025). LDD-VTS: An AI-based framework for lung disease diagnosis using vision transformers and SHAP. *Health Information Science and Systems, 13*(1), Article 47. https://doi.org/10.1007/s13755-025-00363-5

Bhardwaj, P., & Kaur, A. (2025). DLSAC-Net: An automated enhanced segmentation and classification network for lung diseases detection using chest X-ray images. *Multimedia Tools and Applications*, 1-26. https://doi.org/10.1007/s11042-025-20763-9

Bhosale, R. D., & Yadav, D. M. (2024). Customised convolutional neural network for pulmonary multi-disease classification using chest X-ray images. *Multimedia Tools and Applications, 83*(6), 18537-18571. https://doi.org/10.1007/s11042-023-16297-7

Chen, C., Isa, N. A. M., Liu, X., Ding, J., & Lu, L. (2026). MSA-Net: Multi-scale attention-based DenseNet for multi-label chest X-ray image classification. *Biomedical Signal Processing and Control, 113*, Article 109069. https://doi.org/10.1016/j.bspc.2025.109069

Egala, R., & Sairam, M. V. S. (2025a). Multi-instance Riemannian residual neural network with mountaineering team-based chest disease detection using chest X-ray images. *Research on Biomedical Engineering, 41*(3), Article 44. https://doi.org/10.1007/s42600-025-00423-5

Egala, R., & Sairam, M. V. S. (2025b). Multi-layer stacked residual coordinate termite alate network for multi-class lung diseases detection from chest X-ray images. *Applied Soft Computing*, Article 113393. https://doi.org/10.1016/j.asoc.2025.113393

Gonçalves, L. A., Junior, G. B., Bessa, M. L., Matos, C. E., & Fernandes, A. G. (2025). DualAttentionNet: A convolutional neural network for thoracic disease classification in chest X-rays. *Procedia Computer Science, 256*, 797-804. https://doi.org/10.1016/j.procs.2025.02.181

Juneja, A., Kumar, V., Kaur, M., Singh, D., & Lee, H. N. (2024). XcepCovidNet: Deep neural networks-based COVID-19 diagnosis. *Multimedia Tools and Applications, 83*(37), 85195-85225. https://doi.org/10.1007/s11042-024-19046-6

Kaggle. (2024). *COVID-19 radiography dataset*.

Karaddi, S. H., & Sharma, L. D. (2024). Classification of lung disorders in chest multi-modal images using hyper-parameter tuning and modified ResNet50. *Multimedia Tools and Applications*, 1-25. https://doi.org/10.1007/s11042-024-20097-y

Kaushik, S., & Verma, R. (2025). LD-DBGRTM: Vision transformer enabled lightweight attention-based deep learning model for pneumonia detection. *Biomedical Materials & Devices*, 1-25. https://doi.org/10.1007/s44174-025-00548-2

Kumar, S., & Bhowmik, B. (2025). ADConv-Net: Advanced deep convolution neural network for COVID-19 diagnostics using chest X-ray and CT images. *SN Computer Science, 6*(5), 1-22. https://doi.org/10.1007/s42979-025-03923-4

Maniruzzaman, M., Sami, A., Hoque, R., & Mandal, P. (2024). Pneumonia prediction using deep learning in chest X-ray images. *International Journal of Science and Research Archive, 12*(1), 767-773. https://doi.org/10.30574/ijsra.2024.12.1.0880

Odeh, A. A. R., & Mustafa, A. (2024). Explaining transfer learning models for the detection of COVID-19 on X-ray lung images. *International Journal of Electrical and Computer Engineering (IJECE), 14*(4), 4542-4550. https://doi.org/10.11591/ijece.v14i4.pp4542-4550

Okolo, G. I., Katsigiannis, S., & Ramzan, N. (2025). CLN: A multi-task deep neural network for chest X-ray image localisation and classification. *Expert Systems with Applications*, Article 128162. https://doi.org/10.1016/j.eswa.2025.128162

Patel, A. N., Murugan, R., Srivastava, G., Maddikunta, P. K. R., Yenduri, G., Gadekallu, T. R., & Chengoden, R. (2024). An explainable transfer learning framework for multi-classification of lung diseases in chest X-rays. *Alexandria Engineering Journal, 98*, 328-343. https://doi.org/10.1016/j.aej.2024.04.072

Prasath, J., Prabu, S., Mayil, V. V., & Saini, S. (2025). Optimised double transformer residual super-resolution network-based X-ray images for classification of pneumonia identification. *Knowledge-Based Systems, 311*, Article 113037. https://doi.org/10.1016/j.knosys.2025.113037

Satapathy, S. K., Cho, S. B., Mishra, S., Sah, S., & Mohanty, S. N. (2025). A federated learning approach for classifying chest diseases from chest X-ray images. *Biomedical Signal Processing and Control, 100*, Article 107107. https://doi.org/10.1016/j.bspc.2024.107107

Sayeed, A., Khansur, N. O., Srizon, A. Y., Faruk, M. F., Alyami, S. A., Azad, A. K. M., & Moni, M. A. (2024). An effective screening of COVID-19 pneumonia by employing chest X-ray segmentation and attention-based ensembled classification. *IET Image Processing, 18*(9), 2400-2416. https://doi.org/10.1049/ipr2.13106

Shanthi, S., & Mahalingam, M. (2025). Advanced chest X-ray image classification for early detection and treatment monitoring of respiratory conditions. *Biomedical Signal Processing and Control, 110*, Article 107990. https://doi.org/10.1016/j.bspc.2025.107990

Singh, K., Gaur, A., Kumar, S., Shastri, S., & Mansotra, V. (2025). Deep CP-CXR: A deep learning model for classification of COVID-19 and pneumonia disease using chest X-ray images. *Annals of Data Science*, 1-24. https://doi.org/10.1007/s40745-025-00601-3

Slika, B., Dornaika, F., Merdji, H., & Hammoudi, K. (2024). Lung pneumonia severity scoring in chest X-ray images using transformers. *Medical & Biological Engineering & Computing, 62*(8), 2389-2407. https://doi.org/10.1007/s11517-024-03066-3

Slimi, H., Cherif, I., Abid, S., & Sayadi, M. (2025). Enhanced pneumonia classification from chest X-ray images using hybrid capsule and LSTM networks with attention mechanisms and GAN integration. *Progress in Artificial Intelligence*, 1-14. https://doi.org/10.1007/s13748-025-00383-y

Tekin, V., Tekinhatun, M., Özçelik, S. T. A., Fırat, H., & Üzen, H. (2025). A deep learning-based EffConvNeXt model for automatic classification of cystic bronchiectasis: An explainable AI approach. *Journal of Imaging Informatics in Medicine*, 1-26. https://doi.org/10.1007/s10278-025-01688-z

Thavasimuthu, R., Hanumanthakari, S., Sekar, S., & Kirubakaran, S. (2024). Enhancing multi-class lung disease classification in chest X-ray images: A hybrid manta-ray foraging volcano eruption algorithm boosted multi-layer perceptron neural network approach. *Network: Computation in Neural Systems*, 1-32. https://doi.org/10.1080/0954898X.2024.2350579

Varadharajan, I., & Rathinavelayutham, S. (2025). An efficient lung disease classification using dove swarm optimisation based multi-scale faster-RCNN model. *Evolving Systems, 16*(3), Article 89. https://doi.org/10.1007/s12530-025-09708-7

Yadav, S., Rizvi, S. A. M., & Agarwal, P. (2024). Detection of lung diseases for pneumonia, tuberculosis, and COVID-19 with artificial intelligence tools. *SN Computer Science, 5*(3), Article 303. https://doi.org/10.1007/s42979-024-02617-7

Zhao, X., & Wang, X. (2025). Multi-label chest X-ray image classification based on long-range dependencies capture and label relationships learning. *Biomedical Signal Processing and Control, 100*, Article 107018. https://doi.org/10.1016/j.bspc.2024.107018

Zhu, J., Al-Qaness, M. A., Dalal, A. A., Tao, H. Z., & Alsamhi, S. H. (2025). Multiple lung diseases detection using advanced deep learning model with attention mechanisms and upsampling features. *Engineering Applications of Artificial Intelligence, 156*, Article 111038. https://doi.org/10.1016/j.engappai.2025.111038

Pertanika J. Sci. & Technol. 34 (2): 625 - 646 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Skin Cancer Classification from Dermoscopic Images based on Convolutional Neural Network and Vision Transformer

Gan Mengyun[1], Hazlina Hamdan[1]*, Mahmud Dwi Sulistiyo[2], and Abdulrahman Dira Khalaf [1,3]

[1]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
[2]School of Computing, Telkom University, Jl. Telekomunikasi No. 1, Terusan Buah Batu, 40257 Bandung, Jawa Barat, Indonesia
[3]Computer Centre Department, University of Fallujah, 31002 Fallujah, Anbar, Iraq

ABSTRACT

Skin cancer poses a vast worldwide health issue because of its fast progression and high mortality rate. Early diagnosis and precise prognosis are important for the treatment of these patients. In this paper, we propose an enhanced hybrid deep learning model composed of a combination of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for the dermoscopic image classification task. The designed architecture adopts ResNet50, EfficientNet, and ViT as strong feature extractors with an innovative attention module, MambaATT, for precise modelling and utilisation of global context information. Methods for statistical growth (rotation, flips, and scaling) have been tested to improve the model's generalisation and its variety, even if the dataset is short. When evaluating the entire dataset, we found that our proposed version CNN_ViT_MambaATT reaches an advanced-class accuracy of 92%, outperforming the traditional methods for individual networks. The findings highlight the power of combining strong CNN and VIT models for both mechanisms of interest and information growth strategies, offering a robust and accurate approach to early skin cancer analysis.

Keywords: Convolutional neural networks, data augmentation, deep learning, EfficientNet, MambaATT Model, ResNet50, skin cancer, vision transformer

ARTICLE INFO

Article history:
Received: 21 October 2025
Accepted: 26 December 2025
Published: 01 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.02

E-mail addresses:
gsgs84617@gmail.com (Gan Mengyun)
hazlina@upm.edu.my (Hazlina Hamdan)
mahmuddwis@telkomuniversity.ac.id (Mahmud Dwi Sulistiyo)
adk1973@uofallujah.edu.iq (Abdulrahman Dira Khalaf)
* Corresponding author

INTRODUCTION

Cancer is a medical condition consisting of deviant proliferation, division and distribution of cells throughout the human body by way of the lymphatic system and

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

blood, subsequently destroying healthy tissues. Skin cancer is one of the most frequent forms of cancer worldwide. It is caused by the overgrowth of skin cells, usually induced by excessive contact with UV light or carcinogens. Quickened development of modern industry and chemical industry Results to pollution and disaster for environment deployed Environmental factors such as industrial pollution and ozone depletion have led to a global rise in skin cancer occurrence. For 2025, it is expected that there will be approximately 104,960 new cases of melanoma in the United States, with around 60 550 in men and roughly 44,410 in women. Similarly, the projected number of deaths as 5,470 and 2,960 for males and females, respectively (Siegel et al., 2025). New cases in the USA, according to projections by sex, are shown in Figure 1. These numbers highlight the critical need for early and precise diagnostics.

The most lethal are the rapidly advancing, high-mortality tumours. In particular, the first identity complements the survival rate and emphasises that precisely adequate clinical equipment is necessary. The symptoms of skin cancer in the early stage are like those of a non-cancerous mole. "Low inter-class variance and high intra-class variance" in skin lesions is an extraordinary feature that brings a great challenge to dermatology specialists. Diagnosing such diseases by purely visual characteristics assisted by nonspecific histopathological biopsy could be misdiagnosed and have loss of diagnosis, and this would make patients lose their best treatment time. To increase the accuracy of skin

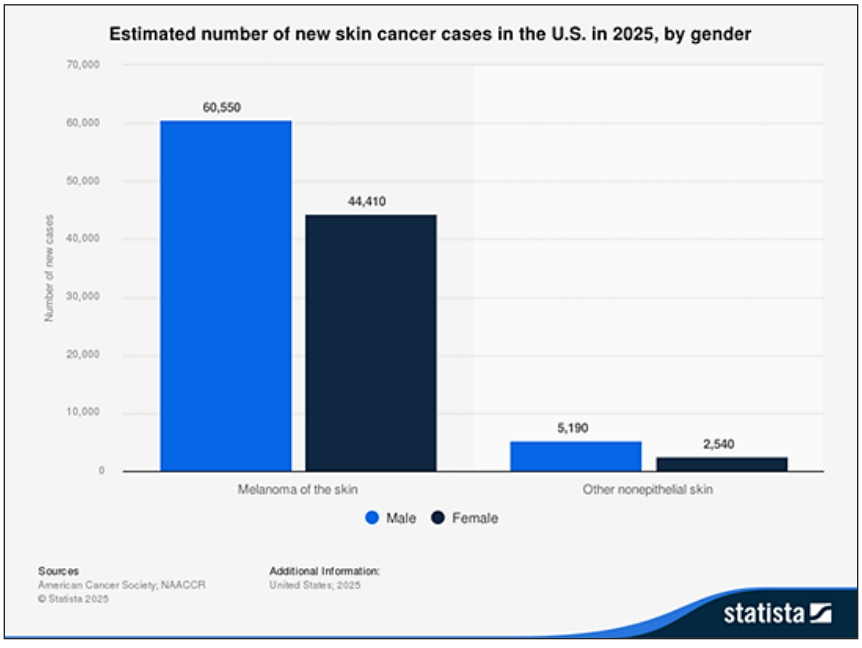


Figure 1. Estimated number of new skin cancer cases in the U.S. in 2025, by gender (Siegel et al., 2025)

disease diagnosis, the dermoscopy diagnostic technique has been developed and has been working well in clinical analysis (Kittler et al., 2002). Dermoscopy is a surface imaging method that utilises optical imaging for deeper skin lesions resulting in appearance of most affected areas.

In comparison to a single visual inspection, dermoscopy allows experts to diagnose and evaluate lesions under the direct observation with high resolution images, thus offering more accurate treatment plans for patients and leading to higher survival of skin cancer patients. Nevertheless, because the structure of dermoscopic images is so complex, its accuracy largely relies on physicians' expertise. The challenge is in the variability of the skin lesions which makes automatic classification very hard. This is primarily due to features common among diverse skin lesions can give rise to misdiagnoses. Second, it changes within the same category partially overlap with those of the lesion subtype (Goyal et al., 2020). A more accurate, rapid and objective method for automatic classification of skin cancer is therefore of great importance for the patients as well as medical professionals.

Deep learning has advanced and skin cancer classification is no exception including convolutional neural networks (CNNs) and vision transformers (ViTs). CNNs can be recognised by utilising local spatial features information in dermoscopic images, while ViTs use a self-attention mechanism. This work critically investigates the development of CNN and ViT architectures by demonstrating impactful hybrid models that harness their individual advantages. These hybrid processes have improved lesion detection and classification, providing a potential solution to current drawbacks of dermatology practices. Skin cancer in particularly melanoma, remains an important global health challenge due to its aggressive behaviour and associated fatal outcome when diagnosed late. Data-driven diagnostic systems have ditched a rule-based framework to enhance clinical performance and robustness by employing deep learning methods. CNN is the conventional backbone for dermoscopic image processing. The CNN structure is partitioned into local spatial patches, and it proves very effective in the image-based skin ulcer classification. As CNNs can capture hierarchical features from image data, they have been significant in the field of medical image analysis. Baykal Kablan and Ayas (2024) proposed an automatic clothing recognition function using a convolutional model with 97.1% accuracy on the ISIC 2019 dataset. It is a paradigm shift to treat images as sequential data instead of CNNs, which enables the recovery of coherence on the global context.

Recently, ViT was introduced in medical imaging, and it has attracted significant attention for its ability to capture global context through self-attention. A combining of CNN, ViT, and Xception introduced model achieving 95.46% precision and 96.74% accuracy on the HAM10000 dataset. The model effectively captured both local details and global contexts, proving robust in complex classification tasks (Ali et al., 2025). Reis and Turk (2024) suggested the MABSCNET model, a hybrid VIT architecture, which

received 100% accuracy on the IC 2018 data set with significant performance (92.74%) with a combination of CNN and ViT architecture. Arshed et al. (2023) reported the efficacy of ViT architecture over traditional CNN models, achieving 92.14% accuracy in multi-class skin cancer classification. A hybrid CNN-transformer model with focal losses to address class imbalance. This model effectively used CNN for early local functional extraction and a transformer model for global meaningful interpretation, and significantly improved classification performance in ISIC 2018 (Nie et al., 2023). Yang et al. (2025) presented a new attention-based approach at multiple levels, increasing the exact speed by 95.05%. Abbas et al. (2023) proposed Assist-Dermo, a lightweight transformer-based CNN hybrid achieving 95.6% accuracy through optimised architecture and preprocessing strategies.

One of the challenges of early detection is that malignant and benign tumours are often indistinguishable from one another. Early detection is vital; if the diagnosis is made at early stage, cure rate is more than 95% while it plunges to approximately 14% in advanced stages (Khan et al., 2021). Their morphological resemblance, size and pigmentation make them frequently being misdiagnosed by visual observation and histological diagnosis only. This limitation might delay treatment and contribute to lower overall survival. Dermoscopy has been widely used to improve the accuracy of diagnosis. This non-invasive imaging tool improves the visualisation of subsurface skin structures and treatment planning (Kittler et al., 2002). Nevertheless, accuracy of dermoscopic diagnosis is largely dependent on the experience of the physician while interobserver variation remains an issue. Because of skin lesions' intrinsic variability, it is challenging to classify skin lesions automatically. ViTs interpolate between the model and classic provisions in case of an optimal amount of regularisation and capacity (Flosdorf et al., 2024). This way skin cancer will be automatically classified with the patient and doctor faster, more accurately, and unbiasedly.

An important aspect of the investigations that have been performed on DL models for skin cancer classification is to compare the metrics such as accuracy, sensitivity, specificity, and F1 score. Research reported that the hybrid models outperform CNN or VIT alone. Despite some progress, various challenges bias, lesion-awareness uncertainty, real-time deployment, and lack of diversity or clarity remain. The fusion of CNNs and ViTs has boosted the skin cancer classification task by a considerable margin. Recent surveys Meedeniya et al. (2024) highlight the absence of integrated architectures combining CNNs, Vision Transformers (ViTs), and state-space models in medical imaging. In this paper, to cope with these issues in automatic diagnosis of skin cancer, we propose the combination of convolutional and transformer models. Despite significant advancements, there are still some important missing links in combining the local and global feature learning into effective and interpretable networks. Three central questions guide the research:

1. Can a hybrid CNN–ViT–MambaATT architecture outperform individual CNN and ViT models for dermoscopic image classification?

2. Does the MambaATT state-space mechanism improve discriminative feature extraction compared to traditional attention?
3. How does multi-branch feature fusion influence performance on limited medical datasets?

Recent state-of-the-art methods for skin cancer image classification have shown good to mixed levels of performance. Huang et al. (2024) used an ensemble-based attention CNN model, and their accuracy was 87.8% while Naz et al. (2025) achieved an accuracy of 86.5% using federated VGG19-XAI architecture. In another study, Lea (2025) used topological data analysis along with machine learning and presented 91.25% accuracy on a binary dataset of a similar kind. Nevertheless, the performance accuracy is still low. This study's novelty lies in combining the strengths of ResNet50, EfficientNet, and ViT via tensor-based feature fusion, and in refining the resulting features with the lightweight MambaATT module. The proposed approach captures both spatial and contextual dependencies with fewer parameters, improving accuracy and interpretability.

The objective of our study is to construct an accurate diagnostic model for differentiating dermoscopic tumour images as benign or malignant, with these three contributions:

1. Applying data augmentation to enhance generalisation with small dataset.
2. A composite model consisting of ResNet50 (He et al., 2016), EfficientNet (Tan & Le, 2019), ViT and MambaATT.
3. Evaluation of the proposed model against bench models to show improved classification accuracy.

This paper is organised as follows: Section 2 provides background on the research domain and motivation to work in the field of skin cancer classification: this includes a discussion of current state-of-the-art research, applied methodology and an overview of relevant literature addressing the problem discussed throughout. Section 3 describes the approach in this study and presents information on algorithms and operations. Implementation steps and results of the implementation are presented in Section 4. The reproduction of the base work and the improvement are described and compared. Section 5 provides the conclusion of the paper. Finally, Section 6 discusses limitations and future research.

RELATED WORK

In recent years, deep learning has brought significant improvement in medical imaging and outperformed conventional machine learning for several tasks, including skin cancer classification. In this section, we briefly describe the main building blocks of the deep learning image classification landscape. The emphasis is on convolutional neural network (CNN) architectures and their development from early ideas to the current state of the

art. The power of transfer learning and ensemble training is explored, emphasising how they can be utilised across a range of tasks. Vision Transformer and Vision Mamba are particularly highlighted, together with discussing techniques that affect model generalisation and feature extraction.

Since the introduction of AlexNet in 2012, deep neural networks have become the workhorse for computer vision applications, especially dermatological image diagnosis (Krizhevsky et al., 2017). Unlike traditional methods, where the features are designed manually, deep learning can automatically learn detailed patterns from input data based on convolutional and pooling operations. CNNs are the most common architecture in this area. They effectively model the spatial hierarchy in images using a localised representation and have shown reliability on different classification problems (LeCun & Bengio, 1998). CNN-based models, including AlexNet (Krizhevsky et al., 2017), DenseNet (Mobiny et al., 2019), and ResNet (He et al., 2016) have been widely used in skin cancer research. Some studies have improved the performance of convolutional neural networks for skin lesion classification. Han et al. (2020) included attention mechanisms into ResNet-50, resulting in a sensitivity of 76.8% and a specificity of 90.6%, while Tschandl et al. (2020) documented an accuracy of 80.3% on the HAM10000 dataset utilising ResNet-34. Mahbod et al. (2020) developed a multi-scale fusion CNN achieving 86.2% accuracy on the same dataset. In a different approach, Ali et al. (2022) employed EfficientNet in conjunction with transfer learning and preprocessing techniques, including hair removal, to attain an accuracy of 87.91%. Huang et al. (2021) developed a lightweight deep learning network for the classification of skin cancer, with an accuracy of 85.8%.

Despite this development, CNN's obstacles to capturing global features, researchers use transformer-based architecture, especially the visual transformer (VIT), which uses a self-attention technique to achieve global features (Chen et al., 2020; Dosovitskiy et al., 2021). Unlike CNNs, which need to stack several layers for perceiving the global patterns, Wits might capture relations between widely separated regions in one layer. The main idea of ViT is to cut the input images into non-overlapping patches and transform them into embeddings, which are subsequently fed through stacked transformer layers aided by self-attention. This enables the model to reason about the relative significance of many image regions. Below are some of the advantages that Vision Transformers have over traditional CNNs:

- **Stringer Long-range Information:** The CNNs are sensitive to the input size and objects in large images, thus making it hard for them to capture the global context. Vision Transformer models can capture long-range feature relations in images by self-attention, hence achieving better global information extraction.
- **Scale Invariances:** Traditional CNN-based models use fixed-size input, which makes it difficult to deal with images of other resolutions. Vision Transformer

models that cut images into small patches can better generalize to the input images of different resolutions.

- **Better Interpretability:** The output of Vision Transformer can indicate the importance of patches, which leads to a better understanding of the decision model.

It has been shown in the recent state of the art that hybrid models using either CNN or transformers are successful for skin cancer detection. For example, Pon Selchiya et al. (2023) presented a CNN-ensemble for melanoma classification and Ali et al. (2025) developed xCViT, where the authors combined a CNN, ViT, and Xception model for better explainability. However, these architectures are based only on self-attention operators and do not consider latent state-space modelling. Abbas et al. (2023) studied ensemble hybrid networks, nevertheless, without a systematic way to efficiently model the long-range features. The CNN_ViT_MambaATT differs by employing the Mamba module, which is a selective state-space mechanism designed to permit token mixing at fewer computational costs, allowing for better context memory.

Another study by Ozdemir and Pacal (2025) demonstrated that VIT can outperform CNN on multiclass classification when regularisation and followed training are enhanced. Vit enhances the generalisation by showing that patches indeed matter in the decision. However, some challenges are faced. They are also less biased than CNNs and need large datasets (Reis & Turk, 2024; Tay et al., 2021). Furthermore, the constrained local modelling capacity and overfitting problem within deep layers would also hamper its performance. More recent studies have attempted to loosen these constraints by looking at state space models (SSMs), such as Mamba (Gu & Dao, 2023), whose goal is to allow the modelling of linear time series with a bigger motivational bias. Zhu et al. (2024) proposed a visual-specific SSM with a good trade-off for 2D-selective scanning, capturing both spatial and sequential information. This changing environment indicates the necessity for a coupling design that allows for taking the advantages of CNNs, transformers, and SSMs.

METHODOLOGY

In this section, the architecture of our project is described, with deep learning model architectures CNN_ViT_MambaATT and transfer learning techniques that employ pre-trained ResNet50, ViT, and EfficientNet. The confusion matrix, as well as accuracy, precision, recall, and F1 score, are used to examine the performance of the model.

Features were taken from the penultimate convolutional block of ResNet50, EfficientNet, and from the final transformer encoder output of ViT. These high-level representations encode complementary visual semantics: local patterns in CNNs and global relationships in ViT. The reference and query features were stacked using a tensor stacking operation to form a single representation along a new dimension. The stacked tensor was further fed into MambaATT for learning sparse spatial-temporal dependencies. MambaATT is used as a post-fusion refinement module and does not serve as a substitute

to any CNN or ViT block. It directly manipulates the concatenated feature tensor, improving inter-model dependencies via selective state transition. To counteract the impact of contributions between backbones, a weighted average operation was performed prior to fusion. A Kaggle community dataset based on the ISIC Archive is used for this study. The dataset contains a total of 3,297 dermoscopic images divided into benign and malignant images. The dataset is balanced between two classes Table 1, allowing for a fair skin cancer classification comparison between different approaches.

Table 1
Distribution of training and test sets in the ISIC dataset

Type	Train	Test	Total
Benign	1440	360	1800
Malignant	1197	300	1497
Total	2637	660	3297

The medical images have been pre-processed and scaled to 224×224 pixels. The dataset is split into a training set 80% and a testing set 20%. This method ensures that a significant fraction of the data is used for model training, enabling the model to learn from a large dataset. Concurrently, allocating 20% of the data for testing facilitates a significant assessment of the model's efficacy on unfamiliar data. This method identifies overfitting when the model excels on training data but underperforms on novel data, thereby ensuring the model's generalisability. Figure 2 shows benign and malignant images taken from the dataset.

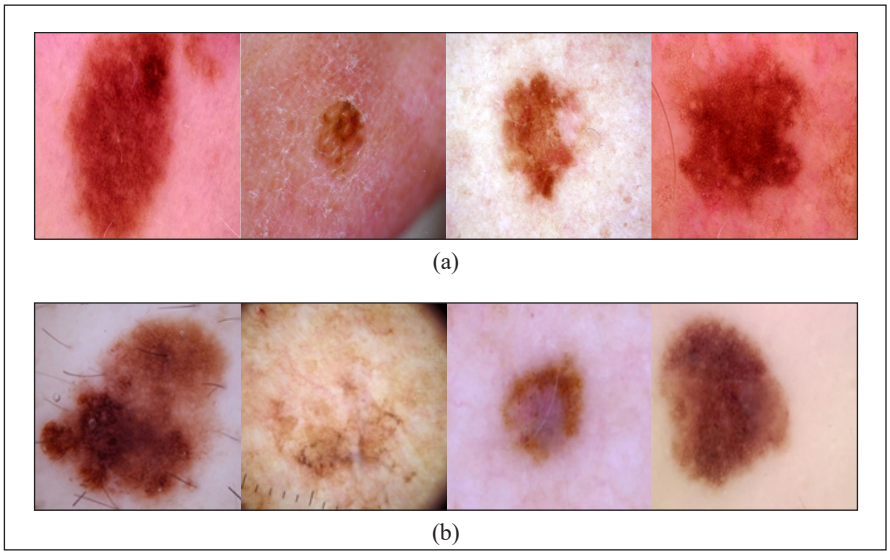


Figure 2. Example of images taken from the ISIC dataset (a-Benign, b-Malignant)

Dataset Preparation

Data augmentation refers to the process of applying a variety of transformation operations on some source images in a small-sized image dataset to create additional training data

that can help enhance the model's generalisation (Mikołajczyk & Grochowski, 2018). Since the dataset size is relatively small, we augment the training data by horizontally and vertically flipping and randomly rotating from 0 to 45 degrees. Figure 3 shows images with different sorts of augmentations applied. By doing so, the dataset is enlarged and enables the trained network to better classify data.

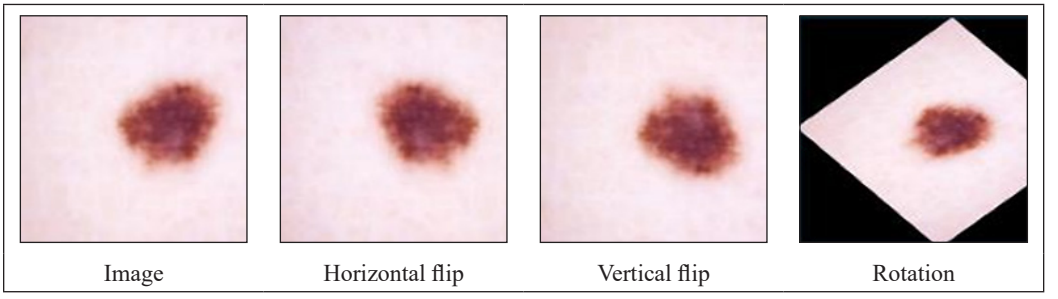


Figure 3. Comparison of the images transformed in different ways

CNN_ViT_MambaATT

Skin cancer image classifiers must use the datasets in a large size, and features extraction from skin image analysis is considered one of the major problems in this application. To solve this, the CNN_ViT_MambaATT module is proposed with a hybrid architecture for image classification on public datasets using transfer and ensemble learning. The CNN_ViT_MambaATT module provides an accurate diagnosis of skin diseases and determination of the malignancy or benignity. Combining CNNs with ViTs exploits the advantages of both models. CNNs are good at extracting local information, while self-attention in ViTs focusses on modelling global interaction information. This combined method will enhance general performance on image classification. On the contrary, CNNs generalise better with less data and achieve higher accuracy compared to ViTs. On the contrary, ViTs are efficient to learn from a small number of images. The general flowchart of the process is illustrated in Figure 4.

Feature Extraction

This study utilises a hybrid feature extraction framework, involving the fusion of 3 already pre-trained deep learning architectures: ResNet50, EfficientNet and Vision Transformer (ViT). These models separately process the input photos and produce high-dimensional feature vectors that capture diverse aspects of the image data. ResNet50 employs residual learning that allows for training of deeper networks and the capture of intricate low-level and mid-level visual patterns. Our EfficientNet introduces a new efficient model scaling method that balances all dimensions of depth/width/resolution and leads to better

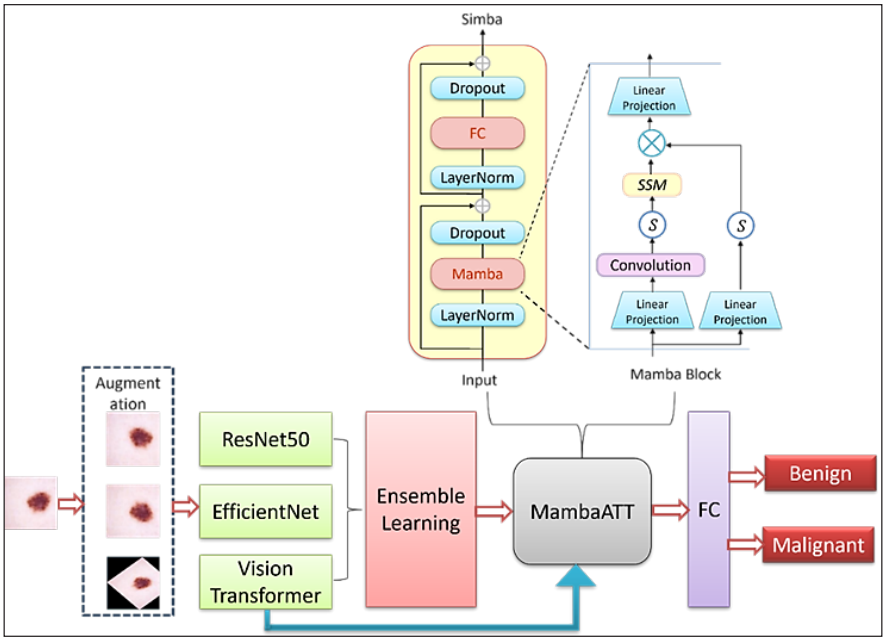


Figure 4. The proposed framework of skin cancer classification

performance. ViT applies self-attention to image patches so that visual structures can be generally recognised. We select the model out of architectural considerations and to have an effective way to benchmark image classification performance. Autonomous processing implies that each model has distinct functionality, leading to diversity and resistance.

In ensemble learning, stack the ResNet50, EfficientNet, and ViT features using tensor stacking. This approach concatenates the feature vectors along one extra tensor dimension, so that subsequent layers can learn from all three models' full spectrum of information. Ensemble learning is essential to facilitate generalisation and predictive power. Architecture combines multiple types to alleviate the shortcomings of individual models while exploiting their advantages. This is especially useful in medical imaging, where lesions may differ greatly in shape, colour, and texture. For example, the ViT may be able to recognise global context while ResNet50 and EfficientNet may have an advantage in detecting localised or hierarchical features.

Attention-based Weighting Mechanism

As illustrated in Figure 5, to reconcile the cooperation of three feature extractors ResNet50, EfficientNet, and ViT, an attention-guided weighting strategy is employed before feature fusion. For both backbones, a feature vector was obtained that was normalised and processed with a learned attention gate. This gate was a pluggable lightweight

MLP with a sigmoid activation function, and it produced an importance score for each feature branch. These coefficients, which expressed the importance of the feature of each model for the given input (i.e., in a local scope), were used to calculate the weighted sum of representations in Eq. 1.

$$F_{\text{fused}} = \sum_{i=1}^{3} \alpha_i \times F_i \quad [1]$$

$$\text{where } \alpha_i = \text{sigmoid}(W_i F_i + b_i) \text{ and } \sum_i \alpha_i = 1$$

Such adaptive weighting enables the model to highlight CNN features for texture-rich regions as well as ViT features for global structural context and leads to dynamic and feature fusion. Architecture allows the model to dynamically focus on texture-based local information extracted with CNNs, while preserving richer and more semantic contextuality represented by ViT, leading to improvements in both representation space richness and classification performance.

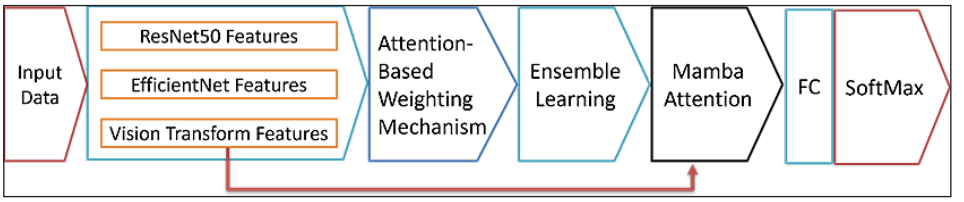


Figure 5. Proposed CNN_ViT_MambaATT framework

MAMBA Attention

We propose MambaATT, an attention mechanism specifically designed to improve the discriminative capacity of combined features. At the expense of creating additional parameters, MambaATT weighs different feature map parts in the pattern that corresponds to their relevance for a classification purpose. The method uses enhancing features that reinforce certain characteristics while suppressing irrelevant details or redundancy. This attention helps the model focusses on the most discriminative regions between benign and malignant tumours, especially for cases with subtle visual clues. MambaATT is intended to work in a distributed manner across the network to decrease the number of parameters of the network while maintaining high precision. That is, since the network can attend to whatever it needs to, it does not require the parameters that would be necessary if all the inputs were being processed. The experimental results show that the proposed extraction strategy based on the ensemble deep learning model and MambaATT modules can

effectively improve classification accuracy and robustness. This method offers a practical and potentially effective downsampling in CADs of dermatology. Figure 6 illustrates the structure of Mamba attention (MambaATT). MambaATT takes the stacked feature tensor and proceeds to process the features, producing an attention-weighted representation. To improve the stability and accelerate the training process, layer normalisation is used to normalise the input among features. A dropout layer, which deactivates a random subset of neurons during training to reduce overfitting, enhances model generalisation. Mamba for token mixing addresses issues of inductive bias and computational complexity. Finally, the attention-weighted features are flattened into a single vector and passed through two fully connected (FC) layers for final classification. Through the steps, the features obtained from MambaATT combine with the feature vectors acquired from the previous ViT model. Finally, these combined features are fed into the FC layer for the result analysis.

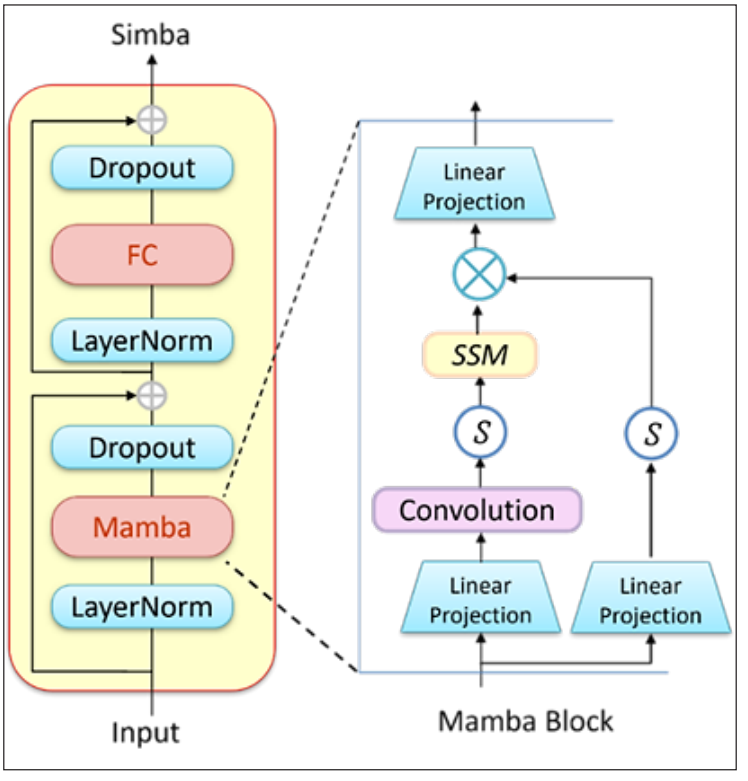


Figure 6. The architecture of MambaATT

Training and Evaluation

The accuracy and loss plots in Figure 7 show that the CNN_ViT_MambaATT model reaches a plateau after only a few epochs, with no sign of overfitting during training.

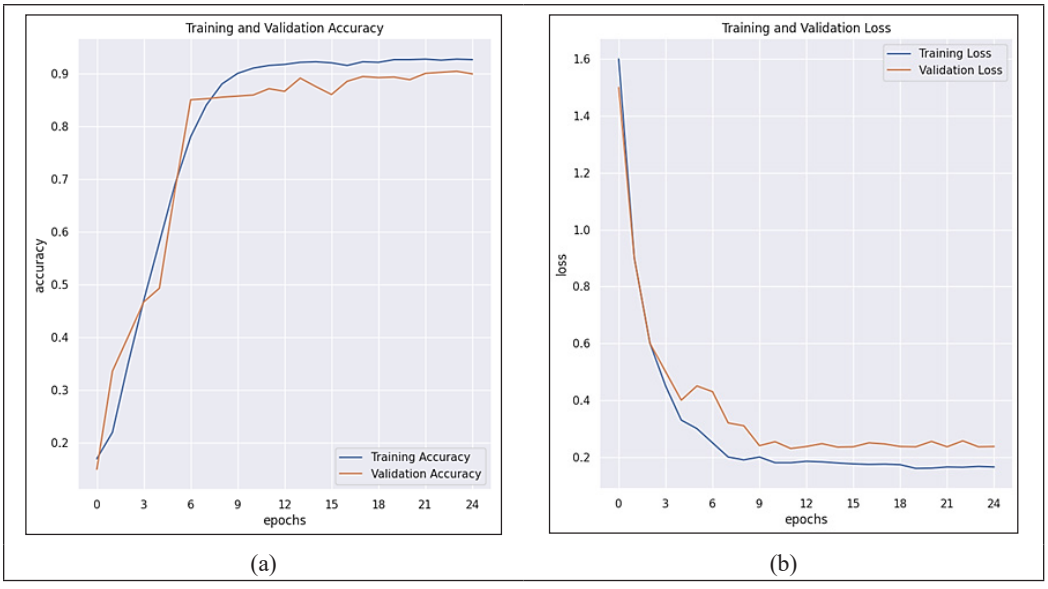


Figure 7. Training and validation (a) accuracy and (b) loss over epochs for CNN_ViT_MambaATT

We employed transfer learning techniques to leverage pre-trained ResNet50, EfficientNet, and ViT models trained on the ImageNet dataset. These models have already acquired valuable hierarchical features from the extensive and diverse ImageNet dataset. Utilising the knowledge acquired during their ImageNet training proves particularly advantageous when handling small datasets. This approach significantly reduces the training time and computational resources required for our research. The only modification that we made was in the last fully connected layer to have 2 output classes. These three models were used as base models for feature extraction. The combined extracted features were processed through "torch. stack," stacking tensors along a new dimension to form their complete feature representation. Then, the custom core module MambaATT was instantiated, and the features output by pre-trained models were concatenated as the input of MambaATT. In MambaATT, features were additionally extracted by its attention mechanism and multi-layer perception to promote the accuracy of image classification.

To assess the power of the feature set learned by CNN_ViT_MambaATT for skin cancer dermoscopic image classification, we compared it in three network architectures. Experiments reused the same datasets and widely used model evaluation metrics to ensure comparability of outcomes. The training employed a scheduling learning rate, starting from 0.001. The learning rate was divided by 10 if the model did not converge within two epochs. The training configuration for all the experiments is shown in Table 2. The MambaATT module was used with a state dimension of 64 an expansion factor was set to 2 and the size of convolution kernel was 3. This process was followed by a dropout rate of 0.1 to alleviate overfitting and the GELU activation function for improved nonlinearity.

Then two MambaATT layers were cascaded to refine the fused features from ResNet50, EfficientNet and ViT. These hyperparameters were determined empirically based on validation performance to balance model complexity and classification accuracy. The evaluation equations are given by:

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad [3]$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad [4]$$

$$\text{F1/Dice} = \frac{2\text{TP}}{2\text{TP+FP+FN}} \quad [5]$$

$$\text{True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP+FN}} \quad [6]$$

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP+FN}} \quad [7]$$

Table 2
Training parameters for each experiment

Parameter Name	Value
Input image dimension	224×224×3
Batch size	64
Learning rate	0.001
Training Epochs	25
Loss function	Cross-Entropy Loss
Optimiser	Adam
State dimension (MambaATT)	64
Expansion factor (MambaATT)	2
Convolution kernel size (MambaATT)	3
Dropout rate (MambaATT)	0.1
Number of MambaATT layers	2

RESULTS AND DISCUSSION

The results demonstrate a notable performance improvement, with the CNN_ViT_MambaATT model achieving an accuracy of 0.92. This signifies improvement over

earlier deep learning research Bechelli and Delhommelle (2022) using the same ISIC dataset, which reported an accuracy of 0.87. The hybrid model improves skin cancer classification accuracy, aligning with new findings that highlight the value of intelligent fusion frameworks in skin lesion analysis (Dorathi Jayaseeli et al., 2025). Pon Selchiya et al. (2023) and Waheed et al. (2023) two models are proposed with 90% and 91% accuracy, respectively. The state-of-the-art (SOTA) model's benchmarks have been compared with our proposed model to show the accuracy of CNN-Vit_MambaATT as displayed in Table 3.

A confusion matrix is included to visualise classification outcomes. The model correctly identifies 589 benign and 603 malignant samples, with minor confusion between visually similar cases (Figure 8). CNN_ViT_MambaATT achieves competitive performance 92% while maintaining higher efficiency and lower parameter complexity. Misclassified cases mainly resulted from image blur or artifacts. Future work will incorporate hair removal and colour normalisation to improve results further. To ensure a comprehensive evaluation, additional metrics were introduced: specificity and a combined receiver operating characteristic (ROC) with area under the ROC curve show that CNN_ViT_MambaATT achieves an AUC of 0.945, as illustrated in Figure 9, which confirms its superior discrimination.

Table 3
Comparison proposed with SOTA methods

Model	Accuracy	Precision	Recall	F1-Score
(Huang et al., 2024)	0.8780	0.8760	0.8790	0.8770
(Naz et al., 2025)	0.8652	0.7645	0.9069	0.8293
(Lea, 2025)	0.9125	**0.9127**	0.9125	0.9125
Proposed (CNN_ViT_MambaATT)	**0.9200**	0.9100	**0.9150**	**0.9180**

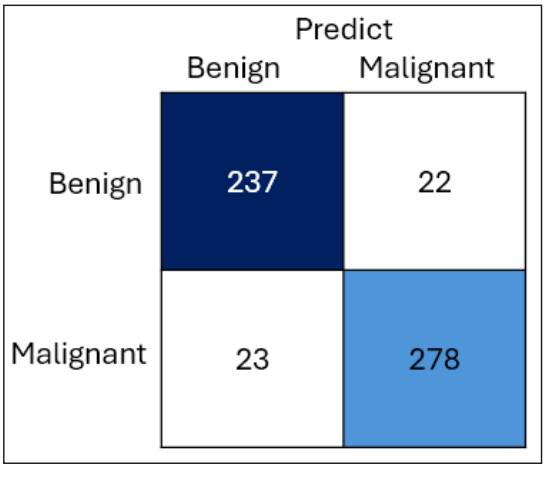


Figure 8. Confused matrix for testing the benign and malignant of dataset

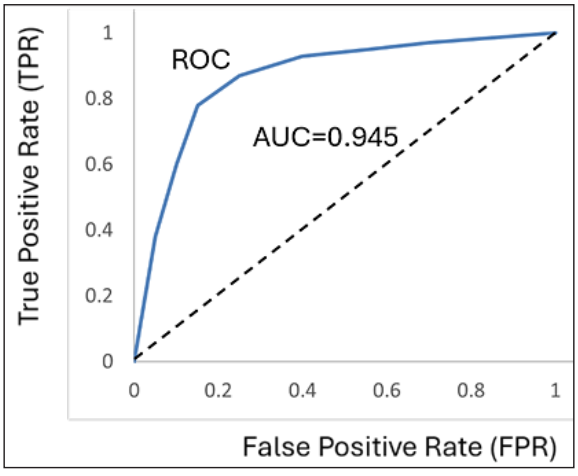


Figure 9. Area under the ROC curve of the proposed model

Ablation Experiment

The ablation experiment is a technique employed to assess the influence of specific components or features on model performance by systematically eliminating them from the model. It is typically applied to investigate the rationality of model design. The results of the ablation experiment for the proposed model are shown in Table 4.

The model achieved an accuracy of 92%. Embedding the MambaATT model resulted in a 1.4% accuracy to show that it improved classification precision through CNN-based presentation of image which counteracted the lack shortcoming from previous work. With the ViT model included, accuracy improved by 2.4%, demonstrating that by introducing global contextual information into models, ViT could improve the accuracy of classification task. Data augmentation was shown to provide a 1.6% improvement on the model, where we were able to generate more training examples through rotation, flipping, cropping and

Table 4
Results of the ablation experiment for the proposed model

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	0.8610	0.8600	0.8580	0.8500
EfficientNet	0.8400	0.8430	0.8430	0.8380
ViT	0.8750	0.8740	0.8740	0.8740
Resnet50 + EfficientNet	0.8660	0.8600	0.8620	0.8630
Resnet50 + EfficientNet + ViT	0.8900	0.8850	0.8890	0.8800
Resnet50 + EfficientNet + ViT + MambaATT	0.9040	0.9010	0.8980	0.8950
(Resnet50 + EfficientNet + ViT + MambaATT + Data Augmentation)	**0.9200**	**0.9140**	**0.9180**	**0.9150**

so forth. The additional training examples helped the model to better learn the distribution and properties of the data leading to an improved performance. The performance measures for the various model configurations are outlined in Figure 10. To justify the effectiveness of data augmentation techniques, we evaluated the pre- and post-augmented results of ResNet50, ViT, and CNN_ViT_MambaATT models. As can be seen from Table 5, data augmentation significantly improves the model performance. This further reinforces the idea that deep learning models typically require large datasets to be trained, and small datasets may not sufficiently utilise their potential.

The research questions defined in the Introduction were effectively addressed. The CNN_ViT_MambaATT model showed better results than pure CNN or ViT networks. The module MambaATT improved the discriminative power by discovering the useful contextual transitions. The proposed feature fusion rule strengthened the robustness

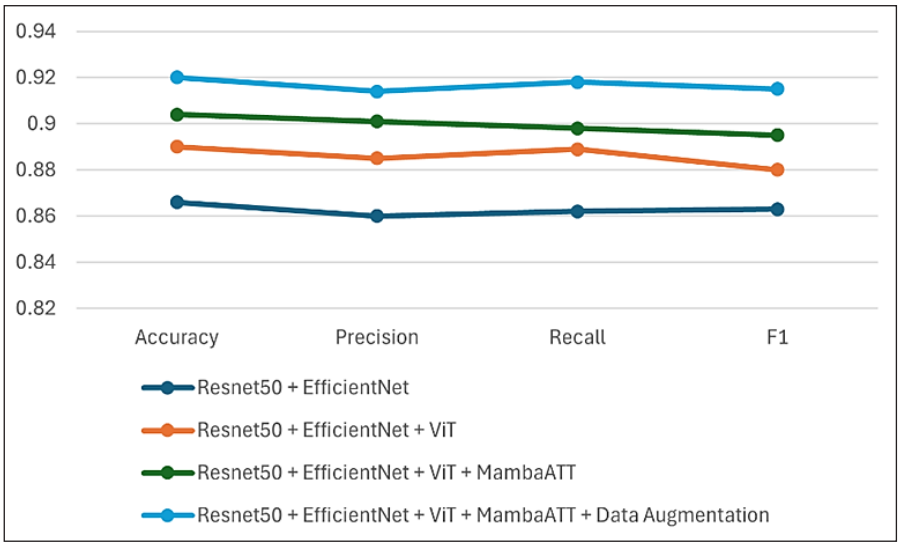


Figure 10. Performance metrics by model configurations

Table 5
Accuracy comparison of models before and after applying data augmentation

Model	Accuracy	Precision	Recall	F1-Score
Resnet50	0.8500	0.8450	0.8480	0.8490
(Resnet50+ Data Augmentation)	0.8610	0.8600	0.8640	0.8610
ViT	0.8600	0.8520	0.8560	0.8520
(ViT+ Data Augmentation)	0.8750	0.8680	0.8720	0.8680
CNN_ViT_MambaATT	0.9000	0.8910	0.8960	0.8920
(CNN_ViT_MambaATT +Data Augmentation)	**0.9200**	**0.9140**	**0.9180**	**0.9150**

of less data using the complementary spatial and global cues. This framework provides a basis for incorporating state-space models into medical image pipelines.

Misclassified samples were reviewed to search for common causes of the errors, such as low contrast of a lesion, overlapping artifacts, and visual similarities between early-stage melanoma and benign lesions. The Grad-CAM visualisation was used for interpreting the model's decision-making process, and it is evident that the proposed hybrid CNN–ViT–MambaATT architecture primarily emphasises lesion boundaries. In some cases, however, the activation of attention was also in harmless neighbouring regions, which might suggest over-sensitivity to background texture. Overall, the AUC of 0.945 and classification accuracy of 92% demonstrated a strong diagnostic performance of the model. Still, these perfusion-based measures should be used carefully in clinical practice. False negatives could result in missed melanomas, while false positives would warrant unnecessary surgical excisions. Therefore, this study frames the proposed model as a triage support tool that aids but does not substitute expert clinical judgment of dermatologists. In the future, we will work toward collaborating with dermatologists for validation and incorporating the model into clinical databases for evaluation of real-world reliability and safety.

CONCLUSION

In this study, we propose a novel model, CNN_ViT_MambaATT for skin cancers classification by utilising advanced DNN architectures such as ResNet50, VIT and MambaATT. This study also analysed the binary categorisation between benign and malignant dermoscopic skin cancer images. The skin cancer classification model CNN_VIT_MAMBAATT was trained and tested with image data from the ISIC dataset. The proposed method showed significant enhancements in malignant skin lesions and had 92% accuracy on the ISIC dataset. Conclusion: Classification focuses on how to couple learning mechanisms with the new attention mechanisms for improved classification accuracy and robustness. Furthermore, using the data-enlargement technique helped to overcome issues associated with small sample size and, in essence allowed the model to 'normalise' the data. We think that these results provide valuable understanding of the potential of hybrid models for medical image classification and encourage further developments towards reliable clinical systems.

LIMITATIONS AND FUTURE WORKS

Even though advances have been made, tasks such as denoising and efficiency are still necessary to balance the dermoscopic dataset. A potential solution incorporates a metaheuristic algorithm for deep renovation, optimal hyperparameter configuration advanced preprocessing methods to deal with these issues.

Hybrid models will particularly be a major direction in future studies for better explanation by visualisation tools, that is, Grade-Cam (Gamage et al., 2024), and for acknowledgements of the real world to make clinical relevance. VIT enriched with our architecture has strong prospects in skin cancer classification. In conjunction, hybrid methods to break pointing strategies and excellent preterm refined transition mechanisms that address current limitations, make the path towards more effective CAD systems, a clinically acceptable data control diagnostic (CAD) system. Future work should also focus on the slow processing speed of the Mamba Vit model. Though skin cancer detection can greatly benefit from high accuracy, its computation is slow and hence not within the reach of a wider population. Advancements in such models will have to entail architectural changes, simplified model structure, and faster training and estimation procedures. Furthermore, using the hardware acceleration of GPUs or TPUs increases performance.

The next study model is designed to be a compromise between computational efficiency and accuracy, as well as a practical, scalable clinical device. The objective will be to develop a hybrid system that combines the best attributes of Mamba and the VIT architecture, together with a model that is well-established for high-speed computations. In future work, we expect to validate the interpretability using clinically trained Dermatologists on Grad-CAM (Gamage et al., 2024) visualisations. train the proposed model on a large dataset, e.g., validate it on the ISIC 2019/2020 datasets. There will be some areas of future work integrating lesion-specific preprocessing hair removal, colour normalisation, testing on ISIC 2019 and ISIC 2020 benchmarks for wider generalisation and optimising computation cost for real clinical use.

ACKNOWLEDGEMENT

The authors wish to acknowledge the Faculty of Computer Science and Information Technology at UPM for the support and funding provided in this work.

REFERENCES

Abbas, Q., Daadaa, Y., Rashid, U., & Ibrahim, M. E. A. (2023). Assist-Dermo: A lightweight separable vision transformer model for multiclass skin lesion classification. *Diagnostics, 13*(15), Article 2531. https://doi.org/10.3390/diagnostics13152531

Ali, A., Shahbaz, H., & Damaševičius, R. (2025). xCViT: Improved vision transformer network with fusion of CNN and Xception for skin disease recognition with explainable AI. *Computers, Materials & Continua, 83*(1), 1367-1398. https://doi.org/10.32604/cmc.2025.059301

Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets-a first step towards preventing skin cancer. *Neuroscience Informatics, 2*(4), Article 100034. https://doi.org/10.1016/j.neuri.2021.100034

Arshed, M. A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M. (2023). Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models. *Information, 14*(7), Article 415. https://doi.org/10.3390/info14070415

Baykal Kablan, E., & Ayas, S. (2024). Skin lesion classification from dermoscopy images using ensemble learning of ConvNeXt models. *Signal, Image and Video Processing, 18*(8-9), 6353-6361. https://doi.org/10.1007/s11760-024-03321-y

Bechelli, S., & Delhommelle, J. (2022). Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. *Bioengineering, 9*(3), Article 97. https://doi.org/10.3390/bioengineering9030097

Chen, M., Radford, A., Child, R., Wu, J., Jun, H., Luan, D., & Sutskever, I. (2020). Generative pretraining from pixels. In H. Daumé III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 1691-1703). PMLR. https://proceedings.mlr.press/v119/chen20s.html

Dorathi Jayaseeli, J. D., Briskilal, J., Fancy, C., Vaitheeshwaran, V., Patibandla, R. S. M. L., Syed, K., & Swain, A. K. (2025). An intelligent framework for skin cancer detection and classification using fusion of squeeze-excitation-densenet with metaheuristic-driven ensemble deep learning models. *Scientific Reports, 15*(1), Article 92293. https://doi.org/10.1038/s41598-025-92293-1

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929

Flosdorf, C., Engelker, J., Keller, I., & Mohr, N. (2024). Skin cancer detection utilising deep learning: Classification of skin lesion images using a vision transformer. *arXiv*. https://doi.org/10.48550/arXiv.2407.18554

Gamage, L., Isuranga, U., Meedeniya, D., De Silva, S., & Yogarajah, P. (2024). Melanoma skin cancer identification with explainability utilising mask guided technique. *Electronics, 13*(4), Article 680. https://doi.org/10.3390/electronics13040680

Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine, 127*, Article 104065. https://doi.org/10.1016/j.compbiomed.2020.104065

Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modelling with selective state spaces. *arXiv*. https://doi.org/10.48550/arXiv.2312.00752

Han, S. S., Moon, I. J., Lim, W., Suh, I. S., Lee, S. Y., Na, J.-I., Kim, S. H., & Chang, S. E. (2020). Keratinocytic skin cancer detection on the face using region-based convolutional neural network. *JAMA Dermatology, 156*(1), 29-37. https://doi.org/10.1001/jamadermatol.2019.3807

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Huang, H., Hsu, B. W., Lee, C., & Tseng, V. S. (2021). Development of a lightweight deep learning model for cloud applications and remote diagnosis of skin cancers. *The Journal of Dermatology, 48*(3), 310-316. https://doi.org/10.1111/1346-8138.15683

Huang, S., Lei, H., Yang, J., Hang, T., Jin, L., Yao, Y., & Li, C. (2024). An attention mechanism and ensemble learning based on dermoscopic image classification. *Journal of Imaging Science and Technology, 68*(4), Article 40401. https://doi.org/10.2352/J.ImagingSci.Technol.2024.68.4.040403

Khan, M. A., Zhang, Y.-D., Sharif, M., & Akram, T. (2021). Pixels to classes: Intelligent learning framework for multiclass skin lesion localisation and classification. *Computers & Electrical Engineering, 90*, Article 106956. https://doi.org/10.1016/j.compeleceng.2020.106956

Kittler, H., Pehamberger, H., Wolff, K., & Binder, M. (2002). Diagnostic accuracy of dermoscopy. *The Lancet Oncology, 3*(3), 159-165. https://doi.org/10.1016/S1470-2045(02)00679-4

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097-1105.

Lea, S. (2025). Malignant vs. benign skin cancer classification using topological data analysis and machine learning. In *2025 IEEE 9th Forum on Research and Technologies for Society and Industry (RTSI)* (pp. 102-107). IEEE. https://doi.org/10.1109/RTSI64020.2025.11212537

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 255-258). MIT Press.

Meedeniya, D., De Silva, S., Gamage, L., & Isuranga, U. (2024). Skin cancer identification utilising deep learning: A survey. *IET Image Processing*. Advance online publication. https://doi.org/10.1049/ipr2.13219

Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)* (pp. 117-122). IEEE. https://doi.org/10.1109/IIPHDW.2018.8388338

Mobiny, A., Singh, A., & Van Nguyen, H. (2019). Risk-aware machine learning classifier for skin lesion diagnosis. *Journal of Clinical Medicine, 8*(8), Article 1241. https://doi.org/10.3390/jcm8081241

Naz, N. S., Mehmood, M. H., Ahmed, F., Ahmad, M., Rehman, A. U., Ismael, W. M., & Adnan, K. M. (2025). Privacy preserving skin cancer diagnosis through federated deep learning and explainable AI.

Nie, Y., Sommella, P., Carratù, M., O'Nils, M., & Lundgren, J. (2023). A deep CNN-transformer hybrid model for skin lesion classification of dermoscopic images using focal loss. *Diagnostics, 13*(1), Article 72. https://doi.org/10.3390/diagnostics13010072

Ozdemir, B., & Pacal, I. (2025). A robust deep learning framework for multiclass skin cancer classification. *Scientific Reports, 15*(1), Article 4938. https://doi.org/10.1038/s41598-025-89230-7

Pon Selchiya, R., Manimegalai, P., & Thomas George, S. (2023). Comparative analysis-based melanoma detection in dermoscopic images with deep learning techniques. In *2023 5th International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICECCT56650.2023.10179820

Reis, H. C., & Turk, V. (2024). Fusion of transformer attention and CNN features for skin cancer detection. *Applied Soft Computing, 164*, Article 112013. https://doi.org/10.1016/j.asoc.2024.112013

Siegel, R. L., Kratzer, T. B., Giaquinto, A. N., Sung, H., & Jemal, A. (2025). Cancer statistics, 2025. *CA: A Cancer Journal for Clinicians, 75*(1), 10-45. https://doi.org/10.3322/caac.21871

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (pp. 6105-6114). PMLR.

Tay, Y., Dehghani, M., Gupta, J., Bahri, D., Aribandi, V., Qin, Z., & Metzler, D. (2021). Are pre-trained convolutions better than pre-trained transformers? *arXiv*. https://doi.org/10.48550/arXiv.2105.03322

Waheed, S. R., Saadi, S. M., Rahim, M. S. M., Suaib, N. M., Najjar, F. H., Adnan, M. M., & Salim, A. A. (2023). Melanoma skin cancer classification based on CNN deep learning algorithms. *Malaysian Journal of Fundamental and Applied Sciences, 19*(3), 299-305. https://doi.org/10.11113/mjfas.v19n3.2900

Yang, G., Luo, S., & Greer, P. (2025). Boosting skin cancer classification: A multi-scale attention and ensemble approach with vision transformers. *Sensors, 25*(8), Article 2479. https://doi.org/10.3390/s25082479

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). Vision Mamba: Efficient visual representation learning with bidirectional state space model. *Proceedings of Machine Learning Research, 235*, 62429-62442. https://proceedings.mlr.press/v235/zhu24f.html.

Pertanika J. Sci. & Technol. 34 (2): 647 - 670 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Design, Manufacturing, and Testing a Versatile Test Stand for MP100S Integration

Huu Danh Tran

Department of Mechanical Engineering, Vinh Long University of Technology Education, 73 Nguyen Hue Street, Vinh Long City, 85100 Vinh Long Province, Vietnam

ABSTRACT

The article designs and manufactures an engine mounting test stand with a structure that can not only mount a single test engine but also mount many different types of engines with an engine support structure that can move along the X, Y, Z axes and rotate around the Z axis, making it very flexible and easy to mount and align the engine connected to the MP100S dynamometer system. Manufacturing and assembling a complete engine mounting test stand that can be linked to the MP100S dynamometer system ensures technical requirements, safe operation, and testing to measure the number of revolutions n (rpm), torque Mx (N·m), and power P (kW) of the engine when operating at no load and with load. The results of the paper can be used as reference data in calculating the design and manufacturing of the engine mounting test stand to measure the number of revolutions, torque and power of the engine, Turbocharger blade efficiency, exhaust gas pressure, fuel consumption corresponding to each engine operating mode.

Keywords: Engine, load, MP100S, open power flow, test stand, turbocharger

INTRODUCTION

Diesel engines are the most economical and durable engines in the world, have the longest history and are widely used in machinery and industrial equipment such as automobiles, excavators, forklifts, and construction machinery. There are many research projects in the world on diesel engines to improve engine performance, reduce emissions, reduce noise, and reduce vibration (Calvo et al., 2006; Chen et al., 2022; Huang et al., 2011; Lloyd & Cackette, 2001; Ni et al., 2020; Norouzi et al., 2022; Visan et al., 2022). Research on turbocharger improvements to reduce turbo lag, emissions, and fuel

ARTICLE INFO

Article history:
Received: 15 October 2025
Accepted: 29 December 2025
Published: 01 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.03

E-mail address:
danhth@vlute.edu.vn

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

consumption aims to optimise the turbocharger (Chen et al., 2022; Visan et al., 2022), use advanced algorithms through deep learning to control the engine to reduce NO_x and fuel consumption, keep the load, which reduces the calculation speed (Norouzi et al., 2022), control the engine using advanced algorithms (Norouzi et al., 2022), and reduce environmental impact through biodiesel fuel (Visan et al., 2022) or emission control (Lloyd & Cackette, 2001). From the above research works with those improvements, it is required that the engine mounting test stands must be modern, capable of connecting to the MP100S dynamometer system to collect, and process accurate data, to have a general overview of the research and manufacture of a multi-purpose test stand that can adjust the mounting of many different types of engines, diverse testing requirements, not only testing the main engine performance but also having the ability to flexibly integrate with sensors and MP100S measuring devices. In addition to research works on diesel engines, there are also research works on other types of engines that have been studied such as hybrid engines (Crispi et al., 2022), ammonia-hydrogen fuel engines (Qi et al., 2023), CNG-SI engines (Tomin et al., 2024), and customised internal combustion engines (Altarazi et al., 2020; Yu et al., 2021) to meet the orientation of green and clean energy development in the current period. Combined with the research on auxiliary details such as mufflers in the air intake system (Liu et al., 2019) and exhaust gas thermoelectric generators (Subramaniam et al., 2019) which also greatly affect engine performance, the research, design and manufacture of a test stand for mounting an engine, especially a diesel engine, is an important field with studies focussing on testing and evaluation (Adkine et al., 2017; Alfaiz et al., 2014; Derbiszewski et al., 2023; Izvorean & Stoica, 2022; Lee et al., 1998; Mohammadi et al., 2024; Oliveira et al., 2024; Park, 2023; Santhosh Kumar et al., 2022; Shao, 2011; Singh, 2000; Tatarynow et al., 2022; Wang et al., 2021). These studies use finite element analysis (FEA) to analyse the surface and vibration to optimise the structural design or analyse the stress/load of the engine test stand (Adkine et al., 2017; Mohammadi et al., 2024; Santhosh Kumar et al., 2022).

The studies provide low-cost, efficient options for performing engine performance analysis such as designing a small engine dynamometer using a hydraulic system with a pressure regulator and pressure gauge, measuring power and torque to develop low-cost engine test stands (Alfaiz et al., 2014) and also serve educational or vocational training purposes (Oliveira et al., 2022), or testing with different fuels (Izvorean & Stoica, 2022; Tatarynow et al., 2022).

Through an overview of research projects in the world, it is found that in order to meet the general development trend of the times, especially the localisation of auto parts, mastering the technological capabilities to move towards self-designing and manufacturing diesel engines in Vietnam, it is necessary to have many research projects on engines such as reducing energy consumption, reducing exhaust toxicity, ensuring smooth and unified

operation of the equipment without causing noise, improving engine efficiency, especially increasing the Turbocharger efficiency of the engine, to serve the above research well, it is necessary to have experiments in the laboratory similar to the actual operating conditions of the engine. The design of the engine mounting test stand in the experiment is like the engine mounted on the vehicle, the test stand is designed and manufactured with the engine mounted horizontally, like the engine used on front-wheel drive vehicles and some 4-wheel drive vehicles. The test stand is designed to mount a variety of engines and is easily connected to other auxiliary structures. It has a structure that can not only mount a single test engine but also mount many different types of engines with a engine support structure that can move forward along the X, Y, Z axes and rotate around the Z axis, so it is very flexible and easy to mount and align the engine connected to the MP100S system to perform different testing objectives. The engine mount, at the same time, the frame structure must ensure technical requirements for manufacturing and assembly, ensuring durability and anti-vibration ability under the conditions of mounting the engine, operating and testing without load and with load.

Although various engine test stands have been proposed in previous studies, most of them are limited to a fixed configuration and cannot be adjusted in multiple degrees of freedom. None of the published designs provide direct compatibility with the MP100S dynamometer system, nor do they combine structural design, FEA validation, vibration analysis, fabrication, and experimental evaluation in a unified workflow. These limitations form the research gap addressed in this study.

This study focusses on the design, structural evaluation, and experimental implementation of an engine mounting test stand integrated with the MP100S dynamometer system, with the primary aim of providing a practical and flexible laboratory solution rather than a comprehensive performance comparison with commercial engine test benches.

MATERIALS AND METHODS

Approach

Overview of diesel engines, MP100S dynamometer system, related domestic and foreign research projects. Proposing the schematic diagramme of the test stand, analysing and determining a reasonable manufacturing plan, building a 3D model of the test stand, applying finite element analysis (FEA) to evaluate the durability and vibration characteristics, establishing the technological process of processing the parts, selecting materials, processing methods and reasonable processing technological parameters for processing the parts. Manufacture and assemble the complete test stand by connecting the test stand with the MP100S system.

Research Method

Theoretical research method: using the method of collecting documents at the same time as the method of synthesising and processing documents, studying an overview of related domestic and foreign research works, proposing a schematic diagramme of the test stand, analysing and determining a reasonable manufacturing plan, building a 3D model of the test stand, applying the finite element analysis (FEA) to optimise the design.

Experimental Research Method

Manufacturing and assembling a complete test stand connecting multiple engines with the MP100S system, determining the research objective function, building an experimental diagramme, determining a set of experimental parameters, conducting experiments to determine the values corresponding to the experimental modes. Based on the results achieved, comparing them with the measured values from the experiment, it is possible to determine several factors affecting the objective function.

Design and Manufacturing Plan of the Test Stand

From the overview of research topics, it can be seen that there are a number of research projects related to engine test stands, most of the test stands are researched to serve the purpose of testing and applying the achieved results to experimentally verify the results of the achieved goals, however, these projects are all theoretical studies, so the task of the article is to fully research the issues of: test stand, MP100S system and the achieved linkage results to meet the technical requirements and ensure good practical application of the test stand. The test stand is designed according to the open power flow principle: based on the theoretical research on open power flow, the proposed operating principle of the test stand must be flexible to mount many different types of engines, applying design software to draw the principle diagramme and structure of the test stand, designing and assembling 3D models, simulating durability testing and analysing the vibration of the test stand, manufacturing, assembling the complete test stand, checking and testing.

Analysis and Selection of Reasonable Manufacturing Options

The design idea is to design the structural diagramme of the engine test stand from the schematic diagramme of the manufacturing options, analyse the advantages and disadvantages of each option to select a reasonable option, then apply Autodesk inventor software to design 3D, assemble the complete test stand, conduct simulation and durability analysis in Ansys, manufacture, install and test

Select Materials and Manufacturing Methods of the Test Stand

The available equipment and main components of the test stand were identified and selected, including the engine, MP100S dynamometer system, cardan coupling, cooling tank, and fuel tank. The engine support frame works under heavy load conditions, so it is necessary to choose the connecting parts as square steel tubing 75 × 75 × 3 mm made of ASTM A36 steel with the following parameters: Tensile strength: 400-550 MPa; Elongation ≥ 20-23%; Hardness: 119-162 HB; Elastic modulus: 200 GPa; Poisson's ratio: 0.26. According to ASTM A36 standard, the chemical composition of steel usually includes the following elements (percentage by mass) shown in Table 1. The test stand consists of many different parts connected, so the main manufacturing method chosen is welding to connect the parts, combined with bolts and nuts.

Table 1
Chemical composition (percentage) of ASTM A36 standard steel

Steel Grade	C	Si	Mn	P	S	Cu
A36 Steel	0.16	0.22	0.49	0.16	0.08	0.01

Note. C = Carbon; Si = Silicon; Mn = Manganese; P = Phosphorus; S = Sulfur; Cu = Copper

The test stand consists of many different parts connected, so the main manufacturing method chosen is welding to connect the parts, combined with bolts and nuts.

RESULTS AND DISCUSSION

Schematic Diagramme of Engine Mounting Test Stand

Figure 1 shows the schematic layout of the engine mounting test stand. The system consists of the engine (1), the MP100S dynamometer system (2), the cardan coupling (3), the engine support frame (4), the cooling tank (5), the fuel tank (6), and the brake controller (7).

The multi-engine mounting test stand is designed, manufactured and tested according to the open power flow principle, characterised by the input and output power not being connected in a closed loop with the main components including the engine (1), the MP100S unit (2) and the cardan coupling (3). In which, the engine is used for driving, the cardan coupling connects the engine to the MP100S unit, the support frame (4) is used to support the engine, the cooling tank (5) is used to cool the engine, the fuel tank (6) supplies fuel when the engine is running, MP100S dynamometer system manufactured by Weinlich of Germany includes the following main components: the control unit includes the MP computer inside and the control panel; the brake and measuring unit is directly connected to the engine; the brake controller (7) to change the braking force to create load and torque acting on the engine and the support frame to perform the test conditions and requirements.

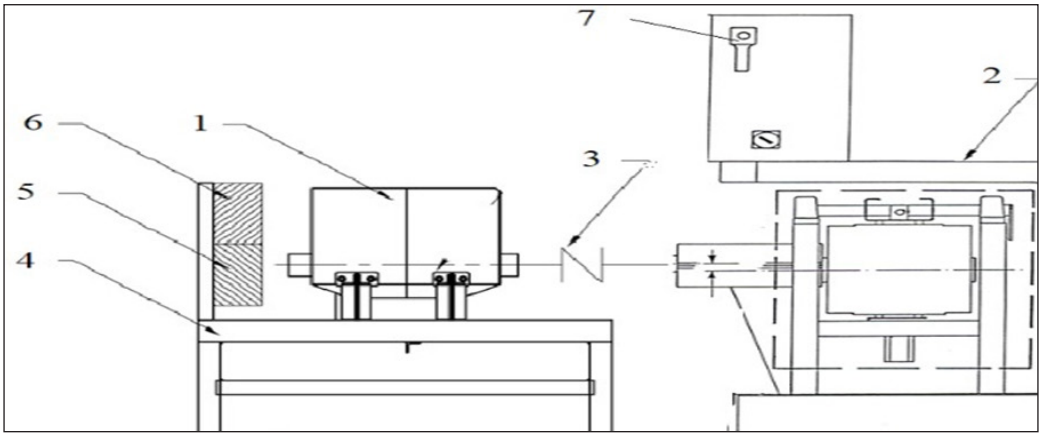


Figure 1. The schematic layout of the engine mounting test stand

Design of the Engine Mounting Test Stand

Figure 2 illustrates the four design alternatives considered for the engine mounting test stand (Alternatives 1-4). Design structural diagrammes of manufacturing options, analyse advantages and disadvantages of each option to choose the appropriate option.

Based on the above design concepts, the four alternatives were compared to identify the most suitable configuration, as shown in Table 2.

Table 2
Comparison of the four design alternatives

Criterion	Alternative 1	Alternative 2	Alternative 3	Alternative 4 (Selected)
Structural rigidity	High	High	Medium	High
Adjustability in X–Y–Z	No	Yes	Yes	Yes
Rotation around Z-axis	No	Yes	Yes	Yes
Safety under load	Medium	Good	Low	Very good
Engine compatibility	Limited	Wide	Limited	Very wide
Ease of fabrication	Medium	Medium	Low	Easy
Shaft alignment accuracy	Poor	Good	Poor	Very good

Note. Alternative 4 offers the most balanced combination of rigidity, adjustability, safety, and compatibility

Alternative 1: Four-point Engine Mounting

The engine is mounted at 4 points at 4 mounting positions at 2 ends of the engine, in which each side is mounted at 2 symmetrical positions across the longitudinal cross-section of the engine. The engine mounting frame consists of a 4-bar support structure, the upper end is welded to the L-shaped engine mounting base to mount to the engine with bolts and nuts, the lower end is fixed to the main frame by welding. The main frame is designed

with a rectangular structure; the 4 legs of the main frame are connected to the base by 4 anti-vibration mounts to reduce shock and vibration. In this design, the advantage of the simple structure makes manufacturing easy, the engine is mounted at 4 points, so it has high rigidity. The engine is rigidly mounted to the support, the frame structure and the base must be designed to fit the engine, so it is only possible to permanently connect the engine to the structure with almost minimal deflection, so there is inconvenience in the displacement of the engine mounting system with the support and the ability to withstand tolerance when the test stand is put into operation. With this design, the engine installation also does not allow any adjustment in the vertical (X direction), horizontal (Y direction), and height (Z direction) when such adjustment is very necessary. Therefore, this connection theoretically makes the engine support frame, and the main frame form a unified block, so it only allows the engine to be connected to the MP100S system by moving the entire test stand, so it is very difficult to align the test stand after mounting to ensure the assembly requirements and safety during operation.

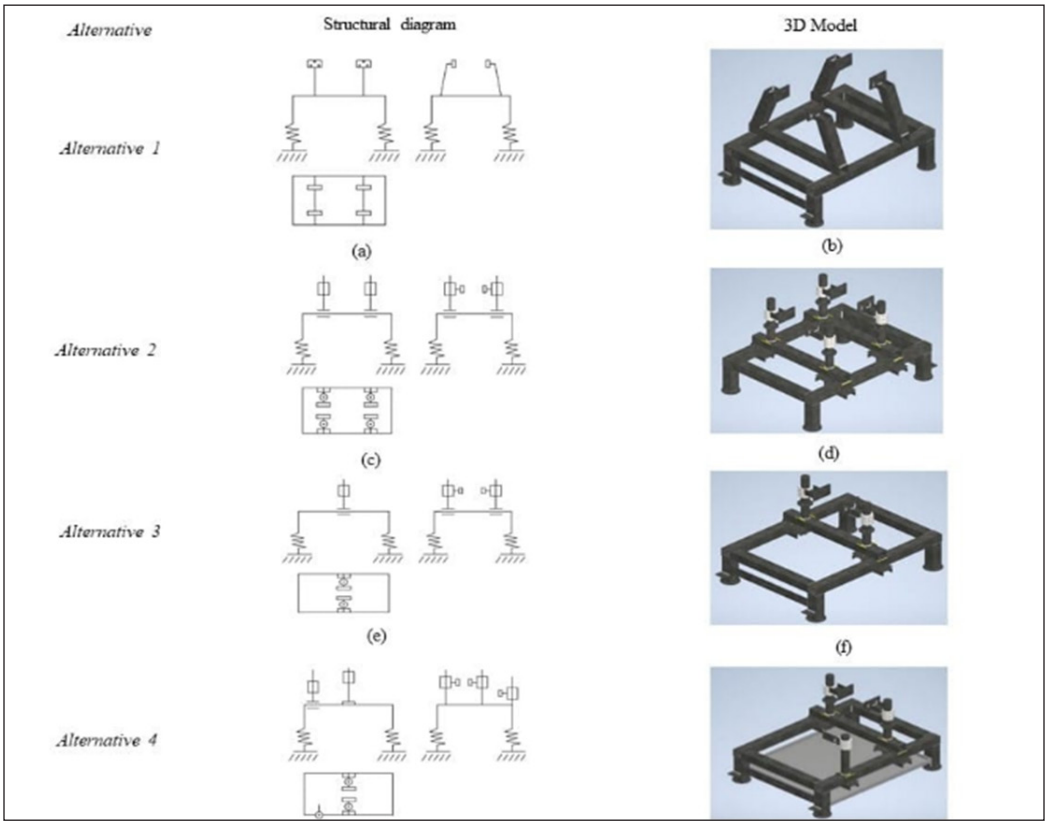


Figure 2. Design alternatives for the engine mounting test stand: (a) Structural diagramme of alternative 1; (b) 3D model of alternative 1; (c) Structural diagramme of alternative 2; (d) 3D model of alternative 2; (e) Structural diagramme of alternative 3; (f) 3D model of alternative 3; (g) Structural diagramme of alternative 4; (h) 3D model of alternative 4

Alternative 2: Adjustable Engine Mounting System

The main frame of the test stand is designed with a rectangular frame structure, the 4 legs of the main frame are connected to the ground by 4 anti-vibration mounts to reduce shock and vibration; The engine support consists of 2 rectangular cross-section bars placed on the upper surface of the main frame, this cross-section bar acts as a horizontal sliding support bar, the cross-section bar has the bottom surface at both ends welded to 02 U-shaped basses to move along the frame (X direction) on the main support bar, each horizontal sliding bar contains 2 sets of round shaft supports + shaft joints, thus 2 horizontal support bars contain 4 sets of round shaft supports + shaft joints, in which the round shaft supports + shaft joints to move up and down (Z direction) and rotate around the centre of the round shaft to the appropriate position are fixed with hexagon socket head locking screws, 04 shaft joints are welded to 04 L-shaped engine mounting basses to mount with the engine with bolts and nuts mounted at 4 points at 4 mounting positions at 2 ends of the engine, in which each side is mounted in 2 symmetrical positions across the longitudinal cross-section plane of the engine, the lower end of the round shaft support is 04 points at 04 mounting positions at 02 ends of the engine, in which each side is mounted in 2 symmetrical positions across the longitudinal cross-section plane of the engine, the lower end of the round shaft support is welded to 04 L-shaped engine mounting bases to mount with the engine. U-shaped bass to connect with the rectangular crossbar on the top of the main frame with bolts + nuts to perform linear movement in the horizontal direction (Y direction). In this design, the advantage is that the engine is mounted at 4 points so it has high rigidity, the engine is rigidly mounted to the engine support mechanism, in which the engine support mechanism can perform translational movement in 3 directions X, Y, Z and rotate around the Z axis, so with this design, the engine will be easily connected to the MP100S performance measuring device by adjusting the dynamic support mechanism vertically (X direction), horizontally (Y direction), and height (Z direction) to the appropriate position, then the fixed connection between the mechanisms and the main frame with bolts + nuts makes it very easy to adjust and mount, expanding the technological capabilities of the fixture, the fixture can mount a variety of different types of engines to ensure assembly requirements and safety during operation. Besides the above advantages, there are also some notes that the mounting structure is too large, so the working space and mounting are limited. The 4 rigid mounting points on the engine create a very high rigidity, so to make the 2 ends of the engine even, the mounting design must be precise. In case the engine rotates around the Y axis, it will be difficult to adjust. However, because the engine is connected to the MP100S by a cardan coupling, the above problem is not significant.

Alternative 3: Centralised Two-point Engine Mounting System

The test stand is designed in terms of main frame structure and engine support mechanism with the same moving principle as option 2, but the engine support mechanism has only

1 rectangular cross-section bar placed on the upper surface of the main frame containing 2 round shaft supports + shaft joints, 02 shaft joints are welded to 02 L-shaped engine mounting basses to mount to the engine with bolts and mounting nuts at 2 points at 2 mounting positions in the middle of the engine. This design has the advantage of overcoming the disadvantages of options 1 and 2. The engine is mounted at 2 points in the middle of the engine, so it bears the entire weight of the engine, making it not very rigid. The engine is connected to the MP100S performance measuring device via a cardan coupling, bearing the load of the joint and the moment when the engine is operating, especially in the case of a load test using the load-inducing and braking mechanism on the MP100S performance measuring device, so it does not ensure safety during operation.

Alternative 4: Three-point Engine Mounting System

The test stand is designed exactly the same as option 3 but has an additional round shaft support + shaft coupling with the lower end of the round shaft support welded with a U-shaped bass to connect to the main frame, the shaft coupling is welded to a fixed bass engine mount with an expansion gap to support the flywheel and the cardan coupling that drives the engine and the MP100S performance measuring device at the engine head. In this design, the engine is mounted at 3 points, including 2 main points in the middle of the engine, so it bears the entire weight of the engine and 1 mounting point at the moving connector of the engine, so it has high rigidity, ensuring safety when operating the test without load and with load on the MP100S performance measuring device. Through analysing the advantages and disadvantages of the 4 design options above, option 04 is the most reasonable, with a simple structure, easy to manufacture, easy to adjust and assemble. The test stand can be assembled with many different types of engines, so option 04 was chosen for design and manufacture.

Complete Assembly of Test Stand and Durability Test Simulation in Inventor Software

Design details and assemble the complete test stand in Autodesk Inventor 3D design software, the results are the scientific basis for the simulation process of analysing and testing the test stand's durability and vibration as shown in Figure 3. Stress analysis in Autodesk Inventor 2025 evaluated structural integrity under a 500 kg engine weight and 100 N·m torque as shown in Figure 4.

Boundary Conditions and Applied Loads

The frame was fixed at the four mounting legs. A distributed load equivalent to the 500 kg engine mass was applied to the mounting brackets, and a torque of 100 N·m was applied through the cardan coupling to represent typical operating conditions.

A tetrahedral mesh with refinement in high-stress regions was used. The maximum von-Mises stress obtained (91.79 MPa) remained far below the 250 MPa yield strength of A36 steel.

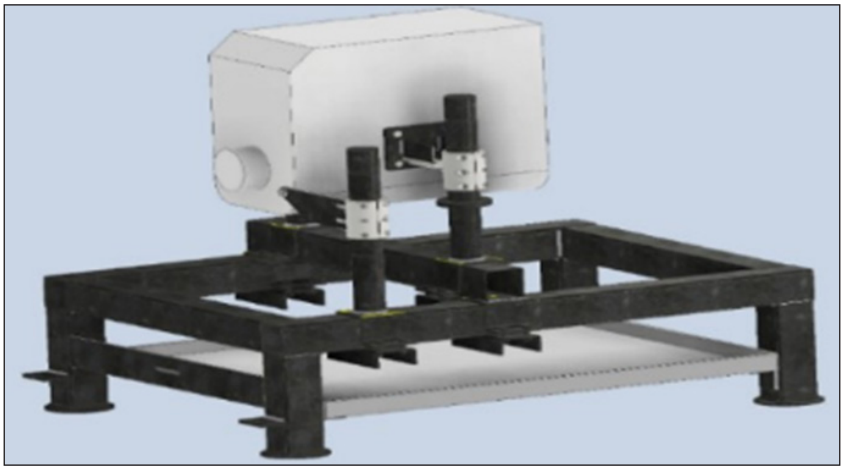


Figure 3. Complete assembly of the design test stand

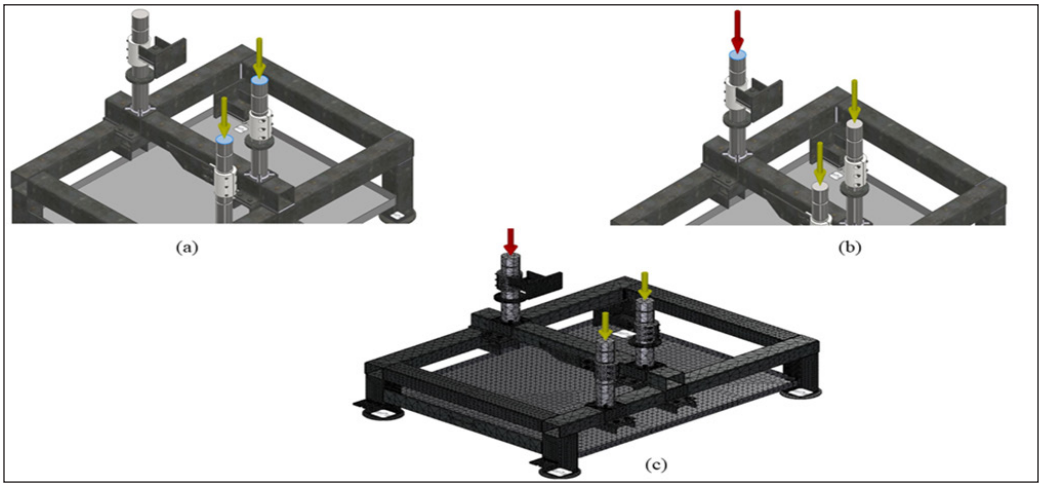


Figure 4. Forces and mesh for stress analysis: (a) F1, F2 (yellow); (b) F3 (red); (c) meshing

Mesh Sensitivity

Several mesh densities were tested. When the element size was refined from 8 mm to 4 mm, the variation in maximum stress decreased from 4.2% to 2.7%, confirming mesh

convergence.

Torque Selection

The applied torque of 100 N·m corresponds to the typical mid-load operating range of the Toyota 3C engine as documented in the thesis, ensuring that the simulated loading condition is realistic.This loading condition was selected as a conservative representation of typical mid-load operation to ensure structural safety under normal use as well as slightly unfavourable working conditions.

Material Selection

A36 steel was chosen for its good weldability, wide availability, and adequate yield strength for medium-duty structural applications. This material is suitable for a welded steel frame such as the proposed test stand.

Durability Test of Engine Support Frame

Stress analysis results showed a maximum von Mises stress of 91.79 MPa, well below the A36 steel yield strength of 250 MPa, ensuring safe operation (Figure 5). Maximum displacement was 0.07026 mm, with 0.06657 mm in the Z-axis, indicating high stiffness (Figure 5). The maximum rotation angle of 0.030 radians was within acceptable limits, though flexible couplings could enhance precision in high-vibration environments.

Calculate Input Parameters

The force exerted by the engine acts on the frame due to gravity

$$F_{ĐC} = \frac{\left(M_{ĐC} \cdot g\right)}{3} = \frac{210.10}{3} = 700\ \left(N\right) \quad [1]$$

(Divided into three sections because the engine is mounted in three different positions relative to the frame system)
Torque is determined based on the engine parameters

$$M_{Obr} = 216\ (\text{N·m}) \quad [2]$$

Force Exerted:

$$F_{M\,obrY} = \frac{M_{Obr} . x_{max}}{\sum x_i^2} = \frac{216000.360}{(360^2 + 250^2 + 260^2)} = 299(N) \quad [3]$$

Moment of Inertia

Moment of inertia of the flywheel:

$$I_{W1} = \frac{1}{2}.m.(R_x{}^2 - R_w{}^2) = \frac{1}{2}.12.\left(\left(\frac{360}{2}\right)^2 - \left(\frac{86}{2}\right)^2\right) = 183306(kgmm^2) \quad [4]$$

Moment of inertia of steel bushing:

$$I_{W1} = \frac{1}{2}.m.(R_x{}^2 - R_w{}^2) = \frac{1}{2}.0,3.\left(\left(\frac{360}{2}\right)^2 - \left(\frac{86}{2}\right)^2\right) = 694(kgmm^2) \quad [5]$$

Total moment of inertia of flywheel and steel bushing:

$$I_2 = I_W 1 + I_W 2 = 183360 + 694 = 184054(kgm^2) \quad [6]$$

Angular acceleration without the additional flywheel:

$$\varepsilon_1 = 34,90(Rad/s^2) \quad [7]$$

Angular acceleration with the additional flywheel:

$$\varepsilon_2 = 23,23(Rad/s^2) \quad [8]$$

Moment of inertia of the crankshaft and piston system in an internal combustion engine:

$$I_1 = \left(\frac{I_2.\varepsilon_1}{\varepsilon_1.\varepsilon_2}\right) = \frac{0,184054.23,23}{(34,9 - 23,23)} = 0,3663(kgm^2) \quad [9]$$

Torque of the engine with an additional flywheel:

$$M_T = \left(I_1 + I_2\right).\varepsilon_2 = \left(0,3663 + 0,184054\right).23,23 = 12,7\ \left(Nm\right) \quad [10]$$

The force exerted by the load of the crankshaft–piston system on the frame system:

$$F_{(M_T Y)} = 12,7.\frac{360}{360^2 + 250^2 + 260^2} = 18(N) \quad [11]$$

The resultant force acting at a given fixed point:

$$F_{1,2} = F_{ĐC} - F_{MOrY} - F_{M_T Y} = \ 700 - 299 - 18 = 383\ \left(N\right) \quad [12]$$

$$F_3 = F_{ĐC} + F_{MOrY} + F_{M_T\ Y}\ F_3 = F_{ĐC} + F_{MOrY} + F_{M_T\ Y}$$
$$= 700 + 299 + 18 = 1017\ (N) \quad [13]$$

where F denotes the gravitational force acting on the frame (N), m is the engine mass (kg), g represents gravitational acceleration (9.81 m/s²), Mx is the engine torque (N·m), I is the moment of inertia (kg·m²), and α is the angular acceleration (rad/s²).

The von Mises stress distribution of A36 steel reaches a maximum value of 91.79 MPa, which is significantly lower than its yield strength of 250 MPa. This ensures that the structure operates safely within the elastic region, preventing plastic deformation.

The maximum displacement of the frame members in any direction is 0.07026 mm, while the maximum displacement along the Z-axis (vertical direction) is 0.06657 mm. These values indicate that the frame exhibits high stiffness and meets load-bearing requirements under normal operating conditions.

The maximum rotation angle of the structure is 0.030 radians, which may introduce minor misalignments. However, this deviation remains within acceptable limits.

Although this small deviation does not significantly affect structural performance, if higher precision is required or the structure operates in a high-vibration environment, the use of flexible couplings or damping solutions should be considered to enhance stability and minimise eccentricity during operation.

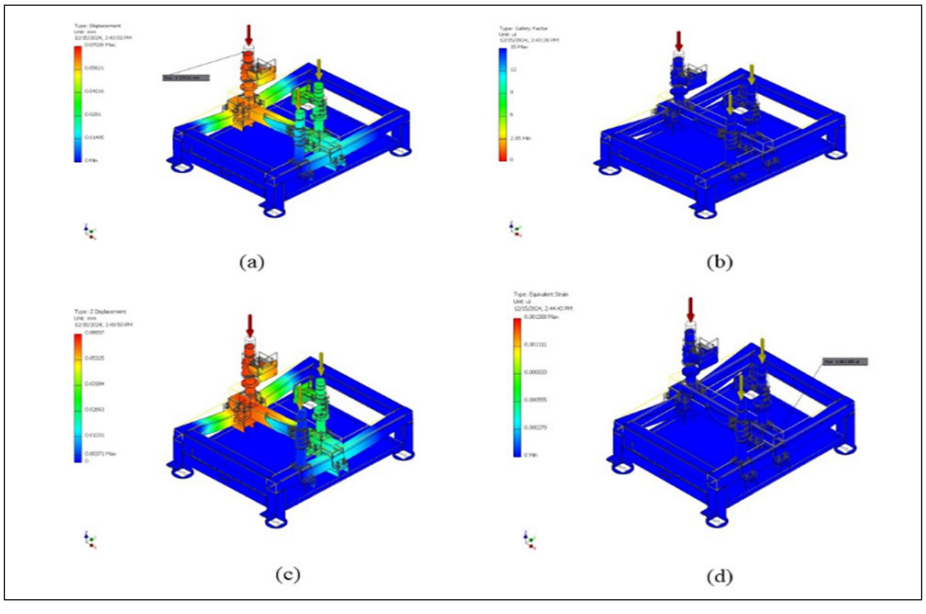


Figure 5. Simulated stress and displacement: (a) total displacement; (b) safety factor; (c) Z-direction displacement; (d) total deformation

Test Stand Vibration Analysis

To conduct a natural frequency analysis for the test stand, follow these steps:

1. Create a natural frequency analysis setup in ANSYS Workbench.
2. Import the designed test stand into the software.
3. Assign material properties to all test stand components.
4. Remove unnecessary details to optimise the simulation.
5. Define contact interactions between components
6. Apply boundary conditions and constraints.
7. Configure simulation parameters.
8. Run the simulation and analyse results.

Vibration analysis in ANSYS Workbench identified 20 natural frequencies, with the lowest at 16.556 Hz, ensuring stability outside typical engine operating ranges (Table 3). The Harmonic Response module showed a peak displacement of 0.2 mm at 38.8 Hz, with minimal impact on measurement accuracy (Figure 6).

Table 3
Natural frequencies of the system

Mode	Frequency [Hz]
1	16.556
2	28.846
3	38.858
4	48.619
5	70.403
6	76.728
7	97.242
8	145.87
9	150.28
10	180.33
11	214.85
12	240.49
13	288.53
14	312.39
15	324.68
16	341.76
17	390.08
18	416.84
19	436.03
20	437.82

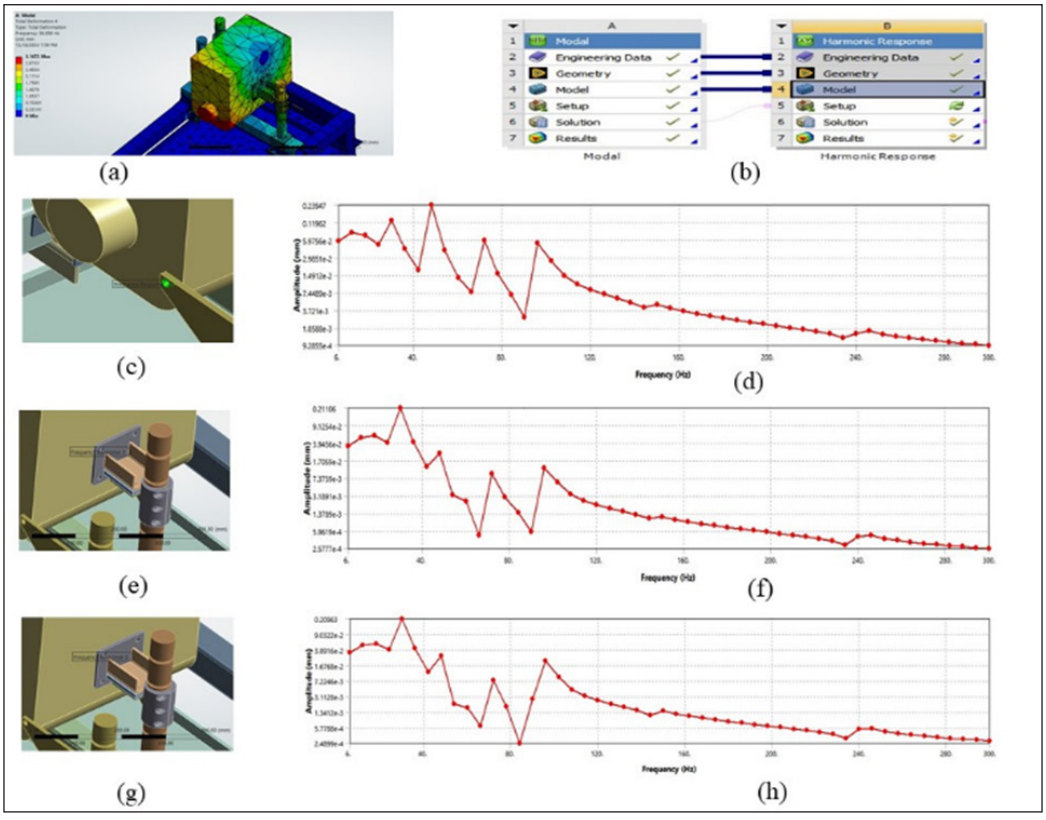


Figure 6. Vibration analysis: (a) ANSYS results; (b) Harmonic response setup; (c) analysis surfaces at Locations 1; (d) total displacement at Locations 1; (e) analysis surfaces at locations 2; (f) total displacement at Locations 2; (g) analysis surfaces at Locations 3; (h) total displacement at Locations 3

Mode Shapes

The first mode exhibited lateral bending of the upper frame, the second mode involved vertical flexing of the mounting platform, and the third mode showed mild torsional deformation. These shapes are typical for lightweight welded steel structures.

Resonance Considerations

The first natural frequency of the test stand was identified as 16.56 Hz (Table 3). This value is slightly higher than the dominant excitation frequency range of the Toyota 3C diesel engine during idle and low-speed operation, which typically lies between 13 and 15 Hz. In practical operation, the test stand is supported by four industrial anti-vibration mounts commonly used for machine tools such as lathes, which are designed to provide effective damping under both static and dynamic loading conditions.

In addition, the cardan coupling used to connect the engine to the MP100S unit introduces a certain degree of torsional flexibility and misalignment tolerance, helping to reduce the direct transmission of dynamic forces. The combined effect of these design features contributes to stable operation of the test stand, and resonance is therefore unlikely to occur under normal testing conditions.

Next, to evaluate the influence of these vibrations on the test stand, the analysis results will be used to examine the displacement of critical components in the test stand using the Harmonic Response module.

The Harmonic Response module is configured based on the data obtained from the natural frequency analysis results.

The total displacement amplitude at the specified positions varies with the frequency of the external force, reaching its maximum value at 38.8 Hz. The peak total displacement is approximately 0.2 mm, which is relatively low and has minimal impact on the experimental measurement results

Manufacturing and Complete Installation of The Test Stand Manufacturing and Installation Requirements

The test stand was manufactured using 75 × 75 × 3 mm A36 steel tubing for the main frame and Ø55 mm round steel for the support shafts (Figure 7). Anti-vibration mounts minimised vibrations, and eccentric bushings with adjustment screws ensured precise positioning. The assembly met requirements for secure connection, concentric alignment, and sufficient cardan coupling insertion (minimum 20 mm). Calibration of the MP100S load-inducing mechanism achieved a standard torque of 250 N·m (Figure 7).

After manufacturing, the engine mounting test stand is connected to the MP100S system and must meet the following conditions:

1. The connection between the test stand and the engine support must be secure.
2. The jack-up protection cover must be firmly installed.
3. The engine and brake assembly must be concentric to ensure smooth operation.
4. The exhaust pipe must be securely installed and airtight.
5. The cardan coupling installation into the brake assembly must ensure sufficient insertion depth (minimum 20 mm) and proper alignment with engagement marks.

The installation position of the entire test stand must allow for a reasonable working space, facilitating ease of experimentation when adjusting engine speed (rpm), loads, and operational conditions.

Complete Assembly of the Test Stand and Calibration

The fabricated test stand has been fully assembled, as shown in Figure 7.

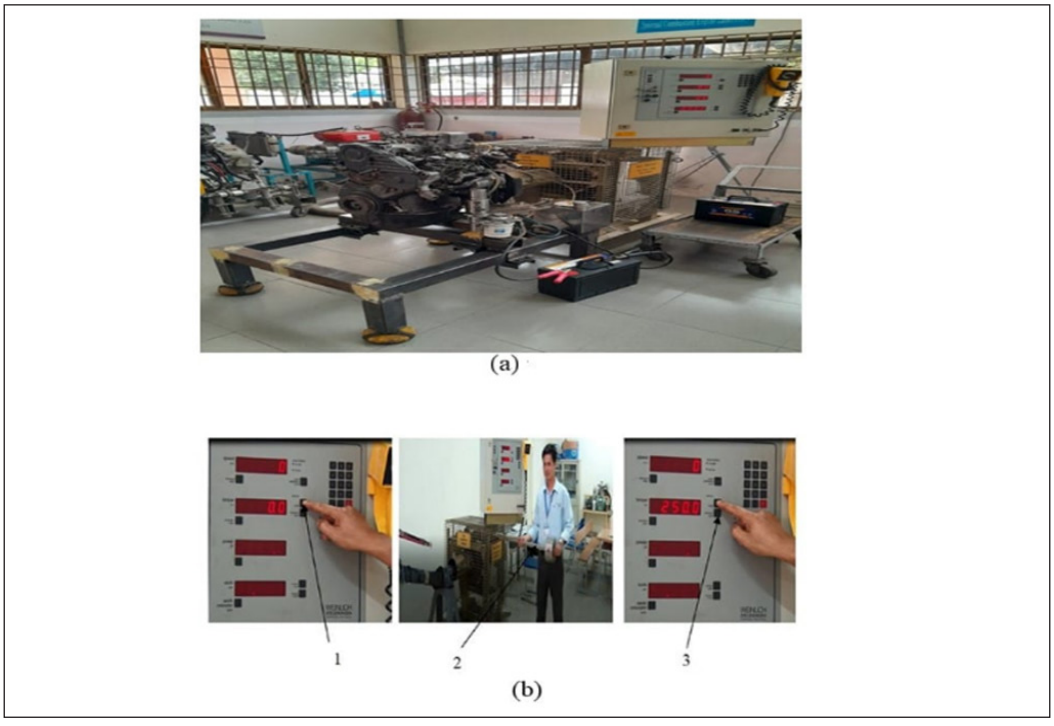


Figure 7. Fabricated test stand and calibration: (a) Fabricated engine test stand and (b) Torque measurement scale

Torque Caliper in the Load-inducing Mechanism

To calibrate the MP100S load-inducing mechanism, follow these steps: Turn on the MP100S load-inducing mechanism and press 0 N·m (Step 1). Insert the calibration tool (Step 2). Select Calibration Value to obtain the standard result of 250 N·m. The standard caliper has a mass of 25 kg, a lever arm length of 1 m, and a resulting torque of 250 N·m, as illustrated in Figure 7.

Experiment

Purpose of the Experiment

The experiment verified the stable connection between the test stand and the MP100S system, measuring engine speed (n, 1000-2400 rpm), torque (Mx), and power (P) (Figure 8).

Test Parameters

Predetermined parameters: Engine speed, n = 1000 to 2400 rpm
Parameters to be measured: Torque, Mx (N·m), engine power, P (kW)

Experimental Setup

The experimental model, as illustrated in Figure 8 (a), used engine speed, n (rpm), as the input parameter, with test modes ranging from 1000 to 2400 rpm. The two output parameters measured were the engine's useful torque, Mx (N·m), and useful power, P (kW). The test stand was fully assembled, and the MP100S dynamometer system was calibrated prior to testing.

Calibration Procedure

Before each test, the MP100S dynamometer was calibrated under no-load conditions to eliminate zero-offset errors, following the manufacturer's recommended procedure

Test Execution

Before operating the engine power tester for the first time, carefully read all instructions for use. Conduct a pre-operation check and ensure compliance with safety principles when using the MP100S device. Pay special attention to preventing the MP100S from exceeding the speed limit specified on the protective cage. Additionally, monitor the load table on the engine tester. Measure the engine power and torque, then record the data using DiaW 1.3 software.

Experimental Results

After fully assembling the test stand, the MP100S performance measuring system was calibrated. The engine was tested under experimental conditions with engine speeds (n) ranging from 1000 to 2400 rpm. The test results determined the useful torque (Mx, N·m) and useful power (P, kW), as shown in Table 4. Additionally, the experimental data were used to generate the corresponding graphs. The power vs. engine speed graph (Figure 8 (b)) indicates that the maximum engine power, Pmax = 11.8 kW, occurs at a speed of 1900 rpm. The torque vs. engine speed graph (Figure 8 (c)) shows that the maximum torque, Mxmax = 66 N·m, is achieved at 1000 rpm, which allows for rapid acceleration, improved heavy-load handling, and fuel efficiency. The power vs. torque graph (Figure 8 (d)) illustrates that maximum power is sustained over a wide range of 1200-2000 rpm, contributing to extended load capacity and enhanced vehicle acceleration. Tests at engine speeds of 1000-2400 rpm yielded a maximum power of 11.8 kW at 1900 rpm and a maximum torque of 66 N·m at 1000 rpm (Table 4; Figure 8). These results align with findings from similar test stands, confirming the system's reliability.

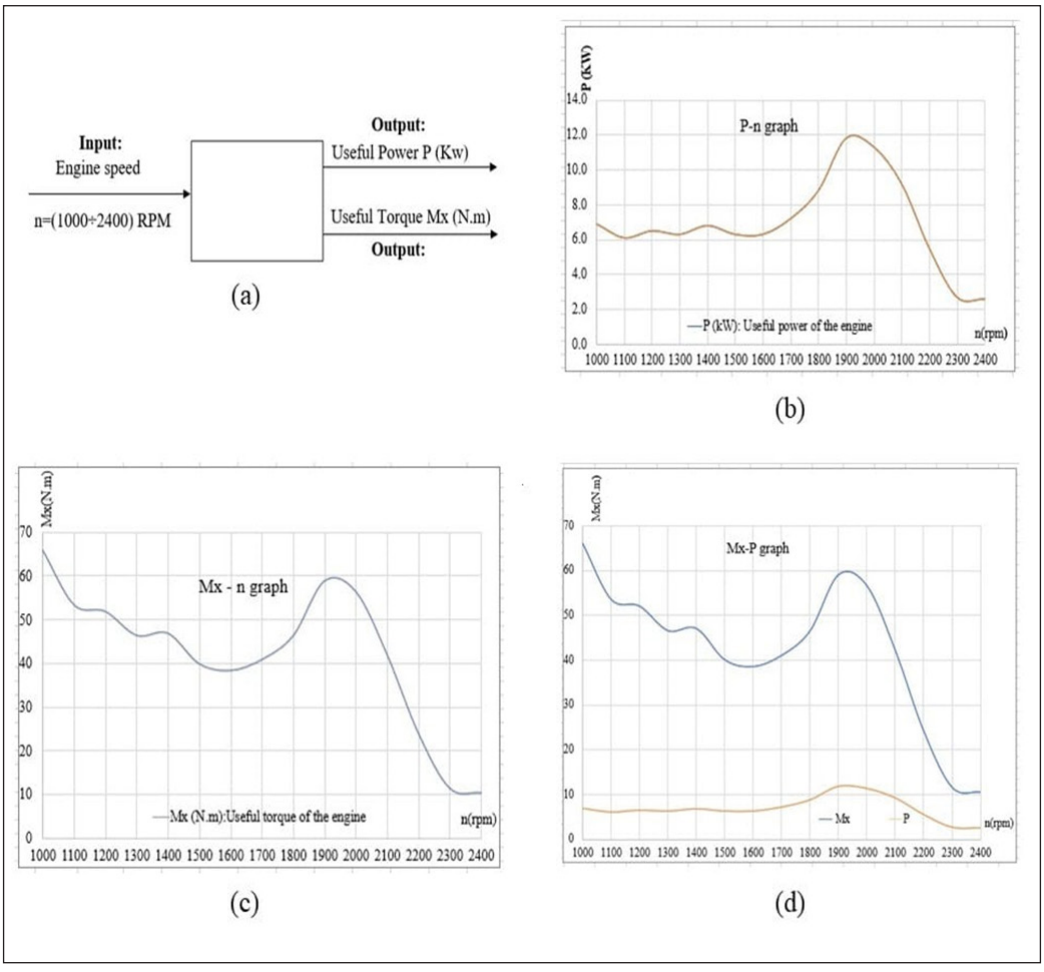


Figure 8. Experimental setup and performance results: (a) Test stand model with input (n = 1000-2400 rpm) and output (P, Mx) via MP100S; (b) Power vs. speed graph with max 11.8 kW at 1900 rpm; (c) Torque vs. speed graph with max 66 N·m at 1000 rpm; (d) Power vs. torque graph (0-11.8 kW, 0-66 N·m)

Table 4
Measured torque (Mx) and power (P) values at engine speeds (n)

n (rpm)	Mx (N·m)	P (kW)
1000	66	6.9
1100	54	6.1
1200	52	6.5
1300	47	6.3
1400	47	6.8
1500	40	6.3
1600	39	6.3

Table 4 (*continued)*

n (rpm)	Mx (N·m)	P (kW)
1700	41	7.2
1800	47	8.8
1900	59	11.8
2000	57	11.3
2100	42	9.2
2200	24	5.5
2300	12	2.7
2400	11	2.6

Measurement Repeatability and Instrument Accuracy

The torque and power values in Table 4 were recorded directly from the MP100S dynamometer system. According to the manufacturer, the MP100S system provides a typical measurement accuracy of ±2%. Although a single measurement was taken at each operating point, this instrument accuracy ensures that the recorded torque and power values remain within the acceptable tolerance range for engine performance evaluation.

It should be noted that the primary purpose of this experiment was to confirm the operational stability of the test stand and its ability to reliably interface with the MP100S measurement system, rather than to conduct a detailed statistical characterisation of engine performance. For this reason, single-point measurements at each operating condition, combined with the certified accuracy of the MP100S system, are considered adequate for the scope of this validation study.

Brake Power Calculation

Brake power was calculated using the relationship:

$$P = \frac{2\pi nM}{60} \quad [14]$$

where n is the engine speed (rpm) and Mx is the torque (N·m).

Engine Specification

The experiment used a Toyota 3C diesel engine with a rated power of 65-67 kW at 4000 rpm and a displacement of 2.184 liters. The torque and power curves obtained from testing

followed the same trend as the manufacturer's specifications, confirming the accuracy of the setup.

Alignment Advantage

Alternative 4 features a symmetric layout that makes it easier to align the engine flywheel with the MP100S shaft. The adjustable bolts allow fine correction of alignment, which helps reduce vibration and transmission losses.

Operational Observation

During testing, the stand ran smoothly with low vibration. The anti-vibration mounts reduced noise and vibration, while the adjustable bolts kept the engine and dynamometer shafts properly aligned throughout the experiment.

Analysis of the Test Results

Design and Manufacturing of a Versatile Engine Test Stand

A test stand has been designed and manufactured to accommodate various engine types and integrate with the MP100S dynamometer system. The test stand ensures proper assembly, operational safety, and the ability to conduct performance tests. It allows measurement of engine speed (n) in rpm, torque (Mx) in N·m, and power (P) in kW across different operating modes (Table 4). Additionally, the data collected enables the generation of power vs. engine speed (Figure 8 (b)) torque vs. engine speed (Figure 8 (c)) and power vs. torque graphs (Figure 8 (d)). Maximum power at 1900 rpm and torque at 1000 rpm indicate robust performance, comparable to results from modular test stands.

The research outcomes extend beyond standard engine testing, enabling various applications such as performance evaluation of turbocharger blades, exhaust pressure analysis, fuel consumption measurement for different operating conditions, emission reduction assessments, ensuring smooth, uniform, and noise-free operation, enhancing efficiency for research, teaching, and practical training.

This test stand provides a valuable tool for universities and technical colleges specialising in automotive technology, supporting both research and education. Furthermore, its application in industrial production contributes to domestic auto parts localisation, improving the industry's capacity for localised manufacturing.

CONCLUSION

The study presents the design and fabrication of a versatile engine test stand that can accommodate different engine types and operate in conjunction with the MP100S

dynamometer system. The proposed configuration satisfies assembly requirements and ensures safe operation across various working conditions. The test stand allows reliable measurement of essential engine performance parameters such as rotational speed, torque, and output power. Experimental testing showed a maximum power of 11.8 kW at 1900 rpm and a peak torque of 66 N·m at 1000 rpm, demonstrating stable and consistent system performance. The developed setup can be applied to a wide range of engine testing and evaluation tasks, supporting studies on performance improvement, efficiency enhancement, emission reduction, as well as applications in research, teaching, and practical engineering activities

Beyond its technical performance, the developed test stand is intended to support laboratory-based education and applied research activities at Vinh Long University of Technology Education, particularly for engine testing experiments and powertrain-related courses.

The present study is limited to structural analysis and functional experimental testing; fatigue assessment and repeated experimental trials were not included and are considered as potential directions for future work.

ACKNOWLEDGEMENT

The author gratefully acknowledges the laboratory facilities provided by Vinh Long University of Technology Education for this research.

LIST OF ABBREVIATIONS

ANSYS	:	Analysis System
ASTM	:	American Society for Testing and Materials
CAD	:	Computer-aided Design
CNG-SI	:	Compressed Natural Gas Spark Ignition
FEA	:	Finite Element Analysis
MP100S	:	MP100S Dynamometer system
NO_x	:	Nitrogen Oxides
RPM	:	Revolutions per Minute
3D	:	Three-dimensional

REFERENCES

Adkine, A., Overikar, P., & Surwase, P. (2017). Modal analysis of engine supporting bracket using finite element analysis. *International Journal of Advanced Engineering Research and Science, 4*(3), 50-55. https://doi.org/10.22161/ijaers.4.3.9

Alfaiz, M. Z., Tahir, M. M., Abu Bakar, R., & Ali, M. S. (2014). New design of a low-cost small engine dynamometer for engine testing. *Applied Mechanics and Materials, 699*, 642-647. https://doi.org/10.4028/www.scientific.net/AMM.699.642

Altarazi, Y. S. M., Saadon, S., Yu, J., Gires, E., Ghafir, M. F. A., & Lucas, J. (2020). On-design operation and performance characteristic of custom engine. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 70*(1), 144-154. https://doi.org/10.37934/arfmts.70.1.144154

Calvo, J. A., Diaz, V., & San Román, J. L. (2006). Controlling the turbocharger whistling noise in diesel engines. *International Journal of Vehicle Noise and Vibration, 2*(1), 17-28. https://doi.org/10.1504/IJVNV.2006.008524

Chen, S., Ma, C., & Zhang, H. (2022). Engine performance improvements through turbocharger matching and turbine design. *Energy Science & Engineering, 10*(9), 3384-3396. https://doi.org/10.1002/ese3.1225

Crispi, M., Cervone, D., Arsie, I., Pianese, C., & Polverino, P. (2022). A methodology for the experimental validation at the engine test bed of fuel consumption and NOx emissions reduction in a HEV. *SAE Technical Paper, 2022-24-0006*. https://doi.org/10.4271/2022-24-0006

Derbiszewski, B., Woźniak, M., Józinkiewicz, D., & Zakrzewski, S. (2023). Research of dynamic phenomena in a model engine stand. *Open Engineering, 13*(1), Article 20220436. https://doi.org/10.1515/eng-2022-0436

Huang, J., Lin, L., Wang, Y., Qin, J., Roskilly, A. P., Li, L., Ouyang, T., & Yu, Y. (2011). Experimental study of the performance and emission characteristics of diesel engine using direct and indirect injection systems and different fuels. *Fuel Processing Technology, 92*(7), 1380-1386.https://doi.org/10.1016/j.fuproc.2011.03.001

Izvorean, S., & Stoica, V. (2022). Experimental test bench with diesel engine for the study of multi fuels usage. *IOP Conference Series: Materials Science and Engineering, 1220*(1), Article 012014. https://doi.org/10.1088/1757-899X/1220/1/012014

Lee, N., Lee, M., Kim, H., & Kim, J. (1998). Design of engine mount using finite element method and optimisation technique. *SAE Technical Paper, 980379*. https://doi.org/10.4271/980379

Liu, C., Cao, Y., Liu, Y., & Wenping, Z. (2019). Numerical and experimental analyses of intake silencer and its effects on turbocharger compressor performance. *Advances in Mechanical Engineering, 11*(3), 1-15. https://doi.org/10.1177/1687814019826677

Lloyd, A. C., & Cackette, T. A. (2001). Diesel engines: Environmental impact and control. *Journal of the Air & Waste Management Association, 51*(6), 809-847. https://doi.org/10.1080/10473289.2001.10464315

Mohammadi, M., Nazemosadat, S. M. R., & Afsari, A. (2024). Modelling and analysis of tractor diesel engine test stand structure using the finite element method. *Biomechanism and Bioenergy Research, 3*(2), 75-87. https://doi.org/10.22103/bbr.2024.24161.1090

Ni, S., Guo, Y., Lv, B., Wang, D., Li, W., & Shuai, Z. (2020). Analysis of torsional vibration effect on the diesel engine block. *Mechanics & Industry, 21*(5), Article 522. https://doi.org/10.1051/meca/2020065

Norouzi, A., Shahpouri, S., Gordon, D., & Winkler, A. (2022). Deep learning-based model predictive control for compression ignition engines. *Control Engineering Practice, 127*, Article 105299. https://doi.org/10.1016/j.conengprac.2022.105299

Oliveira, G. L., de Souza, C. M. A., Arcoverde, S. N. S., Venturin, J. C., Armando, E. J., & Donaire, L. O. (2022). Construction of low-cost benches to assist in teaching machine elements and small engine testing. *International Journal for Innovation Education and Research, 10*(12), 113-122. https://doi.org/10.31686/ijier.vol10.iss12.3911

Park, H. (2023). Structural design and numerical analysis of hoisting device of test bed for aircraft engine. *Applied Mechanics, 4*(2), 407-420. https://doi.org/10.3390/applmech4020023

Qi, Y., Liu, W., Liu, S., Wang, W., Peng, Y., & Wang, Z. (2023). A review on ammonia-hydrogen fueled internal combustion engines. *eTransportation, 18*, Article 100288. https://doi.org/10.1016/j.etran.2023.100288

Santhosh Kumar, H. V., Bharadwaj, K. S. P., Rohith, S., Mridhul, S., & Vishnu, T. (2022). Modal analysis of passive engine mount assembly using finite element analysis. *ARAI Journal of Mobility Technology, 2*(1), 158-167. https://doi.org/10.37285/ajmt.1.1.10

Shao, B. W. (2011). Optimum design of engine test bench based on finite element analysis. *Advanced Materials Research, 338*, 255-258. https://doi.org/10.4028/www.scientific.net/AMR.338.255

Singh, R. (2000). Dynamic design of automotive systems: Engine mounts and structural joints. *Sādhanā, 25*(3), 319-330. https://doi.org/10.1007/BF02703548

Subramaniam, H., Duraisamy, S., Raghavan, G., & Govindan, S. (2019). Design and development of a test rig for the performance evaluation of automotive exhaust thermoelectric generator. *AIP Advances, 9*(6), Article 065004. https://doi.org/10.1063/1.5093587

Tatarynow, D., Longwic, R., Sander, P., Zieliński, Ł., Trojgo, M., Lotko, W., & Lonkwic, P. (2022). Test stand for a motor vehicle powered by different fuels. *Applied Sciences, 12*(20), Article 10683. https://doi.org/10.3390/app122010683

Tomin, S., Keller, K., Wagner, U., Lott, P., Koch, T., & Deutschmann, O. (2024). Innovative engine test bench set-up for testing of exhaust gas aftertreatment and detailed gas species analysis for CNG-SI-operation. *Automotive and Engine Technology, 9*(2), 2-22. https://doi.org/10.1007/s41104-024-00140-8

Visan, N. A., Carlanescu, R., Niculescu, D. C., & Chiriac, R. (2022). Study on the cumulative effects of using a high-efficiency turbocharger and biodiesel B20 fuelling on marine diesel engines. *Journal of Marine Science and Engineering, 10*(10), Article 1403. https://doi.org/10.3390/jmse10101403

Wang, Y. M., Xu, K., & Zhang, Q. (2021). Modal analysis and analysis of motor support based on workbench test bench design of motor dynamic shock absorber. *Journal of Physics: Conference Series, 1986*(1), Article 012096. https://doi.org/10.1088/1742-6596/1986/1/012096

Yu, X., Zhu, L., Wang, Y., Filev, D., & Yao, X. (2021). Internal combustion engine calibration using optimisation algorithms. *Applied Energy, 305*, Article 117894. https://doi.org/10.1016/j.apenergy.2021.117894

Pertanika J. Sci. & Technol. 34 (2): 671 - 693 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Measuring the Performance of Blockchain Systems with an Improved Consensus Mechanism using Scalability Metrics

Kolli Lalitha Kumari and Lalitha Surya Kumari*

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, 500075 Hyderabad, Telangana, India

ABSTRACT

Blockchain is an efficient method to manage and secure data, but scalability remains a limitation. The proposed work concentrates on a novel consensus algorithm, Proof of Useful Work-Authorisation-Storage Availability. The work emphasises the scalability issues and trust-related issues in blockchain-based smart contracts. The primary step in the proposed technique is to verify authorisation by the hash code of the preceding block to generate a digital signature. After information is legitimised, check whether enough storage space is available or not. Transactions are recorded to a block only after they are validated and there is enough space. A transaction is included in a block only after it is mined. To compare how efficient the proposed consensus is, in terms of energy consumption, delay, and the number of transactions it can process, the study uses Python and Solidity. The new method reduced computational energy consumption by 39% and increased transaction capacity by 17%, and decreased network delay by 49% compared to the proof of useful work algorithm. This is an indication that the efficient consensus will help us scale without compromising security and decentralisation. The contribution is beneficial in determining which nodes to apply in the consensus layer. The system can help with banking applications and supply chain systems that are blockchain-based. It highlights three key points: scalability, low delay, and efficient consensus, which are necessary for these systems to function properly.

Keywords: Blockchain, consensus mechanism scalability, computational energy, latency, transaction throughput

ARTICLE INFO

Article history:
Received: 24 May 2025
Accepted: 04 February 2026
Published: 01 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.04

E-mail addresses:
2212030001@kluniversity.in (Kolli Lalitha Kumari)
vlalithanagesh@gmail.com (Lalitha Surya Kumari)
* Corresponding author

INTRODUCTION

Blockchain technology has applications in various fields, such as banking, supply chain management, and the healthcare industry (Andrew et al., 2023).

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

As the network grows in blockchain technology, scalability problems arise. Scalability solutions try to address this difficulty while still ensuring the decentralised nature of the blockchain (Zhang et al., 2021). It is also essential in ensuring interoperability among various blockchain networks. With the increase in the use of blockchain, the ability to exchange information and communicate smoothly across multiple networks is becoming more important. From Figure 1 the working components of blockchain to complete a transaction have been presented. Before adding a transaction to a particular blockchain, it must be verified through mining, consensus, and hashing. These components also enhance the overall performance of a blockchain. Miners validate transactions to be recorded on the global ledger. It should be converted into a fixed-size string of characters when hashing data. Validating blocks of data in a blockchain consensus protocol will be helpful. Based on the review of available literature, scalability (Deng et al., 2024; Fajri et al., 2022; Liu et al., 2021; Ucbas et al., 2023) is a significant issue for handling many nodes that will be added to a blockchain network.

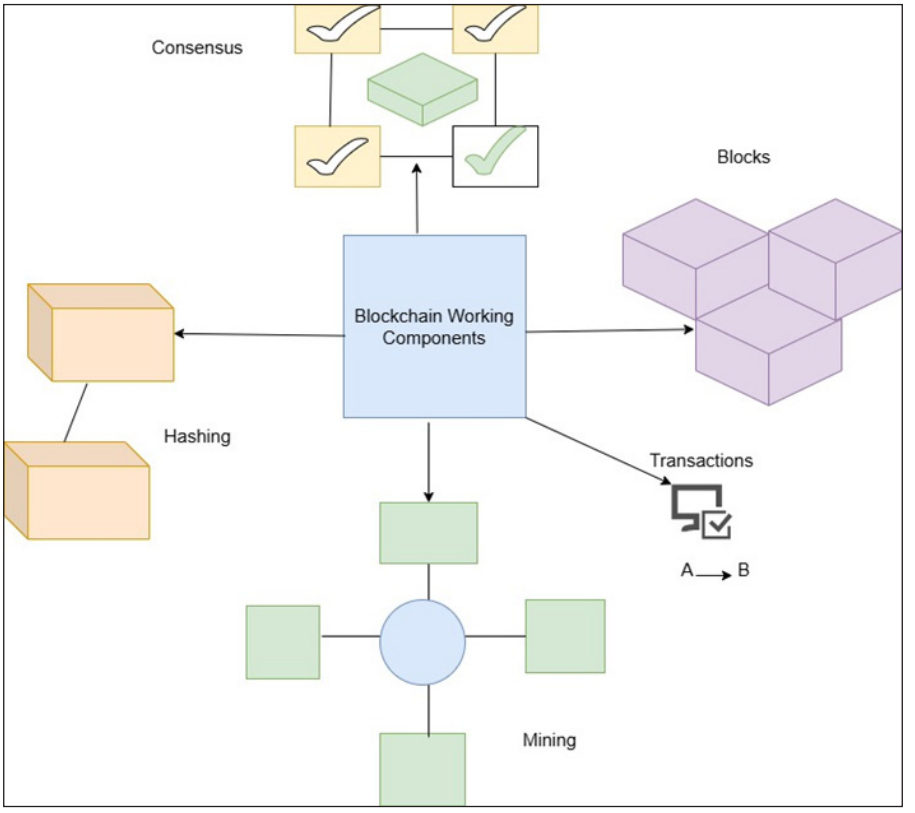


Figure 1. Blockchain working components

Scalability

The ability of a network to perform an increase in the number of transactions without compromising overall efficiency is known as scalability. Because scalability ensures that a network can handle a higher volume of transactions per second. Congestion in a blockchain network can be caused by high demand and limited scalability, resulting in increased transaction fees. Scalability also allows transactions to be more efficient and faster, thus lowering transaction costs. A scalable blockchain provides low latency and fast transaction confirmation, resulting in a more efficient user experience. Strong transaction management will be needed for widespread acceptance of blockchain technology.

Consensus Algorithms

Blockchain networks use the Proof of Work (PoW) consensus protocol to obtain consensus on the state of the distributed ledger as well as protection against nefarious behaviour, such as double spending. In this consensus, the miners race to be the first to solve a computationally difficult puzzle to provide a hash that meets a criterion, such as having a certain number of leading zeros. It is practically not possible to change the blockchain because it requires huge computational resources and energy. After a solution is found, the miner announces the solution or proof to the network, and other nodes can easily verify the proof. The first miner to solve a puzzle correctly is rewarded with a cryptocurrency, and their proposed block is attached to the network.

Proof of Storage (PoS) is a blockchain consensus algorithm that uses storage capacity rather than computational capacity to secure the network. Miners commit idle disk space to hold cryptographic data, e.g., plots or challenges. To produce a block, a miner proves the use of some storage on the network by submitting valid proof computed from their held plots. Proof of Storage is more efficient than Proof of Work because it requires less computational power, instead leveraging the high availability of disk space.

In blockchain technology, the Proof of Authority (PoA) comes to consensus through the delegation of the process to a list of predetermined validators selected according to their verified identity, credibility, and ability to verify transactions. It is based on computational power or token ownership. PoA trusts the validators more, typically verified parties, such as organisations or people with a high reputation and credibility. Validators generate new blocks and validate transactions to maintain high throughput. This approach lowers energy usage and latency when compared to other algorithms specifically meant to apply to consortium and private blockchains. PoUW is an innovative blockchain consensus protocol that addresses the shortcomings of Proof of Work (PoW) by repurposing miner computational work towards meaningful real-world and tangible problems. Miners within a PoW network are performing beneficial activities, such as developing machine learning models, performing scientific simulations, or cracking hard systems, instead of computing

random cryptographic functions. In being effort-driven computationally, this process secures the blockchain but also channels the effort towards beneficial outputs. PoUW is a more efficient and sustainable blockchain consensus choice because it keeps the security and decentralisation benefits of PoW but significantly lowers energy expenditure.

In this paper, the proposed work aims to develop an improved consensus algorithm and measures such as computational energy, latency, and transactional throughput to help alleviate the current limitations. This paper's structure elaborates on the limitations of existing work, including scalability. The second part focuses on a proposed methodology, based on the improved consensus algorithm and the PoUWAS architecture, as well as scalability metrics and an algorithm. The third part mainly focuses on measures related to scalability, then compares the existing PoUW and proposed PoUWAS in terms of latency, transaction throughput, and computational energy.

RELATED WORKS

To minimise the transaction cost and improve scalability, a new fault-tolerant method was proposed (Otsuki et al., 2021). An optimised blockchain solution was developed to minimise the computational complexity and improve the performance (Ghamdi, 2022). A brief discussion on PoW, PoS, DPoS, and PoA was given, along with the performance parameters (Kaur et al., 2021). The study analyses how long transactions take to be confirmed in proof-of-work blockchains like Bitcoin (Malakhov et al., 2023).

By enabling fast, low-cost, and accountable decision-making for private or consortium systems like corporate governance, supply chains, or smart cities, a model was proposed based on PoA (Manolache et al., 2022). Improved PBFT consensus reduces communication overhead and enhances scalability while maintaining strong fault tolerance (Yang et al., 2022). In summary, scalability is the main issue for improving blockchain performance. Various techniques, including sharding and consensus protocols, are available to address this issue. Particularly, enhanced consensus algorithms are also essential to maintain the network as scalable. Latency, transactional throughput, and computational energy are a few scalability metrics to maintain low latency, greater transactional throughput, and minimal computational energy to overcome the scalability issue.

Massive IoT Data Collection (MIDC) system was mproposed to improve efficiency, security, mand trust in mlarge-scale heterogeneous WSNs (Zhang et al., 2022). The trade-offs between security, decentralisation, and scalability in Industry 4.0 applications of blockchain technology are examined, as well as how important consensus processes are to preserving security and functionality, particularly in permissioned blockchains (Nasir et al., 2024). ALB-Chain, a scalable blockchain system that uses lightweight Account Aggregation Messages (AAM) and a replica account method to balance shard workloads by

optimising shard performance, and the MLPA dynamic partitioning technique outperforms current systems in terms of throughput and latency (Wu & Hua, 2025). Blockchain-based healthcare access control frameworks with an emphasis on smart contract-driven security and privacy-preserving strategies were proposed (Tawfik et al., 2025).

According to Table 1, PoS (Sun et al., 2021) decreased power consumption but created fairness problems. PoUW impacts the blockchain consensus activity by applying its effective scalability and low-energy-consuming technique (Daely et al., 2024). Related work has explored how blockchain technology addresses topics such as scalability. The proposed PoUWAS consists of mining, authorisation, and storage availability aspects. This work limits the scalability issue and highlights the comparison between existing PoUW with the proposed work in terms of latency, transactional throughput, and computational energy. This technique concentrates on enhancing the PoUW algorithm. In the end, it concluded with future enhancements and research directions.

Table 1
Algorithm 1: Findings related to existing system

Author & Year	Consensus Algorithm	Key Features	Limitations	Outcome Performance
(Sun et al., 2021)	Delegated Proof of Stake (DPoS)	Elected validators, fast block creation	Partial centralization centralisation	Up to 1,000 TPS, low latency
(Daely et al., 2024)	Proof of Useful Work (PoUW)	Computational efficiency, sustainability	Work verification complexity	Moderate scalability, reduced energy use

Problem Statement

Blockchain technology has a strong basis in decentralised systems, but still, real-time applications have limitations in terms of scalability and energy consumption, particularly in Proof of Work (PoW). To integrate mining-related activities with scientific solutions and AI model training, PoUW came into existence to address the traditional barriers. Still, these techniques lead to high energy consumption, and issues related to task verification complexity, adaptive scalability, and real-time responsiveness are lacking in current PoUW-related applications. A PoUW consensus that not only assures security but also distributes work across nodes to reduce latency and increase computing energy is required to address these gaps. To provide scalable, energy-efficient, and real-time decentralised applications, the proposed work presents an improved PoUW model that combines blockchain mining with security and authorisation. This method offers the best alternative solution for decentralised applications of the future.

Research Gaps

In healthcare and the Internet of Things, the current limitation is the adoption of blockchain technology and latency-related issues. Because existing consensus protocols either achieve speed at the expense of decentralisation (like PoS and DPoS) or suffer from poor scalability and high energy consumption (like PoW), the main gap is the inability to simultaneously solve the blockchain trilemma (decentralisation, security, and scalability). The lack of adaptability and intelligent node involvement is a crucial second gap. Most protocols employ static rules and do not allow for dynamic node selection or validation process adjustments based on trust metrics or real-time network conditions. To make blockchain a viable solution for high-volume, real-time enterprise systems, there is a general research need for flexible, energy-efficient hybrid consensus algorithms that can intelligently manage node resources, maintain high throughput, and provide low latency.

Novelty

To overcome the limitations, a composite blockchain paradigm should be constituted that synergistically integrates the security strength of the cryptographic consensus with the adaptivity of the improved algorithms. The presented solution is to incorporate a node selection process that uses improved consensus techniques to avoid duplication and improve block generation. It is needed for improved throughput, lower energy use, and quicker transaction confirmation times. This is accomplished by modelling the described methodology in Python and Solidity, deploying the model on a local Ethereum Remix, and testing the performance of the model based on energy cost, latency, and block validation time.

Impact

The proposed consensus technique concentrates on making the blockchain efficient, highly fault-tolerant, and less energy-consuming, which leads to the blockchain being scalable. This consensus technique will be the best alternative to apply to real-time applications to address the traditional barriers. The proposed work concentrates on mining-related activities and authorisation, and storage availability checks also. To maintain data efficiency and security, this method will be the best fit to sustain.

Organisational Chart of the Proposed Methodology

Initially introduced the gaps related to existing consensus algorithms, the need for improvement, and an overview of the proposed consensus algorithm. Then it highlights the steps in the improved consensus algorithm as well as its architecture, including node authorisation, selection, creation of a block, validation, and propagation of data. Then the

flowchart describes node-to-node interactions and a validation process for the data and how a block gets added. The proposed consensus algorithm is described in the next step by outlining how node joining, transaction creation, broadcasting, and validation. Then it concludes with a mathematical analysis on energy consumption. Figure 2 presented clearly the organisational chart of the proposed methodology.

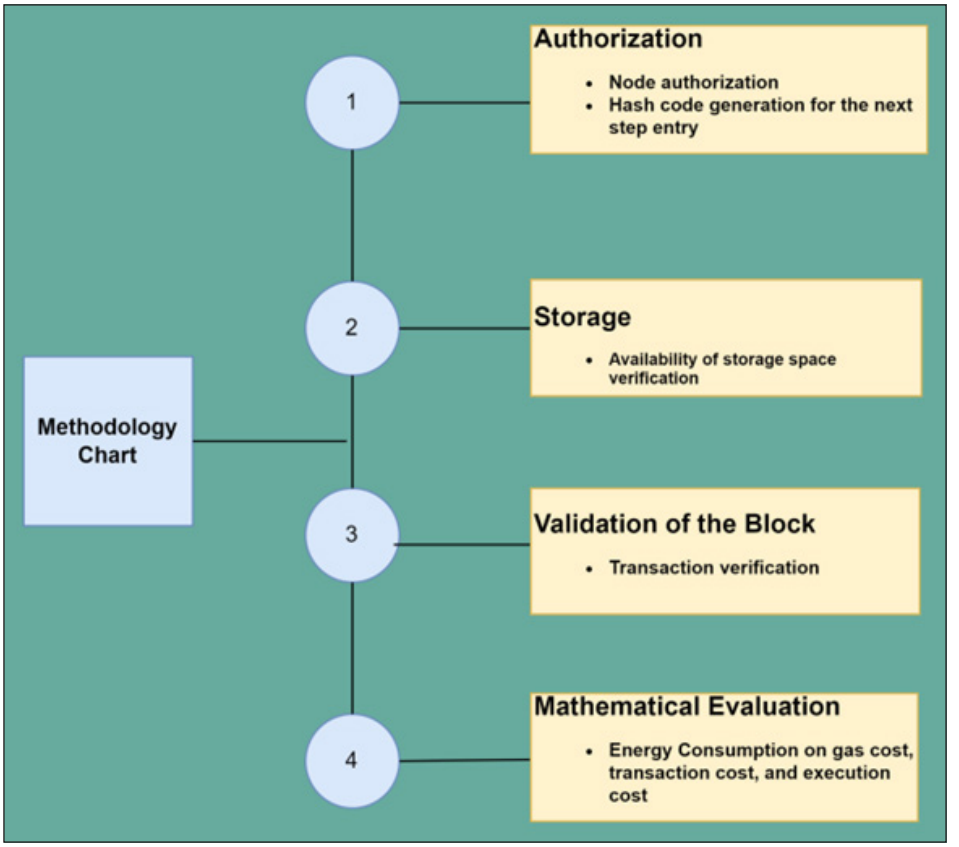


Figure 2. Organisational chart of proposed methodology

PROPOSED METHODOLOGY

Introduction

This approach aimed at designing and developing an improved consensus algorithm to perform a comparative analysis based on the existing algorithms. Based on the parameters, computational energy, latency, and transactional throughput scalability were measured. The motive of this proposed work is to minimise the scalability issues. Apart from this, it is also against fraudulent data on the network, which is lagging in the existing PoUW. Also, the existing algorithm consumes lots of computational power and storage space.

The proposed work will limit the problems and enhance the performance of a network. Figure 3 shows the architecture of the proposed algorithm, PoUWAS. Whenever a block is created, it is added to the blockchain. A specific procedure must be followed to add a block to the blockchain. It checks the transaction's authorisation, storage capacity, and block validation. This proposed architecture combines all three modules: authorisation, storage procedures, and verification.

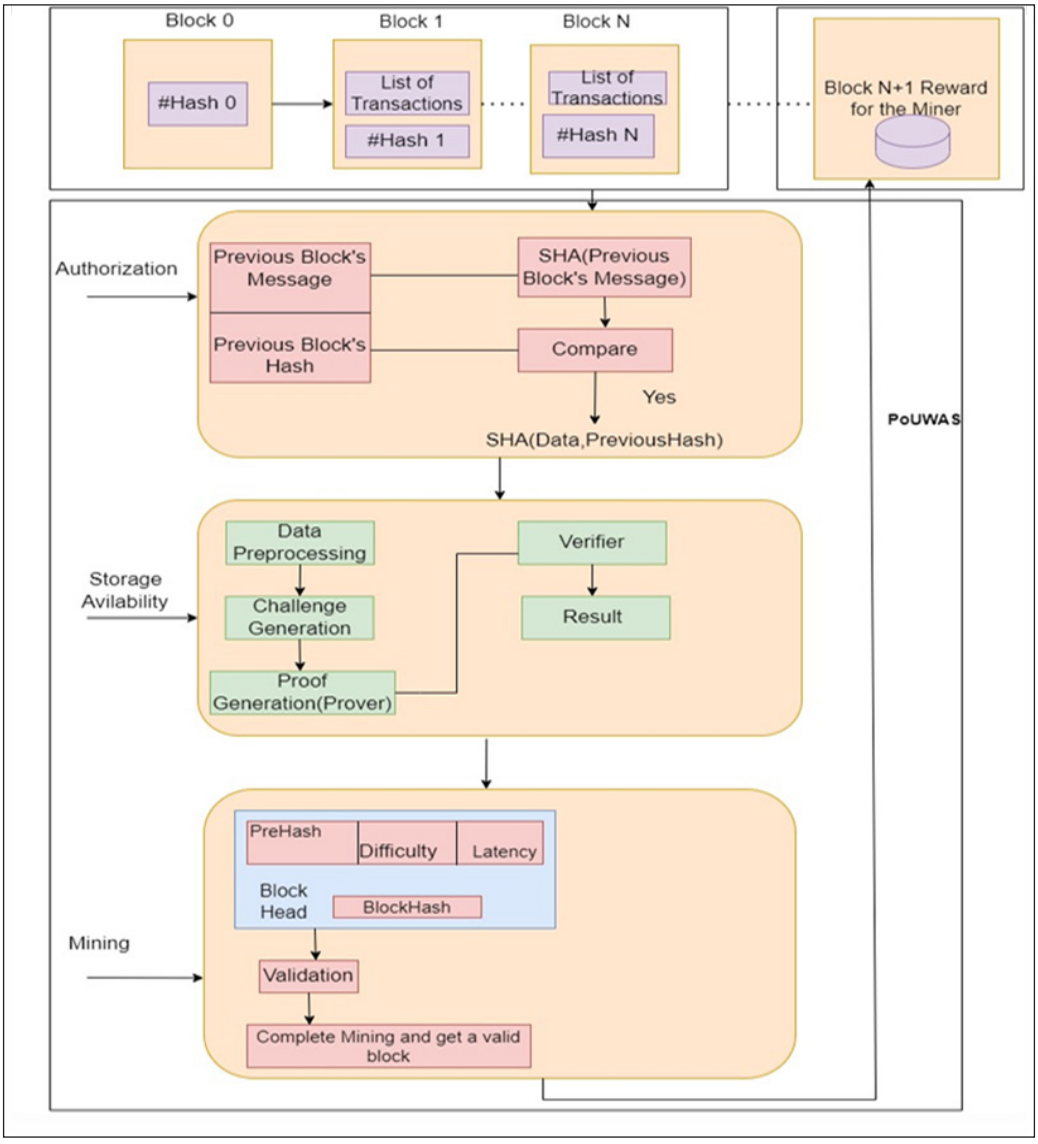


Figure 3. Improved PoUW consensus algorithm architecture

Authorisation

The authorisation method enhances blockchain security by reducing the likelihood of fraudulent activity. Any malicious activity risks harming the real-world reputations of authorisation networks and subjecting them to sanctions, such as revoking their authority, because they depend on a few verified and identifiable validators. Due to this accountability, the validator is discouraged from engaging in dishonest activities, such as altering transaction history or attempting to launch a double-spend attack. The authorisation function takes the following steps. It requires three parameters to proceed, which are the data of the previous block, the hash code of the previous block, and the data of the current block. As a first step, it verifies if the last code hash is indeed the correct one by creating the last code hash and comparing it with the existing previous hash code. Following node authentication, the only hash code that must be relied upon and is needed for the system to work is the one that corresponds to the current block and is generated with the assistance of data and the previous hash code.

Storage

The availability of storage will be verified after block validation. When space is available, the block will be inserted into the blockchain. It also concentrates on reducing scalability problems in a blockchain. Nodes that do not provide on-stored data get penalised. It also maintains efficiency to give robust performance on a blockchain. Plays a vital role in bandwidth-effective usage. If proper storage is not available, then it will follow the rejection process.

Validation of the Block

In this step, the block is validated. When added to the block's transactions, miners compete to find a nonce that meets the required difficulty level to generate a hash. Validation of the block is performed by generating a hash code. It finds latency, which indicates the time taken to validate the block. Difficulty indicates the security level of the block, where zeroes are added as prefixes. It creates a hash code for the current block using the previous hash. By considering these three parameters, the block is validated. This is the concept used by PoUW. To calculate the execution cost, considering not only transaction cost, gas cost, overhead cost, base cost, latency, and computational energy, the proposed work extends the formula to incorporate additional components.

Figure 4 represents how a block is added to a blockchain. Initially, every block-related transaction must be mined, and the verifier must be verified. If the successful execution of every phase is confirmed, the block will be added to the blockchain. Otherwise, it will not be added or included. In this proposed work, verification is done in two different ways after mining. The block is verified based on authorisation and storage availability, allowing scalability issues to be overcome. In Table 2, Algorithm POUWAS available data storage

must be examined for the new block generation, and then the miner's authentication must be verified to proceed with block construction activities. The leader is rejected at once if unauthorised. Lastly, the PoUWAS algorithm proves whether a transaction is valid to add a block, according to the availability of storage, authorisation, and correct mining of the transactions. All these three must be integrated to increase the scalability of the blockchain but also keep it functional against unknown attacks.

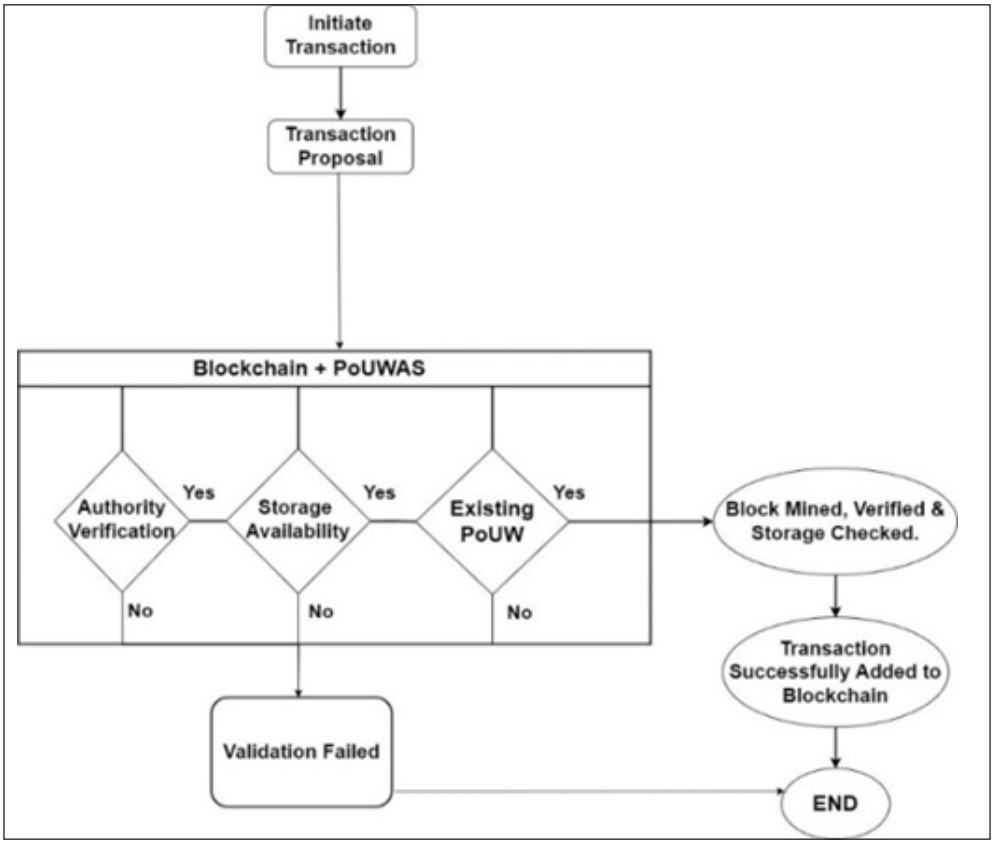


Figure 4. Flowchart for an improved/hybrid consensus algorithm

According to Table 2, the proposed algorithm combines authorisation, storage validation, and mining. It verifies block authorisation by observing whether the hash and current block index accurately follow the previous block. Then compares the requested storage and challenge replies to verify storage capabilities. Lastly, it measures delay as computational effort and executes a Proof of Useful Work (PoUW) by satisfying the specified difficulty level in a limited number of iterations. Finally, a block is added to the blockchain after all verifications are successful otherwise, it is rejected for lack of adequate authority, storage, or work.

Table 2
Proof of useful work-authorisation-storage availability-PoUWAS

```
Inputs:
   prev_idx, curr_blockID, content, prev_content, prev_digest,
   content_digest, node_key, challengeSet [], req_space,
   tx_records, last_digest, validator_key, useful_task
Outputs:
   Confirms block authorisation
   Confirms storage capability
   Adds validated block or rejects insufficient proof
Procedure Verify_BlockAuth (prev_idx, curr_blockID, content, prev_content, prev_digest)
   hash_code ← SHA (content)
   IF prev_idx + 1 = curr_blockID THEN
      IF SHA (prev_content) = prev_digest THEN
         RETURN true
      ELSE
         RETURN false
      END PROCEDURE
Procedure Check_Storage(challengeSet [], req_space, free_space)
   IF free_space ≥ req_space THEN
      IF file_chunk = challengeSet[data_index] THEN
         IF resp_code = SHA (file_chunk) THEN
            IF check_hash = SHA (requested_chunk) THEN
               RETURN true
            ELSE
               RETURN false
END PROCEDURE
Procedure Useful_work (level, max_loops = 100000)
   begin_time ← current_time
   probe_val ← random (1500 to 12000)
   loop_counter ← 0
   WHILE probe_val is NOT prime AND loop_counter < max_loops DO
      probe_val ← probe_val + 1
      loop_counter ← loop_counter + 1
   END WHILE
   IF loop_counter = max_loops THEN
      THROW Exception "Computation Limit Reached"
   END IF
   proof_value ← probe_val
   proof_str ← string(proof_value)
   WHILE proof_str does NOT start with '0' repeated level times AND loop_counter < max_loops DO
      proof_value ← proof_value + 1
      proof_str ← string(proof_value)
      loop_counter ← loop_counter + 1
   END WHILE
   delay_time ← current_time - begin_time
   APPEND delay_time TO self.latency_list
   RETURN proof_value, delay_time
END PROCEDURE

Procedure Main_BlockValidation()
   auth_result ← Verify_BlockAuth(prev_idx, curr_blockID, content, prev_content, prev_digest)
   IF auth_result = false THEN
      RETURN "Authorisation Failure"
   END IF

   storage_result ← Check_Storage(challengeSet[], req_space, free_space)
   IF storage_result = false THEN
      RETURN "Storage Verification Failure"
   END IF
 proof, delay ← Useful_Work(difficulty_level, max_loops)
   IF proof is valid THEN
      ADD content To blockchain
      RETURN "Block Verified and Added"
   ELSE
      RETURN "Block Rejected: Insufficient Useful Work"
   END IF
END PROCEDURE
```

Mathematical Evaluation

It is a crucial part of describing whether a blockchain system is scalable, particularly in the case of improved consensus mechanisms. Mathematical evaluation plays an important role in measuring the scalability of a blockchain system. Performance metrics like block validation time, transaction throughput, computational energy, and network latency can highlight how the system performance varies when the number of nodes or transactions increases. This helps to identify potential risks, optimise resource allocation, and ensure that the proposed algorithm can efficiently handle high-end networks without limiting security or performance. So, it serves as a solid base for proof of the applicability and strength of the enhanced consensus mechanism on a practical blockchain. Table 3 summarises the notations used in mathematical evaluation. Based on the equation 1, 2, 3, 4, values related to gas cost, transaction cost and execution costs are measured. Followed by 5,6, 7, 8, 9, 10, and 11 represent the latency and transaction throughput of the proposed consensus algorithm-based blockchain.

$$TC = GC + C_{base}\,, \quad [1]$$

$$GC = G_{used} \times G_{price}\,, \quad [2]$$

$$EC = E \times C_{energy}\,, \quad [3]$$

$$C_{total} = Transactional\ Cost + Gas\ Cost + Execution\ Cost. \quad [4]$$

$$L1 = ET - ST, \quad [5]$$

$$Tt = No.of\ transactions\ /L, \quad [6]$$

$$Overall\ L1 = \sum_{i=0}^{N}(Li), \quad [7]$$

=0.113+0.063+0.015+0.238+0.097+0.045+0.031+0.319+0.02+0.026,
=0.967

$$Overall\ Tt1 = \sum_{i=0}^{N}(Ti), \quad [8]$$

=8.84+31.59+190.37+16.76+51.5+131.56+219.31+25.02+432.7+376.04
=1483.69

$$L2 = ET - S\,T\,, \quad [9]$$

$$Overall\ L2 = \sum_{i=0}^{N}(Li), \quad [10]$$

= 0.031 + 0.268 + 0.371 + 0.043 + 0.256 + 0.096 + 0.063 + 0.239 + 0.053 + 0.029,
=1.449

$$Overall\ Tt2 = \sum_{i=0}^{N}(Ti) \quad [11]$$

=31.45+7.43+8.08+92.20+19.50+62.44+109.81+33.37+169.27+335.71,
= 869.26

Table 3
Notations related to mathematical evaluation

Notation	Meaning	Description
TC	Transaction Cost	The cost associated with the transaction itself, such as network fees, validator fees, etc.
GC	Gas Cost	The cost of performing computations in a blockchain-based system is calculated as Gas Units × Gas Price.
$C_{latency}$	The cost per unit of latency	(e.g., economic cost or penalty per second of delay).
E	Energy consumed	Measured in Joules or Kilowatt-hours (kWh), depending on the system.
C_{energy}	Cost per unit of energy	cost per kilowatt-hour
C_{total}	Total cost	Measured in Joules or Kilowatt-hours (kWh), depending on the system.
C_{base}	Minimum gas price determined by the network	Measured in currency upon Kilowatt-hours ($/kWh), depending on the system.

Note. Energy consumption interms of Gas cost, Transaction cost and Execution cost with the help of mathematical notations

RESULTS AND DISCUSSION

Blockchain-related activities are conducted with the assistance of Remix IDE and Python Spyder. In Remix IDE, different consensus algorithms like PoUW and PoUWAS were executed, and the computational energy was observed from Figure 5. Then, latency and transactional throughput were observed on PoUW and PoUWAS in Python Spyder from Figure 7 and Figure 8. These platforms are essential to deal with blockchain-related end-to-end transactions. Real-time applications, such as supply chain and healthcare-related frameworks, are also executable. Table 4 represents the recent outcomes related to improved consensus.

In this section, this research mainly explores the scalability metrics regarding computational energy from Table 5, Figure 5, Figure 6, notations, latency, and transaction throughput from Table 3, Table 5, Table 6, Table 7, Figure 7, and Figure 8. The primary motive for this section is to present the PoUW, PoA, PoS, and PoUWAS computational energy details, as well as the comparison between PoUW and PoUWAS latency and transaction throughput details. In Figure 6, this research explores individual algorithm-related gas, transactions, and execution costs and the observed deductions in computational energy.

Table 4
Findings related to related work

Author & Year	Method / Framework	Focus Area	Limitations	Outcome/ Performance
(Pawar, 2025)	Resource-efficient PoUW	Optimising computational resources	Lack of real-time adaptability	Low energy but static scheduling
(Rahman, 2024)	PoUW for scientific computation	Reusing mining power for real computations	Latency in task validation	High utility, moderate delay
(Babaei et al., 2025)	Integrates decision-making techniques	Blockchain on Supply Chain with ML techniques	Scalability	Traceability and security
(Kumari & Kumari, 2025)	integrated PoA+PBFT	Improved Consensus	Limited Scalability metrics	Improved efficiency
Proposed Model (2025)	PoUWAS	Real-time scalable consensus	Under testing for large-scale deployment	Improved scalability, low energy footprint, and real-time DApp suitability

Table 5
Parameters related to scalability

Algorithm	Gas Cost	Transaction Cost	Execution Cost	Total Cost
PoUW (Daely et al., 2024)	1232884	1072403	938293	3243580
PoUWAS	756234	657766	563478	1977478

Table 6
Notations related to the algorithm

Notation	Meaning
L	Latency
Tt	Transactional Throughput
N	Number of Block 0 to Block 9
PoUW	Proof of Useful Work
PoUWAS	Proof of Useful Work Authority Storage
ET	End Time
ST	Start Time

Table 7
Parameters related to the scalability of PoUWAS

Block	Latency	Transaction Throughput
Block 0	0.113	8.84
Block 1	0.063	31.59
Block 2	0.015	190.37
Block 3	0.238	16.76
Block 4	0.097	51.5
Block 5	0.045	131.56
Block 6	0.031	219.31
Block 7	0.319	25.02
Block 8	0.02	432.7
Block 9	0.026	376.04
	Total=0.967	Total=1483.69



Figure 5. Computational energy in terms of gas cost, transactional cost, and execution cost: (a) existing PoUW; (b) Proposed PoUWAS

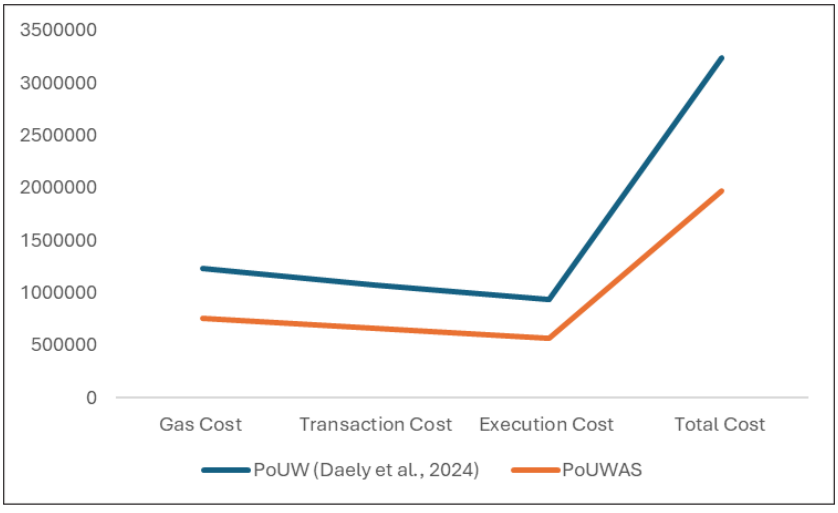


Figure 6. PoUWAS, PoUW computational energy



Figure 7. Latency, and transactional throughput: (a), (b), (c) Proposed PoUWAS; (d), (e) existing PoUW

Computational Energy

In blockchain communities, "gas" quantifies the processing needed to perform tasks like the execution of smart contracts. This metric determines the computational resources allocated to transaction verification. On platforms like Ethereum, participants use the network's inherent digital tokens to facilitate operations. The entire amount of cryptocurrency that a user pays for a transaction to be executed on the blockchain is known as the transaction cost. It covers the price of gas as well as any other charges levied by the network or service provider. The costs incurred in carrying out a particular operation or job inside a blockchain-based smart contract are called execution costs. For smart contracts, every individual

function or instruction carries its own computational charge. Since these contracts are made up of multiple steps, the total gas required is shaped by the sum of all these execution costs, as noted in 1, 2, 3 and 4. As a result, the combined computational demands will ultimately determine the final amount a user pays to complete the transaction.

Table 5 compares the scalability parameters for individuals and proposed consensus algorithms. The proposed work aims to enhance the improved consensus algorithm compared to the existing PoUW. The improved consensus algorithm combines three consensus algorithms, which are PoUW (Dong et al., 2019), PoA (Hanggoro et al., 2024), and PoS (Heo et al., 2024). In the observation, deductions in terms of gas cost, transaction cost, and execution cost for PoUW and PoUWAS were noted at 38.66%, 38.66%, and 39.94%, respectively. Therefore, deductions in these parameters lead to improvements in the performance of a blockchain. Figure 8 also represents the comparison between all the algorithms mentioned.

From Table 5, Table 8, Figure 7, and Figure 8, Proof-of-Useful-Work (PoUW) is examined using the proposed methodology-based consensus algorithm, specifically PoUWAS, which analyses the gas cost, transaction cost, and execution cost. Ultimately, it was concluded from this observation that combining PoUW with PoA and PoS, i.e., PoUWAS, results in minimal computational energy compared to individual PoUW, and it also focuses on the authorisation and storage availability of a node.

Table 8
Parameters related to the scalability of PoUW

Block	Latency	Transaction Throughput
Block 0	0.031	31.45
Block 1	0.268	7.43
Block 2	0.371	8.08
Block 3	0.043	92.2
Block 4	0.256	19.5
Block 5	0.096	62.44
Block 6	0.063	109.81
Block 7	0.239	33.37
Block 8	0.053	169.27
Block 9	0.029	335.71
	Total=1.449	Total=869.26

Latency

The duration from when a transaction is initiated to fully confirmed and recorded on the blockchain is known as latency in blockchain systems. For instance, blockchains such as Bitcoin that use the Proof of Work (PoW) system have greater latency because it takes

approximately ten minutes to include a new block in the chain. Blockchains that use Proof of Stake (PoS) or other newer consensus algorithms can achieve faster confirmation of transactions by lessening the time it takes for a transaction to be included in a block and confirmed.

Transaction Throughput

Transaction throughput in terms of transactions per second on blockchain is the quantity of transactions that the network can process in each time. Scalability and the ability to keep up with the demands of blockchain applications require higher throughput. The throughput of a blockchain is determined by several factors, such as the size of the individual blocks, the frequency of new block generation, and the consensus mechanism employed by the network.

Latency and Transaction Throughput for PoUWAS

To calculate the blockchain latency and transactional throughput, one would have to sum up the latency and throughput of each block. For example, latency (L) and transaction throughput (Tt). Table 6 summarises the different notations and their associated meanings that will be used in the following mathematical equations 5, 6, 7, 8, 9, 10 and 11.

According to the proposed methodology, latency and transactional throughput must be calculated to measure the latency and transactional throughput of the proposed consensus algorithm. Blockwise calculation helps evaluate the overall performance of an algorithm using various metrics, such as latency and transactional throughput, as shown in Table 7. Proposed work research mainly focuses on the latency and transactional throughput for the 10 blocks. Measuring each block's latency and transactional throughput starts from block 0 to block nine and yields values of 0.967 and 1483.69. Here, the latency for 10 blocks is L1, and the overall transaction throughput for 10 blocks is Tt1.

According to the proposed methodology, as observed in Table 7, Figure 7, and the existing consensus algorithm presented in Table 8, Figure 8, latency and transactional throughput were measured. After comparing the proposed methodology with existing methods, a change in improvement was observed in various metrics, including latency, which was reduced by 0.482, and transactional throughput, which improved by 614.43 TPS. The proposed work has reduced latency compared to existing work and increased transactional throughput. The existing consensus algorithm from Table 5 was used to observe the overall latency and transactional throughput for 10 blocks, and its performance was compared with that of the proposed algorithm methodology from Table 8. Overall performance was also compared regarding latency and transaction throughput, and the proposed method achieved better results.

Latency and Transaction Throughput for PoUW

The experimental findings show that, in comparison to conventional Proof of Work (PoW) and Proof of Stake (PoS) processes, there is a notable decrease in block validation time and an increase in throughput. Previous research has shown that hybrid consensus algorithms that combine voting and reputation measures can improve network scalability. (Kumari & Kumari, 2025; Li et al., 2020) demonstrated similar results. These investigations are supported by the current results, which further demonstrate that adaptive or hybrid consensus techniques can maximise resource efficiency and sustain increased transaction throughput without sacrificing security. Financial performance and computational energy were accomplished within the supply chain simultaneously (Goli & Babaee Tirkolaee, 2023). Flexibility, efficiency, and coordination among blockchain applications in supply chains were examined in this study (Haffar & Ozceylan, 2024). A viable supply chain model by integrating a vendor-managed inventory approach and blockchain technology to reduce costs is proposed (Lotfi et al., 2022).

CONCLUSION

This research is unique in that it proposes intelligent decision-making as part of the consensus layer of the blockchain, which has not been explored extensively in the literature. In contrast to previously presented solutions, where merely cryptographic or structural aspects are adjusted, the proposed approach presents a hybrid consensus protocol capable of adapting the participation and validation process at runtime. This solution, besides being more scalable and energy-efficient, is also more trust-adaptable, which is a step forward towards a more sustainable design of a blockchain system that can be used in several fields.

This paper proposes work that mainly concentrates on measuring the improved consensus algorithms with the help of metrics related to scalability in blockchain, as well as comparing the differences in scalability between existing work and the proposed work. Mainly concentrates on transaction throughput, latency, computational energy, and consensus models. After comparing the existing algorithm, proof of useful work with the improved consensus algorithm Proof of useful work, authorisation and storage availability scalability improved regarding latency and transaction throughput in the proposed work. Still, there is considerable scope to improve the overall performance of the blockchain network by applying additional techniques. Future work will focus on this area and integrate with other emerging technologies to utilise blockchain more effectively in real-time applications. Here, the research scope is limited to measuring the proposed consensus algorithm. However, future research will continue with additional metrics related to scalability. After a careful review of the literature on consensus algorithms in blockchain technology, scalability is the primary issue hindering the overall performance of

a blockchain. Consensus algorithm improvements are also essential to validate transactions most effectively. Latency and transaction throughput improvements were also crucial for enhancing the overall growth of a blockchain.

ACKNOWLEDGEMENT

The authors thanked Koneru Lakshmaiah Education Foundation, especially the Computer Science and Engineering Department, which provided all the facilities that supported this research. This article is part of the first author's thesis. No funding or grants were provided for this research.

LIST OF ABBREVIATIONS

PoW	:	Proof of Work
PoA	:	Proof of Authority
PoS	:	Proof of Storage
PoUWAS	:	Proof of Useful Work-Authorisation-Storage Availability
DPoS	:	Delegated Proof of Stake
PBFT	:	Practical Byzantine Fault Tolerance

REFERENCES

Andrew, J., Isravel, D. P., Sagayam, K. M., Bhushan, B., Sei, Y., & Eunice, J. (2023). Blockchain for healthcare systems: Architecture, security challenges, trends and future directions. *Journal of Network and Computer Applications, 215*, Article 103633. https://doi.org/10.1016/j.jnca.2023.103633

Babaei, A., Tirkolaee, E. B., Ali, S. S., Roy, S. K., & Weber, G. W. (2025). An efficiency-uncertainty-based consensus optimisation framework with dual-perspective clustering analysis for blockchain adoption in supply chains. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2025.3626659

Daely, P. T., Lee, J. M., & Kim, D. S. (2024). Vehicle routing based mining for proof-of-useful-work blockchain. *Journal of the Korean Institute of Communications and Information Sciences, 49*(1), 124-133. https://doi.org/10.7840/kics.2024.49.1.124

Deng, Z., Li, T., Tang, C., He, D., & Zheng, Z. (2024). PSSC: Practical and secure sidechain construction for heterogeneous blockchains orienting IoT. *IEEE Internet of Things Journal, 11*(3), 4600-4613. https://doi.org/10.1109/JIOT.2023.3302291

Dong, Z., Lee, Y. C., & Zomaya, A. Y. (2019). Proofware: Proof of useful work blockchain consensus protocol for decentralised applications. *arXiv*. https://doi.org/10.48550/arXiv.1903.09276

Fajri, A. I., & Mahananto, F. (2022). Hybrid lightning protocol: An approach for blockchain scalability issue. *Procedia Computer Science, 197*, 437-444. https://doi.org/10.1016/j.procs.2021.12.159

Ghamdi, M. A. A. (2022). An optimised and secure energy-efficient blockchain-based framework in IoT. *IEEE Access, 10*, 133682-133697. https://doi.org/10.1109/ACCESS.2022.3230985

Goli, A., & Babaee Tirkolaee, E. (2023). Designing a portfolio-based closed-loop supply chain network for dairy products with a financial approach: Accelerated Benders decomposition algorithm. *Computers & Operations Research*, Article 106244. https://doi.org/10.1016/j.cor.2023.106244

Haffar, S., & Ozceylan, E. (2024). Blockchain in resilient and leagile supply chains: Research themes and opportunities. *Journal of Industrial and Management Optimisation*. https://doi.org/10.3934/jimo.2024109

Hanggoro, D., Windiatmaja, J. H., Muis, A., Sari, R. F., & Pournaras, E. (2024). Energy-aware proof-of-authority: Blockchain consensus for clustered wireless sensor network. *Blockchain: Research and Applications, 5*(3), Article 100211. https://doi.org/10.1016/j.bcra.2024.100211

Heo, J. W., Ramachandran, G., & Jurdak, R. (2024). nPPoS: Non-interactive practical proof-of-storage for blockchain. *Blockchain: Research and Applications*, Article 100221. https://doi.org/10.1016/j.bcra.2024.100221

Kaur, M., Khan, M. Z., Gupta, S., Noorwali, A., Chakraborty, C., & Pani, S. K. (2021). MBCP: Performance analysis of large-scale mainstream blockchain consensus protocols. *IEEE Access, 9*, 80931-80944. https://doi.org/10.1109/ACCESS.2021.3085187

Kumari, K. L., & Kumari, P. L. S. (2025). Design and analysis of the improved consensus algorithm of the blockchain technology. *International Research Journal of Multidisciplinary Scope, 6*(2), 833-844.https://doi.org/10.47857/irjms.2025.v06i02.03506

Liu, Y., Qian, K., Wang, K., & He, L. (2021). Effective scaling of blockchain beyond consensus innovations and Moore's law: Challenges and opportunities. *IEEE Systems Journal*. https://doi.org/10.1109/JSYST.2021.3087798

Lotfi, R., Rajabzadeh, M., Zamani, A., & Rajabi, M. S. (2022). Viable supply chain with vendor-managed inventory approach by considering blockchain, risk and robustness. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-05119-y

Malakhov, I., Marin, A., & Rossi, S. (2023). Analysis of the confirmation time in proof-of-work blockchains. *Future Generation Computer Systems, 147*, 275-291. https://doi.org/10.1016/j.future.2023.04.016

Manolache, M. A., Manolache, S., & Tapus, N. (2022). Decision making using the blockchain proof of authority consensus. *Procedia Computer Science, 199*, 580-588. https://doi.org/10.1016/j.procs.2022.01.071

Nasir, N. M., Hassan, S., & Zaini, K. M. (2024). Securing permissioned blockchain-based systems: An analysis on the significance of consensus mechanisms. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3465869

Otsuki, K., Nakamura, R., & Shudo, K. (2021). Impact of saving attacks on blockchain consensus. *IEEE Access, 9*, 133011-133022. https://doi.org/10.1109/ACCESS.2021.3115131

Pawar, T. (2025). Cyber-shield machine learning: An intrusion detection system with blockchain-powered secure log storage across network nodes. *Sigma Journal of Engineering and Natural Sciences*, 714-725. https://doi.org/10.14744/sigma.2025.00066

Rahman, R. U. (2024). A novel machine learning-based artificial intelligence approach for log analysis using blockchain technology. *Sigma Journal of Engineering and Natural Sciences*, 1391-1409. https://doi.org/10.14744/sigma.2024.00004

Sun, Y., Yan, B., Yao, Y., & Yu, J. (2021). DT-DPoS: A delegated proof of stake consensus algorithm with dynamic trust. *Procedia Computer Science, 187*, 371-376. https://doi.org/10.1016/j.procs.2021.04.113

Tawfik, A. M., Al-Ahwal, A., Eldien, A. S. T., & Zayed, H. H. (2025). Blockchain-based access control and privacy preservation in healthcare: A comprehensive survey. *Cluster Computing, 28*(8). https://doi.org/10.1007/s10586-025-05308-x

Ucbas, Y., Eleyan, A., Hammoudeh, M., & Alohaly, M. (2023). Performance and scalability analysis of Ethereum and Hyperledger Fabric. *IEEE Access, 11*, 67156-67167. https://doi.org/10.1109/ACCESS.2023.3291618

Wu, A., & Hua, C. (2025). ALB-chain: Achieving scalability and load balance blockchain sharding. *Cluster Computing, 28*(6). https://doi.org/10.1007/s10586-024-04949-8

Yang, J., Jia, Z., Su, R., Wu, X., & Qin, J. (2022). Improved fault-tolerant consensus based on the PBFT algorithm. *IEEE Access, 10*, 30274-30283. https://doi.org/10.1109/ACCESS.2022.3153701

Zhang, J., Tian, R., Cao, Y., Yuan, X., Yu, Z., Yan, X., & Zhang, X. (2021). A hybrid model for central bank digital currency based on blockchain. *IEEE Access, 9*, 53589-53601. https://doi.org/10.1109/ACCESS.2021.3071033

Zhang, L., Li, F., Wang, P., Su, R., & Chi, Z. (2022). A blockchain-assisted massive IoT data collection intelligent framework. *IEEE Internet of Things Journal, 9*(16), 14708-14722. https://doi.org/10.1109/JIOT.2021.3049674

Pertanika J. Sci. & Technol. 34 (2): 695 - 720 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Association Rule Mining to Detect Significant Factors of Autism Spectrum Disorder at Early Stage

Tania Akter[1,2], Farzana Tasnim[2], Shahriar Hassan[2], Mohammad Hanif Ali[1], and Mohammad Shorif Uddin[1]*

[1]Department of Computer Science and Engineering, Jahangirnagar University, 1342 Savar, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Gono Bishwabidyalay, 1344 Savar, Dhaka, Bangladesh

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex, long-lasting neurodevelopmental condition characterised by communication difficulties. However, early identification of ASD may aid in the design of appropriate treatments to improve communicative development. This study employs association rule mining algorithms, such as apriori, predictive apriori, and Tertius, to identify age-specific behavioural markers across four developmental stages: toddlers, children, adolescents, and adults. These algorithms identified several key behavioural indicators associated with ASD, including pretending games (A5), activity switching (A4), difficulty in making friends (A10), difficulty in conversations (A5), detection of listener boredom (A6) and easy reading of emotions (A9). The apriori algorithm achieved the highest confidence of 100.00% for the adolescent dataset, whereas the predictive apriori algorithm demonstrated the highest accuracy of 99.50% for the toddler dataset, 99.40% for the child dataset and 99.39% accuracy for adult dataset. In contrast, the Tertius algorithm showed the highest confirmation rate of 66.8% for the child dataset, although it required the most time for manipulation across all datasets. Our findings demonstrate that rule mining effectively uncovers clinically relevant patterns, with behavioural questions proving more significant than demographic factors, such as gender or birth complications. Furthermore, no connection was observed between ASD and a history of jaundice at birth. Performance comparisons among the three algorithms are also presented. The rules generated through our investigation will help physicians in the early detection of ASD, thus paving the way for a timely and targeted intervention.

Keywords: AQ-10 screening dataset, association rule mining, autism spectrum disorder (ASD), Q-Chat-10 toddler

ARTICLE INFO

Article history:
Received: 13 September 2024
Accepted: 15 December 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.05

E-mail addresses:
shilpy339cse@gmail.com (Tania Akter)
rekha.tasnim@gmail.com (Farzana Tasnim)
shahriar.hassan303@gmail.com (Shahriar Hassan)
hanif_ju03@juniv.edu (Mohammad Hanif Ali)
shorifuddin@juniv.edu (Mohammad Shorif Uddin)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by a deficiency of social behaviour and interaction (Mazumdar et al., 2021; Rubio-Martín et al., 2024; Uddin et al., 2023). Currently, the number of people with ASD is rapidly increasing, with almost 1 in every 68 children diagnosed with it. Therefore, approximately 1.5% of the entire population suffers from ASD (Towle & Patrick, 2016; Wingate et al., 2014), which has become a concern. Although Autism Spectrum Disorder (ASD) cannot be completely cured, research has shown that early intervention can substantially improve developmental progress and long-term outcomes. However, obtaining a timely diagnosis remains a significant challenge due to the diverse range of symptoms and the dependence on subjective behavioural assessments. Structured tools for monitoring progress, such as screening questionnaires, play a vital role in identifying early risk indicators. However, the complexity of multidimensional data analysis often impedes the generation of actionable insights.

Association Rule Mining (ARM) techniques (Agrawal et al., 1993; Kaoura et al., 2024) are used to extract significant correlations, frequent patterns, associations, or causal structures between all items in datasets (Zhao & Bhowmick, 2003). By analysing behavioural and demographic variables in the ASD screening data, the ARM can recognise combinations of risk factors that are strongly related to ASD diagnosis, allowing for more precise and objective early detection. Despite its potential, ARM remains underexplored in ASD research compared to conventional machine learning approaches.

The application of association rule mining (ARM) methods to the diagnosis of autism spectrum disorder (ASD) has recently received growing attention (Singh & Mantri, 2024). Several studies have investigated the use of ARM to detect patterns related to autism characteristics and probable risk factors. For example, Thabtah and Peebles (2020) proposed a new rule-based machine learning method that detected autism traits and extracted significant rules. In addition, Al-Diabat (2018) evaluated a fuzzy rule-based data mining method for forecasting autistic symptoms in children that can be automatically extracted from previous instances to form a screening classification system. Wiratsin and Narupiyakul (2021) proposed a novel feature selection method to identify important relationships between ASD symptoms in three different age groups. Engle and Rada (2011, 2017) used association rule mining techniques to generate rules. Engle (2013) also proposed a novel approach to combine domain knowledge with association rule mining methods. In addition to association rule mining, many researchers have also worked using various machine learning algorithms such as Random Forest, Support Vector Machine, Gradient Boost, Neural Networks, etc. to the mentioned datasets. In their work, they showed high accuracy to predict autism, ranging from approximately 85% - 98% in these datasets (Akter et al., 2021; Uddin et al., 2023). But they could not show which behaviours were especially

responsible for detecting autism. But in the case of association rules, these algorithms identify important specific behavioural markers and explained relationship among them. For small datasets, association rule mining is better than machine learning in terms of interpretability and feature discovery.

In previous work, researchers have used ARM in the child, adolescent, and adult data sets or used a single cohort to find key factors related to ASD, but no one has used these techniques in the Q-Chat-10 toddler data set yet. In our study, we conducted a comparative and multicohort study of four developmental age stages, such as toddler, child, adolescent, and adult, in which the same parameters of association rule mining were utilised. This analysis provides a broad perspective on behavioural development and the importance of these factors.

While existing research has made contributions, there is a scarcity of studies that specifically utilise association rule mining for ASD diagnosis. This study seeks to fill this research gap by applying association rule mining techniques to analyse various ASD datasets, including the Q-Chat-10 toddler, AQ-10 child, AQ-10 adolescent, and AQ-10 adult screening datasets. Through the implementation of ARM methodologies, we derived rules that differentiate ASD from non-ASD conditions, aiming to identify the most effective strategies for early detection. This study is anticipated to offer crucial insights for medical professionals and caregivers, enabling simpler and more precise identification of ASD.

The vital contributions of our work are mentioned below:

- Various risk factors of ASD and non-ASD individuals are identified using association rule mining algorithms.
- Associations among different risk factors that indicate ASD have been mined using different association rule mining algorithms.
- The performance of different association rule mining algorithms is compared.

The rest of the paper goes through the following organised. Section II presents the details of the datasets used and the methodology of this work. In Section III, the extracted rules are presented, and the performance analysis is provided. A thorough discussion of the findings is provided in Section IV. Lastly, Section V concludes the work.

MATERIALS AND METHODS

In this study, we employed toddlers, children, adolescents, and adults to analyse ASD using the association rule mining technique. Next, we pre-processed these datasets and applied apriori, predictive apriori, and tertius algorithms to uncover relationships and patterns within the data. The rules identified through this analysis were further explored to determine the most significant rules that could assist in early diagnosis and interventions. To detect risk factors for ASD, the basic pipeline of this study is shown in Figure 1.

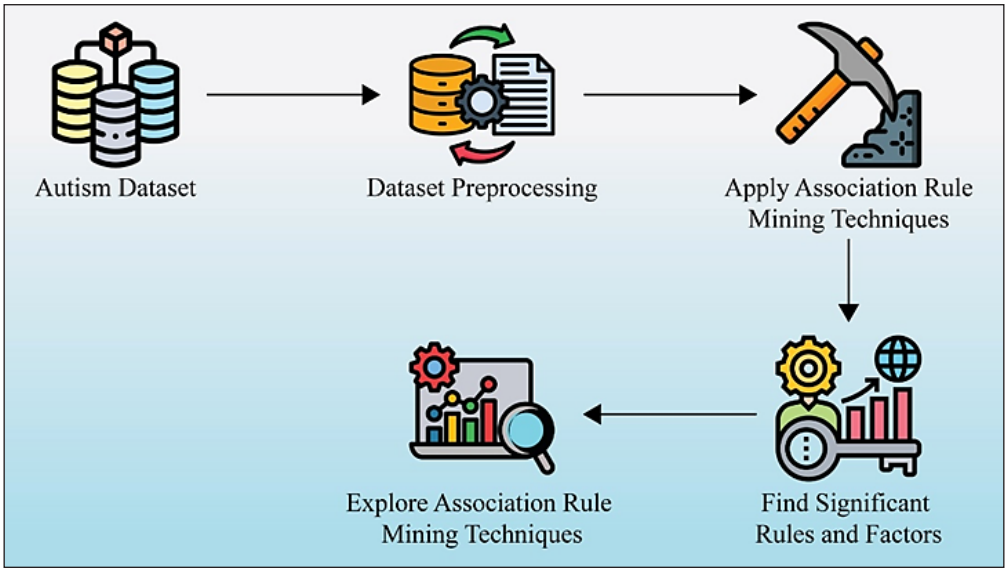


Figure 1. Association rule mining framework for identifying key factors in autism spectrum disorder

Dataset Description

To detect significant factors of autism, we gathered Q-Chat 10 toddler (Delja, n.d.), AQ-10 child, AQ-10 adolescent, and AQ-10 adult datasets from Kaggle and UCI data repositories, respectively, which are created based on AQ-10 screening tool questionnaires (UCI Machine Learning Repository, n.d.-a, -b, -c). These data have been collected using a mobile application called ASDTests. All datasets contained 20 features, including class labels. This label contains two values, one is ”0”, which indicates the individuals with no ASD traits, and ”1”, which indicates that they contain ASD traits. The toddlers screening dataset contains 1054 cases in the age group 18-36 months where 30.26% female and 69.73% male, the child screening dataset contains 292 cases in the age group 4-11 years where 28.77% female and 71.23% male, the adolescent screening dataset contains 104 cases with the age group 12-16 years where 51.92% female and 48.08% male and adult screening dataset contains 704 cases with the age group above 18 years where 47.87% female and 52.13% male. Thus, the toddler and child datasets are not balanced, but the adolescent and adult datasets are balanced (Figure 2).

Figure 3 illustrates the distribution of ASD cases based on the presence of jaundice at birth across all four datasets. The number of cases with jaundice was lower than that without jaundice, suggesting that there was no strong association between jaundice at birth and ASD occurrence.

Figure 4 shows the distribution of ASD cases based on family ASD across all four datasets. In all datasets, the number of cases without a family history of ASD was greater

than that of cases with a family history. This suggests that there is no strong association between family ASD and the occurrence of ASD. The description of the dataset is shown in Table 1.

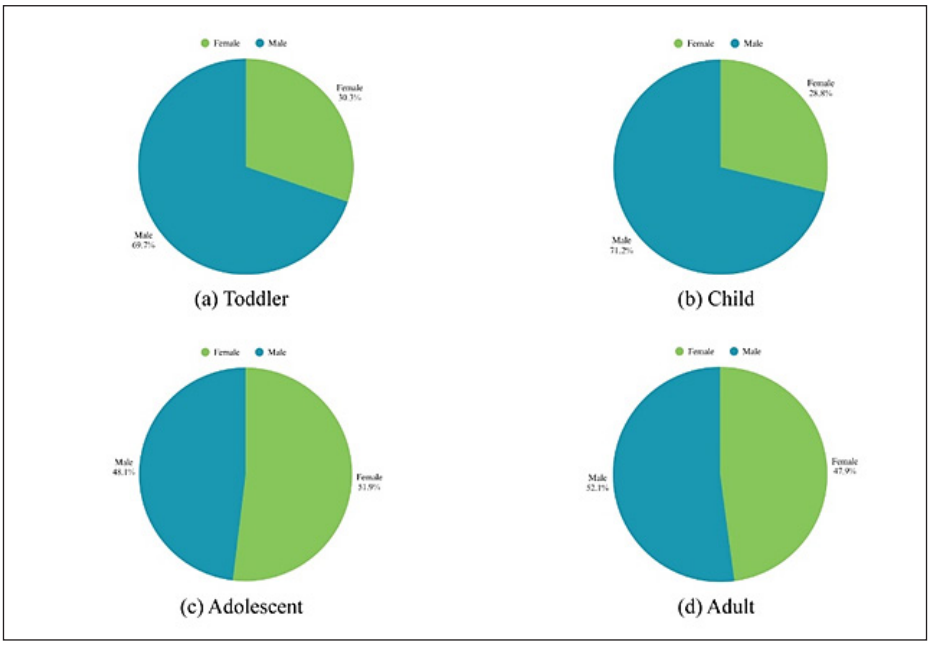


Figure 2. Gender based distribution of ASD cases

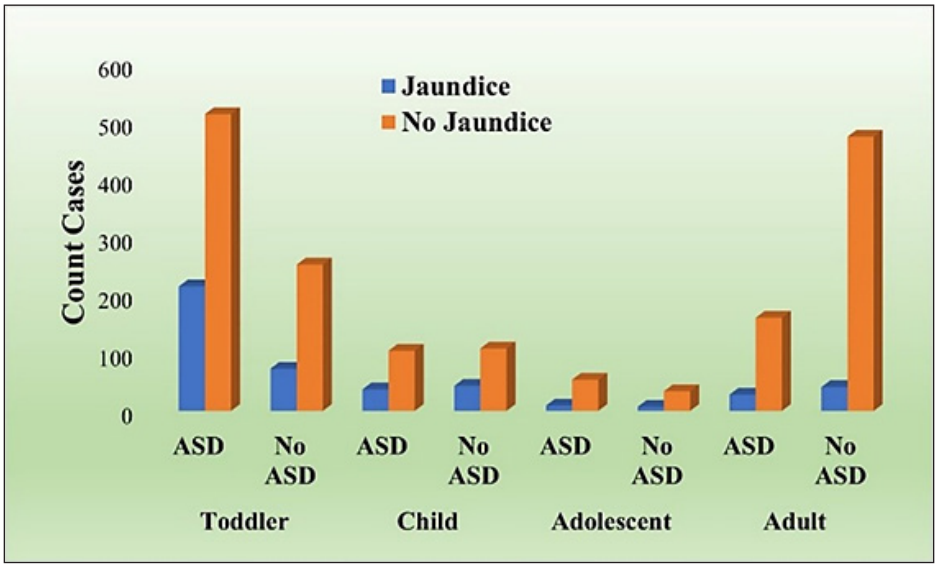


Figure 3. Association between jaundice at birth and autism spectrum disorder

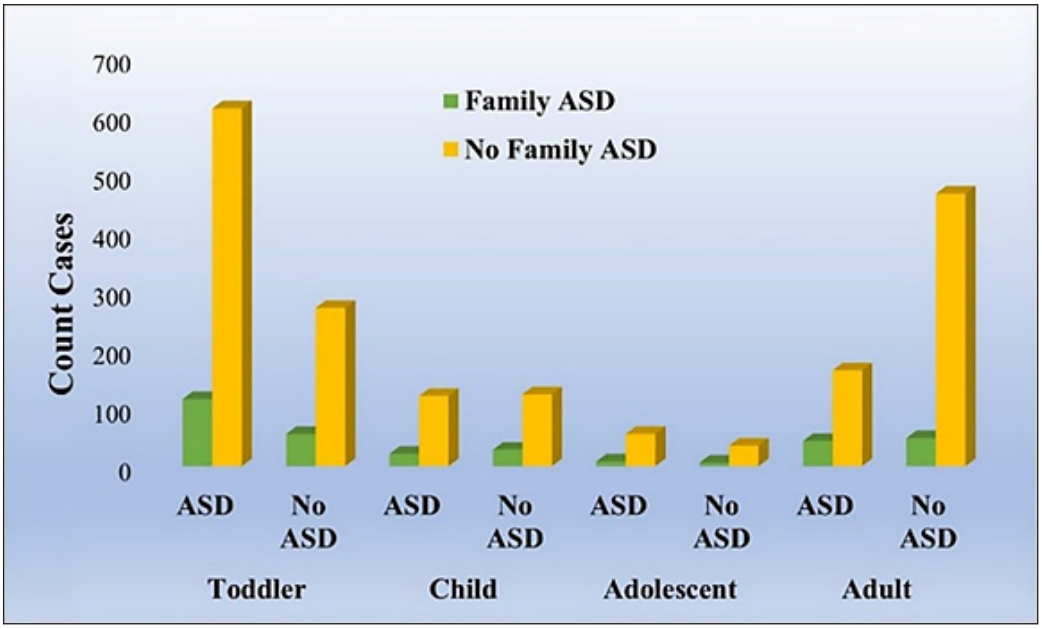


Figure 4. Impact of family ASD on offspring's autism spectrum disorder

Table 1
Features description

Feature	Type	Description
Age	Number	toddlers (month), children, adolescents, and adults(year)
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean	Whether the case was born with jaundice
Family member with ASD	Boolean	Whether any immediate family member has an ASD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician, etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean	Whether the user has used a screening app
Screening Method Type	Integer	The type of screening methods chosen is based on the age category
A1	Binary	The response to the question depends on the screening method employed.
A2	Binary	The response to the question depends on the screening method employed.

Table 1 (*continued*)

Feature	Type	Description
A3	Binary	The response to the question depends on the screening method employed.
A4	Binary	The response to the question depends on the screening method employed.
A5	Binary	The response to the question depends on the screening method employed.
A6	Binary	The response to the question depends on the screening method employed.
A7	Binary	The response to the question depends on the screening method employed.
A8	Binary	The response to the question depends on the screening method employed.
A9	Binary	The response to the question depends on the screening method employed.
A10	Binary	The response to the question depends on the screening method employed.
Scoring Result	Integer	The response to the question depends on the screening method employed.
ASD	Boolean	toddlers, children, adolescent or adults diagnosed with ASD

Data Preprocessing

After collecting the primary datasets, we found that they contained several missing values. If the data are numeric, the missing value is generally replaced by the mean/median. In our dataset, the nominal attributes contained missing values. Therefore, we replace it with a constant (Filter: ReplaceMissing WithUserConstant). This filter works as follows: if any nominal attribute contains two classes and class one contains 90 per cent of the data, then this filter replaces the missing value with the constant value of class one. We also converted the numerical data into nominal values. The correlation method was then used to eliminate highly correlated features from these datasets (Akter et al., 2021). Therefore, QChat10 Score is eliminated from the toddlers' dataset, age desc, and results are removed from the child and adolescent datasets, and the result attribute is reduced from the adult dataset.

Association Rule Mining

In this study, association rule mining was used to analyse the data from autism spectrum disorder. Every association rule mining method has a basic structure to fulfil the minimum support and confidence (Ishibuchi et al., 2007). To extract significant features of autism more accurately, different association rule mining algorithms, such as apriori, predictive apriori, and tertius, were employed in all autism datasets in this study. These algorithms have been extensively used to analyse the significant factors of different datasets (Nahar et al., 2013; Niranjan et al., 2016).

Additionally, these algorithms have their complementary analytical properties. The apriori algorithm has high efficiency to identify confident frequent patterns, the predictive apriori has the capability to use probabilistic optimisation for improved accuracy, and the tertius has capability to extract confirmatory and relational rules. By using them together, we can get a balanced assessment through confidence, accuracy, and confirmation metrics.

Apriori Algorithm

The apriori algorithm (Cai, 2020; Maitra et al., 2021) is an association rule mining technique that employs frequent itemsets to create association rules. This algorithm calculates the level of relationship among objects using the association rule and works well on transactional databases. To estimate the associations of items proficiently, a breadth-first search and a hash tree are applied. In this algorithm, the support of the itemset is first calculated, and the minimum support and confidence are fixed. Next, all item sets whose support values are greater than the minimum support value is selected. Finally, rules for subsets with higher confidence values than the minimum confidence value are created. This algorithm is easy to understand, and a complex formula does not need to be derived. However, it works gradually, and its time and space complexities are high. In the apriori algorithm, the rules are ranked by the confidence, like the accuracy value of the rules (Mutter et al., 2004; Wang & Guo, 2010).

Predictive Apriori Algorithm

The predictive apriori algorithm, first introduced by Scheffer (2005), builds upon the traditional apriori algorithm. Unlike apriori, which generates association rules based solely on confidence (Aher & Lobo, 2012), predictive apriori considers both support and confidence to rank rules more effectively. Within a Bayesian framework, it estimates the expected accuracy of a rule using its support and confidence, thereby enhancing its ability to generalise the results. When applied to a database D with a set of records r and a static process s, the predictive accuracy of the rule: I >= J is the conditional probability of J being a subset of r, where I is a subset of r (Nahar et al., 2013).

Tertius Algorithm

Tertius is an algorithm that finds the rule according to the measurement of confirmation and a top-down rule-finding system using first-order logic representation to deal with structured, relational information (Flach & Lachiche, 2001; Hareendran & Chandra, 2017). Several confirmation measures have been examined, with the simplest being weighted relative accuracy. A confirmation measure reveals the degree to which a rule is unexpected, as well as the percentage of expected counter-instances. The algorithm computes two values: the expected probability and the observed probability (Nahar et al., 2013).

Evaluation Metrics for Association Rule Mining

In this work, we justified the corresponding rules by considering some statistical evaluation metrics, such as confidence (Kane, 2018; Yan et al., 2009), accuracy, confirmation, and support (Hamid et al., 2024) of ASD and non-ASD. This analysis reflects the usefulness and certainty of a rule using these metrics.

When we consider a rule $(X \Rightarrow Y)$,

$$Confidence\ (X \Rightarrow Y) = P(Y|X) = \frac{(Support\ (X \cap Y))}{(Support\ (X))} \quad [1]$$

$$Support\ (X \Rightarrow Y) = P(X \cap Y) \quad [2]$$

where $P(X)$ is defined as the probability of randomly selected transactions of the dataset including all items in X and $X \cap Y$ is indicated as all transactions which contained both X and Y (Hasanpour et al., 2019). On the other hand, support and confidence are used as a single measure called accuracy, which is denoted as follows:

$$Accuracy = \{Support, Confidence\} \quad [3]$$

Then, confirmation is used to measure the novelty of the associated rules. Calculating results requires expected and observed frequency values.

$$Confirmation = \frac{expected - observed}{\sqrt{expected - expected}} \quad [4]$$

To make sure the reproducibility of our results, we applied a consistent minimum support and confidence threshold to all age groups. In the case of apriori algorithm, the minimum support is set at 0.05, and the confidence threshold is set at 0.90. In the case of apriori predictive algorithm, the minimum support is set at 0.04, and the confidence threshold is set at 0.85. For the tertius algorithm, the confirmation value is set at 0.75. Such threshold values were chosen to balance the quality of the rules and the number of rules in all the datasets.

RESULTS

We used the three most prevalent association rule mining algorithms, named apriori, predictive apriori, and tertius, to conduct this study using Weka 3.8 (Aksu & Dogan, 2019).

These algorithms generate several rules, but we only considered those containing class attributes on the right-hand side or left-hand side. Each rule contained a set of features linked to the ASD outcome, accompanied by confidence accuracy, and confirmation scores representing the rule's reliability. The most significant rules were reported based on confidence, accuracy, and confirmation. Tables 2 to 13 represent the results of the experiment. In all results tables, the symbol "∩" denotes logical AND (intersection), while "∪" denotes logical OR (union) between conditions in association rules. (A1- A9) are personal behavioural questions on the ASD screening data. The score of 1 represents the presence of the behaviour, whereas 0 represents the absence of the behaviour.

Results for Identifying ASD in Toddlers

Table 2 illustrates the association rules generated by the apriori algorithm utilised for a dataset of toddlers diagnosed with ASD. Each rule contained a set of features linked to the ASD outcome, accompanied by confidence scores representing the rule's reliability. For instance, Rule 1 (A5 Score=1 ∩ A6 Score=1 → ASD=Yes) has a confidence of 0.990, indicating that 99.00% of the time, when a toddler scores 1 on both A5 and A6, they tend to have ASD. All rules showed the confidence value above 0.920, reflecting robust relationships with ASD. The analysis reveals robust correlations among several attributes, including A1, A2, A5, A6, A7, and A9. This algorithm gave no such rules for healthy toddlers.

Table 3 presents association rules derived from the predictive apriori algorithm, revealing patterns related to ASD diagnosis in toddlers. These rules demonstrate high accuracy, ranging from 0.9940 to 0.9950, indicating strong predictive capability. For example, Rule 1 states that if A4, A5, and A9 scores are all 1, there is a 99.50% probability

Table 2
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Confidence
1	A5_Score=1 ∩ A6_Score=1 → ASD=Yes	0.990
2	A5_Score=1 ∩ A7_Score=1 → ASD=Yes	0.980
3	A1_Score=1 ∩ A7_Score=1 → ASD=Yes	0.970
4	A1_Score=1 ∩ A6_Score=1 → ASD=Yes	0.970
5	A6_Score=1 ∩ A7_Score=1 → ASD=Yes	0.960
6	A9_Score=1 → ASD=Yes	0.960
7	A5_Score=1 → ASD=Yes	0.940
8	A5_Score=1 ∩ autism=0 → ASD=Yes	0.930
9	A6_Score=1 ∩ autism=0 → ASD=Yes	0.930
10	A2_Score=1 → ASD=Yes	0.930

Table 3
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Accuracy
1	A4_Score=1∩A5_Score=1∩A9_Score=1→ASD=Yes	0.9950
2	A5_Score=1∩A9_Score=1∩Sex= male →ASD=Yes	0.9950
3	A1_Score=1∩A5_Score=1∩A9_Score=1→ASD=Yes	0.9950
4	A3_Score=1∩A4_Score=1∩A7_Score=1→ASD=Yes	0.9950
5	A1_Score=1∩A6_Score=1∩A10_Score=1→ASD=Yes	0.9950
6	A3_Score=1∩A4_Score=1∩A6_Score=1→ASD=Yes	0.9950
7	A1_Score=1∩A4_Score=1∩A5_Score= 1 → ASD =Yes	0.9950
8	A1_Score=1∩A5_Score=1∩A6_Score=1∩A7_ Score=1→ASD=Yes	0.9949
9	A2_Score=1∩A4_Score=1∩A7_Score=1→ASD=Yes	0.9949
10	A5_Score=1∩A6_Score=1∩A8_Score=1→ASD=Yes	0.9949
11	A5_Score=0∩A6_Score=0∩A9_Score=0→ASD=No	0.9946
12	A3_Score=0∩A5_Score=0→ASD=No	0.9944
13	A4_Score=0∩A5_Score=0→ASD=No	0.9943
14	A3_Score=0∩A4_Score=0∩A6_Score=0→ASD=No	0.9943
15	A5_Score=0∩A7_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.9940

of ASD being present. Rule 11, on the other hand, indicates that if A5, A6, and A9 scores are all 0, the probability of ASD being absent is 99.46%. These rules emphasise the critical attribute combinations that are strongly associated with ASD, providing valuable insights for early diagnosis and intervention.

Table 4 highlights the association rules obtained through the Tertius algorithm, which identifies additional patterns related to ASD in toddlers. The rules show that certain behavioural scores, like A5 and A9, in combination with factors such as age or ethnicity, are linked to ASD diagnosis. For instance, toddlers aged 16 or 26 months, or those from the Pacific ethnicity, who had A5 and A9 scores of 1, were more likely to be diagnosed with ASD, with confirmation scores ranging from 0.6043 to 0.6079. In contrast, toddlers with A2, A5, and A9 scores of 0 were less likely to have ASD. The highest confirmation score (0.6137) was associated with the rule A2_Score=0 ∩ A5_Score=0 ∩ A9_Score=0 → ASD=No, indicating a moderately reliable predictive association for the absence of ASD. Across all rules, confirmation scores ranged from 0.6042 to 0.6137, indicating moderate predictive strength. Comparing the three algorithms, it becomes evident that positive responses to pretending games (A5) and tracking vision (A6) are indicative of ASD presence, while negative responses to pretending games (A5) and simple gestures (A9) suggest the absence of ASD.

Table 4
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Confirmation
1	A5_Score=1∪A9_Score=1∪Age_Mons=26→ ASD=Yes	0.6079
2	A5_Score=1∪A9_Score=1 ∪ Ethnicity = Pacifica → ASD=Yes	0.6057
3	A5_Score=1∪A9_Score=1∪Age_Mons=16→ ASD=Yes	0.6043
4	A2_Score=0∩A5_Score=0∩A9_Score=No→ ASD=No	0.6137
5	A5_Score=0∩A9_Score=0→ASD=No	0.6071
6	A2_Score=0∩A9_Score=0→ASD=No	0.6055

Results for Identifying ASD in Children

Table 5 presents the association rules generated by the apriori algorithm to identify the patterns associated with ASD in children. The results indicate that specific combinations of characteristics, such as positive responses to A3, A4, A5, and A10 (Rule 1), show the highest confidence level (95.00%) in predicting ASD. Rules 2–10 further highlight other feature pairings with confidence levels of 92.00%–94.00%. Notably, the algorithm did not generate any rules for the healthy children.

Table 6 shows the association rules derived from the predictive apriori algorithm, which demonstrates a high level of predictive accuracy, illustrating the algorithm's effectiveness in distinguishing between ASD and non-ASD cases. Rule 1, with positive responses for A1, A3, A4, A5, and A10, offers the highest accuracy of 99.44% for predicting ASD. Rule 2, which includes A3, A4, A9, and A10, follows closely with an accuracy of 99.43%. Additionally, several rules accurately predicted non-ASD cases by identifying the absence of certain attributes, such as A1, A4, and A9 in Rule 3 (A1_Score=0 ∩ A4_Score=0 ∩ A9_Score=0 → ASD=No), which achieved an accuracy of 99.34%.

According to Table 7, the tertius algorithm identified A4 and A10 scores as significant factors for predicting ASD in children, with a confirmation rate of 66.86%. To predict absence of ASD, the key indicators included A4_Score=0, A10_Score=0, or the child's country of residence (e.g., United Arab Emirates), with a confirmation rate of 67.56%. These factors were consistently present across the generated association rules, underscoring their importance in distinguishing between ASD and non-ASD patients. When comparing the three algorithms, it becomes clear that a positive response for quick activity swapping (A4) and difficulty in making friends (A10) can signal the presence of ASD, while negative responses for these same factors can indicate the absence of ASD.

Table 5
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Confidence
1	A3_Score=1∩A4_Score=1∩A5_Score=1∩A10_ Score=1→ASD=Yes	0.950
2	A1_Score=1∩A4_Score=1∩A5_Score=1→ASD =Yes	0.940
3	A1_Score=1∩A4_Score=1∩A10_Score=1→ASD =Yes	0.940
4	A3_Score=1∩A4_Score=1∩A5_Score=1∩A6_ Score=1→ASD=Yes	0.930
5	A4_Score=1∩A5_Score=1∩A6_Score=1∩A10_ Score=1→ASD=Yes	0.930
6	A4_Score=1∩A5_Score=1∩A10_Score=1→ASD =Yes	0.920
7	A3_Score=1∩A4_Score=1∩A6_Score=1∩A10_ Score=1→ASD=Yes	0.920
8	A5_Score=1∩A9_Score=1∩A10_Score=1→ ASD=Yes	0.920
9	A3_Score=1∩A4_Score=1∩A10_Score=1→ ASD=Yes	0.920
10	A1_Score=1∩A3_Score=1∩A5_Score=1∩A10_ Score=1→ASD=Yes	0.920

Table 6
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Accuracy
1	A1_Score=1∩A3_Score=1∩A4_Score=1∩A5_ Score =1∩A10Score=1→ASD=Yes	0.9944
2	A3_Score=1∩A4_Score=1∩A9_Score=1∩A10_ Score=1→ASD=Yes	0.9943
3	A1_Score=0∩A4_Score=0∩A9_Score=0→ASD=No	0.9934
4	A1_Score=0∩A4_Score=0∩A8_Score=0→ASD=No	0.9933
5	A4_Score=0∩A10_Score=0→ASD=No	0.9932
6	A1_Score=0∩A10_Score=0→ASD=No	0.9924
7	A3_Score=0∩A6_Score=0==>Class/ASD=0→ ASD=No	0.9922
8	A5_Score=0∩A10_Score=0→ASD=No	0.9922
9	A1_Score=0∩A2_Score=0∩A8_Score=0→ASD=No	0.9915
10	A3_Score=0∩A7_Score=0→ASD=No	0.9915

Table 7
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Confirmation
1	A4_Score=1∩A10_Score=1→ASD=Yes	0.6686
2	A3_Score=1∩A4_Score=1∩A10_Score=Yes→ ASD=Yes	0.6651
3	A4_Score=0∪A10_Score=0 ∪ country_of_res= United_Arab_ Emirates →ASD=No	0.6756
4	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Italy → ASD=No	0.6756
5	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Bangladesh → ASD=No	0.6688
6	A4_Score=0 ∪ A10_Score=0→ASD=No	0.6686
7	A3_Score=0 ∪ A4_Score=0 ∪ A10_Score=0→ASD=No	0.6651
8	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Jordan → ASD=No	0.6623
9	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Russia → ASD=No	0.6620

Results for Identifying ASD in Adolescents

Table 8 illustrates that the apriori algorithm identified key characteristics strongly correlated with ASD in adolescents, particularly when A3, A5, A6, and A9 scores were all labelled as "Yes", resulting in a 100.00% confidence level. Other rules also showed significant correlations, with confidence levels ranging from 94.00% to 96.00%. Notably, the algorithm did not generate any rules for healthy adolescents, suggesting that these trends are specific to those with ASD and could serve as critical markers for early diagnosis and treatment.

Table 8
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Confidence
1	A3_Score=1∩A5_Score=1∩A6_Score=1∩A9_ Score=1→ASD=Yes	1.000
2	A3_Score=1∩A5_Score=1∩A6_Score=1→ASD =Yes	0.960
3	A3_Score=1∩A5_Score=1∩A9_Score=1→ASD =Yes	0.960
4	A5_Score=1∩A6_Score=1∩A9_Score=1∩A10_ Score=1 → ASD=Yes	0.960
5	A3_Score=1∩A4_Score=1∩A5_Score=1→ASD =Yes	0.960
6	A4_Score=1∩A5_Score=1∩A8_Score=1→ASD =Yes	0.950
7	A5_Score=1∩A6_Score=1∩A10_Score=1→ASD=Yes	0.940

For the predictive apriori algorithm, as shown in Table 9, the primary factors for identifying ASD in adolescents were the A3, A5, A6, and A9 scores being "Yes", achieving an accuracy of 99.46%. In contrast, for healthy adolescents, the absence of A3 and A10 scores (both marked as "No") was a key indicator, with an accuracy of 99.30%. These findings highlight the clear distinctions between ASD and non-ASD adolescents, offering valuable insights into accurate classification and early intervention.

Table 10, featuring the tertius algorithm, shows that for adolescents with ASD, the significant factors (with a confirmation of 66.52%) were the A3, A5, and A6 scores being "Yes". For healthy adolescents (with a confirmation rate of 69.09%), key indicators included A6_Score=0, A10_Score=0, or the adolescent being 13 years old, with a confirmation rate of 69.09%. A comparison of the three algorithms revealed that a positive response for A3 (tracking conversation) and A5 (conversation difficulty) could indicate the presence of ASD, while a negative response to tracking conversation (A3) and difficulties with conversations (A5) could indicate that the adolescent was non-ASD.

Results for Identifying ASD in Adults

In Table 11, the apriori algorithm identified key factors indicative of healthy adults with a confidence level of 97.00%, particularly when A6_Score, A9_Score, and jaundice status were "No". These combinations strongly suggest the absence of ASD (ASD=No).

Table 9
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Accuracy
1	A3_Score=1∩A5_Score=1∩A6_Score=1∩A9_Score=1→ASD=Yes	0.9946
2	A3_Score=1∩A5_Score=1∩A8_Score=1→ASD=Yes	0.9945
3	A7_Score=1∩A10_Score=1→ASD=Yes	0.9944
4	A1_Score=1∩A4_Score=1∩A5_Score=1∩A8_Score=1→ASD=Yes	0.9943
5	A3_Score=1∩A5_Score=1∩A6_Score=1∩A10_Score=1∩gender=Male → ASD=Yes	0.9932
6	A1_Score=1∩A7_Score=1∩gender=Female→ ASD=Yes	0.9925
7	A3_Score=0∩A10_Score=0→ASD=No	0.9930
8	A4_Score=0∩A5_Score=0→ASD=No	0.9917
9	A3_Score=0∩A5_Score=0→ASD=No	0.9912
10	A1_Score=0∩A5_Score=0→ASD=No	0.9906

Note. The symbol "∩" denotes the logical AND (intersection) between multiple conditions in an association rule. (A1- A9) are personal behavioural questions on the ASD screening data. The score of 1 represents the presence of the behaviour, whereas 0 represents the absence of the behaviour

Table 10
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Confirmation
1	A3_Score=1∩A5_Score=1∩A6_Score= 1→ASD=Yes	0.6652
2	A5_Score=1∩A6_Score=1∩A10_Score= 1→ASD=Yes	0.6398
3	A6_Score=0 ∪ A10_Score = 0 ∪ age =13 → ASD=No	0.6909
4	A3_Score=0 ∪ A5_Score=0 ∪ A6_Score= 0 → ASD=No	0.6652
5	A3_Score=0 ∪ A5_Score=0 ∪ country_of_res = Bahrain → ASD=No	0.6568
6	A3_Score=0 ∪ A5_Score=0 ∪ A9_Score=0 → ASD=No	0.6515
7	A3_Score=0 ∪ A5_Score=0 → ASD=No	0.6322

Table 11
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Confidence
1	A6_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.970
2	A6_Score=0∩A9_Score=0→ASD=No	0.960
3	A6_Score=0∩A9_Score=0∩autism=0 →ASD=No	0.960
4	A9_Score=0∩jundice=0∩autism=0 →ASD=No	0.930
5	A9_Score=0∩jundice=0→ASD=No	0.930
6	A9_Score=0	0.930
7	A9_Score=0∩autism=0 →ASD=No	0.930
8	A6_Score=0∩jundice=0∩autism=0 →ASD=No	0.910
9	A6_Score=0∩autism=0 →ASD=No	0.910
10	A6_Score=0∩jundice=0→ASD=No	0.900

Notably, the algorithm did not generate similar rules for adults with ASD, implying that these patterns were distinct from those of healthy individuals. Overall, the confidence levels ranged from 90.00% to 97.00%.

In Table 12, the predictive apriori algorithm highlighted crucial factors for identifying ASD in adults, achieving an accuracy of 99.44%. These factors included A1, A4, A6, A8, and A9 scores, all being "Yes". In contrast, the absence of A5, A6, and A9 scores predicted non-ASD adults with an accuracy of 99.46%.

Similarly, in Table 13, the Tertius algorithm revealed that the significant factors for identifying ASD in adults (with a confirmation of 57.54%) included A6_score =Yes, or A9_score =Yes, or the country of residence being Malaysia. For healthy adults, the significant factors (with a confirmation of 56.76%) were A6 Score =No and

A9 Score =No. A comparison of the three algorithms revealed that a positive response for detecting listener boredom (A6) and easily reading emotions (A9) may indicate the presence of ASD, while a negative response for grasping implied meanings (A5) and easily reading emotions (A9) could suggest the absence of ASD in adults.

Table 12
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Accuracy
1	A1_Score=1∩A4_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1→ASD=Yes	0.9944
2	A4_Score=1∩A5_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1→ASD=Yes	0.9943
3	A1_Score=1∩A5_Score=1∩A6_Score=1∩A7_Score=1∩A9_Score=1→ASD=Yes	0.9943
4	A1_Score=1∩A5_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1∩A10_Score=1→ASD=Yes	0.9942
5	A1_Score=1∩A3_Score=1∩A4_Score=1∩A5_Score=1∩A6_Score=1∩A9_Score=1∩A10_Score=1∩jundice=0→ASD=Yes	0.9939
6	A5_Score=0∩A6_Score=0∩A9_Score=0→ASD=No	0.9946
7	A3_Score=0∩A5_Score=0→ASD=No	0.9944
8	A4_Score=0∩A5_Score=0→ASD=No	0.9944
9	A3_Score=0∩A4_Score=0∩A6_Score=0→ASD=No	0.9943
10	A5_Score=0∩A7_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.9940

Table 13
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Confirmation
1	A6_Score=1∪A9_Score=1 ∪ country_of_res=Malaysia → ASD=Yes	0.5754
2	A6_Score=1∪A9_Score=1 ∪ country_of_res=Italy → ASD=Yes	0.5722
3	A6_Score=1∪A9_Score=1 ∪ country_of_res=Bangladesh → ASD=Yes	0.5708
4	Class/ASD=1==>A6_Score=1∪A9_Score= 1 ∪ country_of _res=Ireland → ASD=Yes	0.5694
5	A6_Score=1 ∪ A9_Score=1 ∪ age = 30 → ASD=Yes	0.5692
6	A6_Score=1 ∪ A9_Score=1 ∪ country_of_res = Canada → ASD=Yes	0.5681
7	A6_Score=1 ∪ A9_Score=1 ∪ country_of_res = Russia → ASD=Yes	0.5681
8	A6_Score=1 ∪ A9_Score=1→ASD=Yes	0.5676
9	A6_Score=0 ∩ A9_Score=0→ASD=No	0.5676

Performance Comparisons

Figure 5 demonstrates a comparative analysis of the outputs produced by three association rule mining algorithms apriori, predictive apriori, and tertius across four distinct age groups. Figure 5(a) displays the confidence values generated by the apriori algorithm, indicating that adolescents exhibit the highest confidence (100.00%) and children the lowest (95.00%). Figure 5(b) shows the accuracy values obtained from the predictive apriori method, with all age groups demonstrating consistently high accuracy, ranging between 099.40% and 99.50%. In contrast, Figure 5(c) illustrates the confirmation values derived from the Tertius method, which revealed greater variability across age groups, spanning from 57.50% in adults to 66.80% in children.

We evaluated the computational efficiency of the algorithms on four distinct ASD datasets (Figure 6). The apriori algorithm demonstrated instantaneous execution (0.0 s) across all age groups. In contrast, the tertius algorithm exhibited significantly higher computational demands, with the longest runtime observed in the adult group (5250s). Furthermore, the computational time varied by developmental stage: the toddler group showed the fastest execution, whereas the adult group required the most substantial processing time.

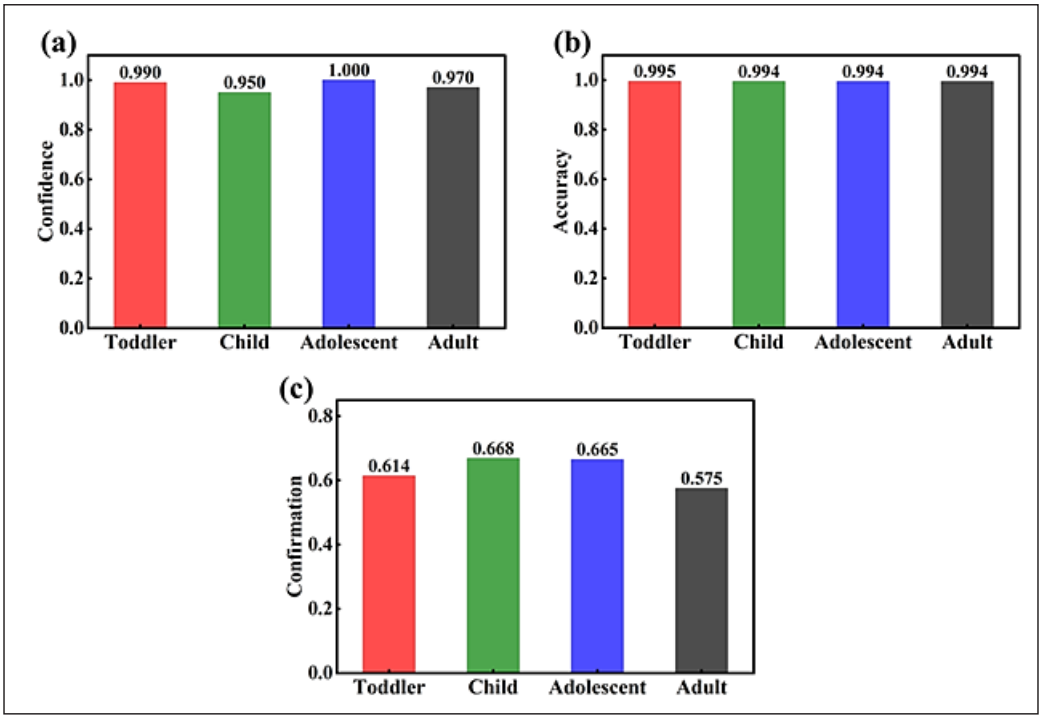


Figure 5. Comparison of the performance of three association rule mining techniques across four developmental age groups: (a) Apriori; (b) Predictive Apriori; and (c) Tertius

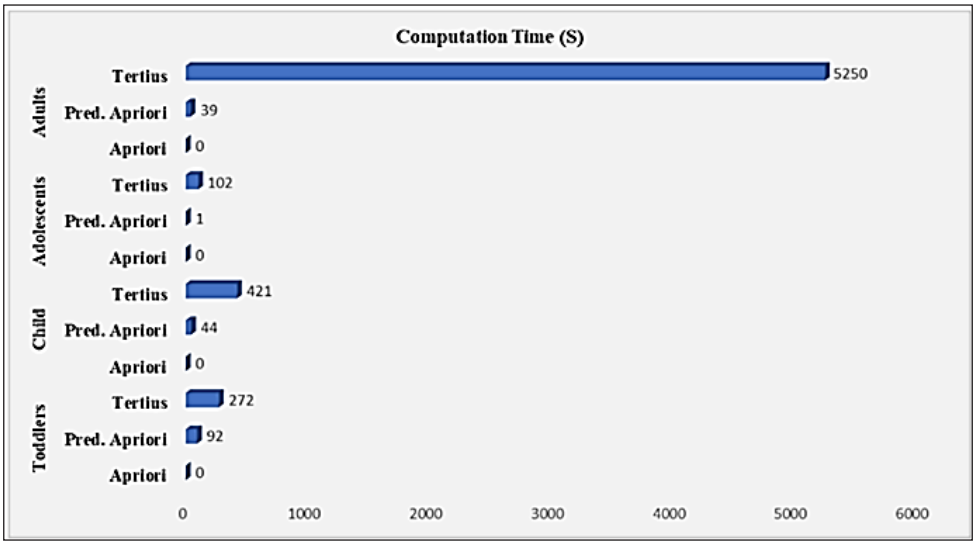


Figure 6. Computational time comparison of different association rule mining algorithms among four age groups

DISCUSSION

In this section, we interpret our findings for each dataset based on the rules found by applying different association rule mining algorithms. Thus, we tried to find some interesting insights into ASD at different ages.

Association Rules and Risk Factors to Detect ASD and Healthy Toddlers

Table 2,3 and 4 show the results of the different rule mining methods on the toddler dataset. We can see that the largest rule for identifying toddlers with ASD was extracted by the predictive apriori algorithm with 99.49% accuracy. The interpretation of this rule shows positive responses for "Sensing Small Sound (A1 Score)", "Pretending Games (A5 Score)", "Tracking Vision (A6 Score)", and "Sensing Others Grief and Comforting (A7 Score)", indicating a toddler with probable ASD. On the other hand, the apriori algorithm provides the smallest rule (single factor) to identify ASD toddlers with 96.00% accuracy, which indicates that an individual positive response to "Pretending Games (A5 Score)" or "Only Simple Gesturing Capability (A9 Score)" is associated with ASD toddlers. According to the apriori algorithm, a toddler might have ASD, although there is no "family member with PDD". Again, based on the predictive apriori algorithm, we found that positive responses for "Pretending Games (A5 Score)" and "Only Simple Gesturing Capability (A9 Score)" were associated factors for male toddlers. According to the Tertius algorithm, positive responses for "Pretending Games (A5 Score)" and "Only Simple Gesturing Capability (A9 Score)" are always risk factors associated with "26/16 Months of Age (Age Mons=26/16)"

and "Pacifica Ethnicity Group (Ethnicity=Pacifica)". However, rules for toddlers with no ASD or healthy toddlers indicate that the negative response for both "Pretending Games (A5 Score)" and "Only Simple Gesturing Capability (A9 Score)" is the predominant factor for healthy toddlers' identification in all algorithms. In addition, a negative response for "Born with Jaundice (Jaundice)" is another contributing factor indicating that toddlers have no ASD. Nonetheless, it was found that the "Country_of_Residence" was not used to identify ASD and healthy toddlers. In addition, it is noticeable that responses to the screening test questions are more important than other features in defining rules to differentiate between ASD and healthy toddlers.

Association Rules and Risk Factors to Detect ASD and Healthy Children

Association rules and risk factors to detect ASD and healthy Children Tables 5-7 show the results of the different rule mining methods on the child dataset. We can see that the largest rule for identifying children with ASD was extracted by the predictive apriori algorithm with 99.44% accuracy. Interpretation of this rule shows a positive response for "Sensing Small Sound (A1 Score)", "Tracking Conversation (A3 Score)", "Quick Activity Swapping (A4 Score)", "Difficulty in Continuing Conversation (A5 Score)", and "Difficulty in Making Friends (A10 Score)", indicating a child with probable ASD. Among all the rules found using all the algorithms here to identify children with ASD, the "Quick Activity Swapping (A4 Score)" and "Difficulty in Making Friends (A10 Score)" features were the most common. In the apriori algorithm, "Difficulty in Continuing Conversation (A5 Score)" is also an important feature. In the predictive apriori algorithm, "Tracking Conversation (A3 Score)" with a positive response is another important feature. In addition, features other than the responses to the screening test questions do not contribute to the identification of children with ASD.

However, rules derived from the Tertius algorithm indicate that negative responses for both "Quick Activity Swapping (A4 Score)" and "Difficulty in Making Friends (A10 Score)" are the predominating factors for identifying healthy children regardless of the "Country of Residence (country_of_res)". Besides that, a negative response to "Sensing Small Sound (A1 Score)", "Lack of Details Capturing (A2 Score)", "Tracking Conversation (A3 Score)", "Difficulty in Continuing Conversation (A5 Score)", "Social Chit Chat (A6 Score)", "Difficulty in Tracking Character's Intention (A7 Score)", "Pretending Games (A8 Score)", and "Interpreting Facial Expression (A9 Score)" is also important in identifying healthy children according to the predictive apriori algorithm. A negative response to "Tracking Conversation (A3 Score)" was also found to be important in the tertius algorithm. Again, other features than responses to the screening test questions did not contribute to the identification of healthy children.

Association Rules and Risk Factors to Detect ASD and Healthy Adolescents

Tables 8-10 present the results of various rule mining methods applied to the adolescent dataset. Among the rules generated by the three algorithms to identify adolescents with ASD, the most common traits observed included a positive response to "Tracking Conversation (A3 Score)", "Difficulty in Continuing Conversation (A5 Score)", "Social Chit-Chat (A6 Score)", and "Social Situations (A9 Score)". Notably, the predictive apriori algorithm derived two distinct rules for identifying male and female adolescents with ASD. For males, a positive response to "Tracking Conversation (A3 Score)", "Difficulty in Continuing Conversation (A5 Score)", "Social Chit-Chat (A6 Score)", and "Difficulty in Making Friends (A10 Score)" was the most indicative. In contrast, for females, positive responses to "Finding Patterns (A1 Score)" and "Pretending Games (A7 Score)" were more relevant. The most concise rule for identifying adolescents with ASD, according to the predictive apriori algorithm (with an accuracy of 99.45%), suggests that a positive response to both "Pretending Games (A7 Score)" and "Difficulty in Making Friends (A10 Score)" is sufficient, regardless of sex. Additionally, positive responses to "Quick Activity Swapping (A4 Score)" and "Lack of Empathy (A8 Score)" are also critical features for identifying adolescents with ASD. On the other hand, rules for identifying healthy adolescents indicate that negative responses to both "Tracking Conversation (A3 Score)" and "Difficulty in Continuing Conversation (A5 Score)" are key factors for distinguishing healthy adolescents. Furthermore, for adolescents aged 13 years, negative responses to "Social Chit-Chat (A6 Score)" and "Difficulty in Making Friends (A10 Score)" were found to be indicative of healthy adolescents.

Association Rules and Risk Factors to Detect ASD and Healthy Adults

Association rules and risk factors to detect ASD and healthy adults Tables 11 to 13 show the results of different rule mining methods on the adult dataset. We can see that the largest rule for identifying ASD adults was extracted using the predictive apriori algorithm with 99.39% accuracy. Interpretation of that rule shows the positive response for "Small Sound Sense (A1 Score)", "Multitasking (A3 Score)", "Quick Activity Swapping (A4 Score)", "Read Between Lines (A5 Score)", "Interpreting Facial Expression (A6 Score)", "Empathy (A9 Score)", and "Difficulty in Tracking Character's Intention (A10 Score)" and the negative response of "Born with Jaundice (Jaundice)" indicates an adult with probable ASD. However, the smallest rule derived from the tertius algorithm (with 56.76% confirmation) suggests that the positive response to "Interpreting Facial Expression (A6 Score)" and "Empathy (A9 Score)" can solely identify adults with ASD regardless of their country of residence and age. However, the apriori algorithm had the smallest rule for finding healthy adults (with a confidence level of 93.00%). This shows that an adult does not have ASD if they give a negative answer for "Empathy (A9 Score)". In addition, the negative responses

for both "Interpreting Facial Expression (A6 Score)" and "Empathy (A9 Score) were the predominant factors for healthy adult identification in all algorithms. Moreover, a negative response for "Born with Jaundice (Jaundice)" and "Diagnosed with Autism (autism)" is another contributing factor that indicates that adults have no ASD, according to the apriori algorithm. As the same response to this feature is present in both the "ASD=Yes" and "ASD=No" classes, it can be said that "Born with Jaundice (Jaundice)" is not a strong discriminatory factor in defining rules to differentiate adults with ASD and healthy adults.

From the above discussions on all datasets, screening test responses are more helpful than other features in identifying patients with ASD. In addition, those born with jaundice may have no relationship with ASD. It is also notable that male and female ASD patient identification rules might differ. Although the screening test questionnaires were slightly different in different datasets, it is noticeable from the above discussion that there are some similarities among the important features found to identify patients with ASD in all datasets. As the questionnaires were different, we may consider the factors that appeared at least twice in the entire experiment as common risk factors for potential ASD patients regardless of age, sex, and country of residence. These common factors include: "Sensing Small Sounds", "Pretending Games", "Tracking Conversations", "Difficulty in Continuing Conversation", "Difficulty in Making Friends", "Difficulty in Tracking Character's Intentions", and "Quick Activity Swapping". Besides, in case of all datasets, we tried to compute Jaccard index value from the rules that we found among the top N rules extracted by apriori, predictive apriori and tertius algorithms. We have obtained a Jaccard index of 0.0 for all algorithms pair apriori vs predictive apriori, apriori vs tertius, and predictive apriori vs tertius of each dataset in this case which indicates that there is no overlap in the sets of extracted rules when top N-rules are considered. This indicates that every algorithm records diverse behavioural characteristics because of variations in evaluation strategies as well as structural rule constraints.

Conflicts among Rule Sets across Different Algorithms

From the above discussion, we can see that different algorithms that we used for association rule mining find out partially overlapping rule sets, although not fully identical. However, we can also observe some conflicts among the rule sets. For instance, the apriori and predictive apriori algorithms associate Pretending Games (A5 Score) with ASD toddlers, while the tertius algorithm associates not ASD toddlers with co-existing Pretending Games (A5 Score) and Pacifica Ethnicity. Again, apriori algorithm only mines rule sets for ASD=no, while the other two mine rules for both ASD = yes and ASD = no. These conflicts among rule sets across algorithms are shown because of the differences in the evaluation metrices used in different algorithms. apriori algorithm focuses on maximising confidence while predictive apriori maximises expected predictive accuracy. On the other hand, tertius

maximises confirmation. So, the conflicts we mentioned can be seen as different approaches of defining interestingness. Where apriori and predictive apriori algorithms find more frequent rules, tertius tries to mine surprising (e.g., high confirmation) rules. For clinical uses we can use some ensemble techniques where we can accept any factor as a high risk of ASD only if it appears in rule sets of at least two algorithms. Thus, we can overcome any biases introduced by a specific algorithm and can make more robust predictions.

Caution in Interpreting Demographic Factors

Although behavioural screening scores are the most important indicators in studying ASD, we still get some association rules with demographic factors like age, country of residence, and ethnicity, especially in the rules derived from the tertius algorithm. Unlike behavioural screening scores, demographic factors may come from uneven sampling and imbalance in the dataset. And so, association rules with demographic factors have quite a high chance to represent correlation rather than causal relation. This correlation can either be spurious or incidental. While interpreting the rules with demographic factors, caution should be taken. Without further controlled validation, these factors should not be interpreted as clinically meaningful.

CONCLUSION

Autism Spectrum Disorder (ASD) remains a critical developmental issue, emphasising the necessity of identifying the key determinants influencing its diagnosis. In our study, we explored the associations between various characteristics and ASD classification by utilising three well-established association rule mining algorithms: Apriori, Predictive Apriori, and Tertius. The apriori algorithm demonstrated the highest confidence of 99.00%, 95.00%, 100.00%, and 97.00% for toddlers, children, adolescents, and adults, respectively. In contrast, predictive apriori exhibited superior accuracy, with peak values of 99.50% (toddler) and a consistently high performance of 99.40% across the child, adolescent, and adult datasets. The tertius algorithm yielded varying confirmation scores: 61.40% for toddlers, 66.80% for children, 66.50% for adolescents, and 57.50% for adults. The association rules uncovered several key behaviours indicative of ASD across developmental stages. For toddlers, “pretending games (A5)” and “tracking vision (A6)” emerged as significant factors, while “quick activity swapping (A4)” and “difficulty in making friends (A10)” were significant for children with ASD. For adolescents, “tracking conversations (A3)” and “difficulty in conversations (A5)” were prominent indicators, and for adults, factors such as “grasping implied meanings (A5)”, “detecting listener boredom (A6)”, and “easily reading emotions (A9)” were significant in distinguishing individuals with ASD from their healthy counterparts. It is evident that responses to screening tests are more informative than other features for identifying patients with ASD. In addition, a history

of jaundice at birth was not associated with ASD. There was also no strong correlation between a familial ASD history and ASD occurrence. Furthermore, we observed that the apriori algorithm required the least manipulation time, followed by the Predictive apriori algorithm, while the Tertius algorithm required the most manipulation time. This research most directly benefits clinicians and diagnosticians by providing data-driven behavioural markers for more accurate and earlier ASD screening across age groups. A short list of red flag behaviours for ASD is the output of this research that can be used by clinicians to assess patients in minutes during check-up. For instance, when screening a patient, if one or more high-risk factors of his/her age group are found in him/her, clinicians can promptly suspect the patient of ASD, and the further evaluation plan would be easier for the clinician. Thus, emphasising these high-risk factors will reduce diagnosis delays. This work is also aiding researchers through a comparative analysis of the effectiveness of rule mining algorithms. Although each of the three algorithms are theoretically complete, later we also evaluated the quality of the extracted rules in terms of some additional evaluation metrics, like lift and reverse confidence to determine the strength of the identified associations. Additionally, future studies will build on this work by examining larger datasets and investigating gender-specific association rules to improve the diagnostic accuracy.

ACKNOWLEDGEMENT

The authors would like to thank Jahangirnagar University and Gono Bishwabidyalay for their support in completing this research. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207-216. https://doi.org/10.1145/170035.170072

Aher, S. B., & Lobo, L. (2012). A comparative study of association rule algorithms for course recommender system in e-learning. *International Journal of Computer Applications*, *39*(1), 48-52. https://doi.org/10.5120/4788-7021

Akter, T., Khan, M. I., Ali, M. H., Satu, M. S., Uddin, M. J., & Moni, M. A. (2021). Improved machine learning based classification model for early autism detection. In *Proceedings of the 2nd International Conference on Robotics, Electrical and Signal Processing Techniques* (pp. 742-747). IEEE. https://doi.org/10.1109/ICREST51555.2021.9331013

Al-Diabat, M. (2018). Fuzzy data mining for autism classification of children. *International Journal of Advanced Computer Science and Applications*, *9*(7), 11-17. https://doi.org/10.14569/IJACSA.2018.090702

Aksu, G., & Dogan, N. (2019). An analysis program used in data mining: WEKA. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(1), 80-95. https://doi.org/10.21031/epod.399832

Cai, Q. R. (2020). Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm. *IEEE Access*, *8*, 75607-75615.https://doi.org/10.1109/ACCESS.2020.2988288

Delja, F. (n.d.). *Autism screening data for toddlers* [Data set]. Kaggle.

Engle, K., & Rada, R. (2017). Knowledge-guided mutation in classification rules for autism treatment efficacy. *Health Informatics Journal*, *23*(1), 56-68. https://doi.org/10.1177/1460458215627187

Engle, K. M. (2013). *Data mining and domain knowledge: An exploration of methods to advance medical research* (Doctoral dissertation, University of Maryland, Baltimore County).

Engle, K. M., & Rada, R. (2011). A top-k analysis using multi-level association rule mining for autism treatments. In C. Stephanidis (Ed.), *Universal access in human-computer interaction: Applications and services* (Part IV, Vol. 6768, pp. 328-334). Springer. https://doi.org/10.1007/978-3-642-21657-2_35

Flach, P. A., & Lachiche, N. (2001). Confirmation-guided discovery of first-order rules with tertius. *Machine Learning*, *42*(1-2), 61-95. https://doi.org/10.1023/A:1007656703224

Hamid, Z., Khalique, F., Mahmood, S., Daud, A., Bukhari, A., & Alshemaimri, B. (2024). Healthcare insurance fraud detection using data mining. *BMC Medical Informatics and Decision Making*, *24*(1), Article 112. https://doi.org/10.1186/s12911-024-02512-4

Hareendran, S. A., & Chandra, S. S. V. (2017). Association rule mining in healthcare analytics. In Y. Tan, H. Takagi, & Y. Shi (Eds.), *Data mining and big data: DMBD 2017* (Vol. 10387, pp. 31-39). Springer.https://doi.org/10.1007/978-3-319-61845-6_4

Hasanpour, H., Meibodi, R. G., Navi, K., Shams, J., Asadi, S., & Ahmadiani, A. (2019). Fluvoxamine treatment response prediction in obsessive-compulsive disorder: Association rule mining approach. *Neuropsychiatric Disease and Treatment*, *15*, 895-904. https://doi.org/10.2147/NDT.S200569

Ishibuchi, H., Kuwajima, I., & Nojima, Y. (2007). Prescreening of candidate rules using association rule mining and Pareto-optimality in genetic rule selection. In B. Apolloni, R. J. Howlett, & L. Jain (Eds.), *Knowledge-based intelligent information and engineering systems: KES 2007-WIRN 2007, Part II, Proceedings* (Vol. 4693, pp. 509-516). Springer. https://doi.org/10.1007/978-3-540-74829-8_63

Kane, C. (2018). *Classification using association rules* [Thesis]. Technological University Dublin.

Kaoura, G., Kovas, K., Boutsinas, B., & Hatzilygeroudis, I. (2025). A new hybrid reasoning model based on rules, cases and processes: Application to care of individuals facing autism spectrum disorders. *Knowledge and Information Systems*, *67*(1), 371-401. https://doi.org/10.1007/s10115-024-02228-x

Maitra, S., Akter, N., Zahan Mithila, A., Hossain, T., & Shafiul Alam, M. (2021). Apriori-backed fuzzy unification and statistical inference in feature reduction: An application in prognosis of autism in toddlers. In K. Saeed, N. Gazem, & R. Chaki (Eds.), *Progress in advanced computing and intelligent engineering: Proceedings of ICACIE 2020* (pp. 233-254). Springer. https://doi.org/10.1007/978-981-33-4299-6_20

Mazumdar, P., Arru, G., & Battisti, F. (2021). Early detection of children with autism spectrum disorder based on visual exploration of images. *Signal Processing: Image Communication*, *94*, Article 116184.https://doi.org/10.1016/j.image.2021.116184

Mutter, S., Hall, M., & Frank, E. (2004). Using classification to evaluate the output of confidence-based association rule mining. In G. I. Webb & X. Yu (Eds.), *AI 2004: Advances in artificial intelligence* (Vol. 3339, pp. 538-549). Springer. https://doi.org/10.1007/978-3-540-30549-1_47

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, *40*(4), 1086-1093. https://doi.org/10.1016/j.eswa.2012.08.028

Niranjan, A., Nitish, A., Shenoy, P. D., & Venugopal, K. R. (2016). Security in data mining: A comprehensive survey. *Global Journal of Computer Science and Technology: C Software & Data Engineering*, *16*(5).

Rubio-Martín, S., García-Ordás, M. T., Bayón-Gutiérrez, M., Prieto-Fernández, N., & Benítez-Andrades, J. A. (2024). Enhancing ASD detection accuracy: A combined approach of machine learning and deep learning models with natural language processing. *Health Information Science and Systems*, *12*(1), Article 20. https://doi.org/10.1007/s13755-024-00281-y

Scheffer, T. (2005). Finding association rules that trade support optimally against confidence. *Intelligent Data Analysis*, *9*(4), 381-395. https://doi.org/10.3233/IDA-2005-9405

Singh, K. N., & Mantri, J. K. (2024). An intelligent recommender system using machine learning association rules and rough set for disease prediction from incomplete symptom set. *Decision Analytics Journal*, *11*, Article 100468. https://doi.org/10.1016/j.dajour.2024.100468

Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*, *26*(1), 264-286.https://doi.org/10.1177/1460458218824711

Towle, P. O., & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: A systematic review. *Autism Research and Treatment*, *2016*, Article 4624829. https://doi.org/10.1155/2016/4624829

UCI Machine Learning Repository. (n.d.-a). *Autistic spectrum disorder screening data for adolescent dataset* [Data set]. University of California, Irvine.

UCI Machine Learning Repository. (n.d.-b). *Autistic spectrum disorder screening data for children data set* [Data set]. University of California, Irvine.

UCI Machine Learning Repository. (n.d.-c). *Autism screening adult data set* [Data set]. University of California, Irvine.

Uddin, M. J., Ahamad, M. M., Sarker, P. K., Aktar, S., Alotaibi, N., Alyami, S. A., Kabir, M. A., & Moni, M. A. (2023). An integrated statistical and clinically applicable machine learning framework for the detection of autism spectrum disorder. *Computers*, *12*(5), Article 92. https://doi.org/10.3390/computers12050092

Wang, T. H., & Guo, F. (2010). Associating IDS alerts by an improved Apriori algorithm. In F. Yu, X. Peng, H. Liu, J. Shu, & R. Ng (Eds.), *Third International Symposium on Intelligent Information Technology and Security Informatics* (pp. 478-482). IEEE. https://doi.org/10.1109/IITSI.2010.47

Wiratsin, I.-O., & Narupiyakul, L. (2021). Feature selection technique for autism spectrum disorder. In *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence* (pp. 53-56). ACM. https://doi.org/10.1145/3448218.3448241

Wingate, M., Kirby, R. S., Pettygrove, S., Cunniff, C., Schulz, E., Ghosh, T., Robinson, C., Lee, L. C., Landa, R., Constantino, J., Fitzgerald, R., Zahorodny, W., Daniels, J., Nicholas, J., Charles, J., McMahon, W., Bilder, D., Durkin, M., Baio, J., Christensen, D., Van, K., Braun, N., Clayton, H., Goodman, A., Doernberg, N., & Yeargin-Allsopp, M. (2014). Prevalence of autism spectrum disorder among children aged 8 years-Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. *MMWR Surveillance Summaries*, *63*(2), 1-21.

Yan, X., Zhang, C., & Zhang, S. (2009). Confidence metrics for association rule mining. *Applied Artificial Intelligence*, *23*(8), 713-737. https://doi.org/10.1080/08839510903208062

Zhao, Q., & Bhowmick, S. S. (2003). *Association rule mining: A survey* (Technical Report No. 2003116). Nanyang Technological University, Centre for Advanced Information Systems.

Pertanika J. Sci. & Technol. 34 (2): 721 - 742 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

JaNarX: Smart Soil Moisture Prediction using Jackal Optimiser Enabled Machine Learning Model in IoT Coupled Agricultural Applications

Seema Jitendra Patil[1,2]* and B. Ankayarkanni[1]

[1]*Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, 600119 Chennai, Tamil Nadu, India*

[2]*School of Computer Science and Engineering, Dr. Vishwanath Karad, MIT World Peace University, 411038 Pune, Maharashtra, India*

ABSTRACT

Soil moisture can be used to provide predictive information important for precision farming, irrigation management, and pollution monitoring. Nonetheless, existing approaches often struggle to handle non-linear temporal relationships, environmental uncertainties, and poor real-time integration of sensor data, thereby degrading prediction performance. To overcome these problems, in this article, we introduce JaNarX, a soil moisture prediction framework (SMPF) based on IoT and the NARX model, optimised with the Jackal Apis Optimisation (JAO) algorithm. The approach utilises time-series radar satellite variables, meteorological terms, and continuous in situ sensor data to better capture dynamic changes of soil moisture. The Wazihub Soil Moisture Dataset (WSMD), which is the aggregation of multi-sensor environmental data measured in in-field conditions from real agricultural fields in Senegal, was used for training and validation. JAO was used to speed up convergence and tune NARX hyperparameters for solving the local minima problem,and increase the generalisation performance of the model. Experimental results in MAE, MSE, and RMSE evaluations had MAE = 1.53, MSE = 2.89, and RMSE = 1.70, which outperformed the state-of-the-art base-models such as SVM, LSTM, GLM XGBR, and traditional NARX network. The simulation results confirm that the proposed IOT-based, JAO-optimised NARX structure can achieve more accurate prediction than the traditional method and has good stability and computational efficiency. This paper presents a scalable, high-accuracy model for the prediction of soil moisture, suitable to enable real-time planning decisions for precision irrigation and sustainable water management in agriculture.

Keywords: Internet of things, Jackal APIs optimisation, machine learning, precision agriculture, soil moisture

ARTICLE INFO

Article history:
Received: 16 October 2025
Accepted: 12 December 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.06

E-mail addresses:
seema.patil@mitwpu.edu.in (Seema Jitendra Patil)
ankayarkanni.cse@sathyabama.ac.in (B. Ankayarkanni)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Soil moisture is an important factor in the hydrologic, ecological, and agricultural systems, affecting plant growth and nutrient uptake as well as crop yield (Liu et al. 2014). Maintaining adequate soil moisture content is especially crucial in precision agriculture because both drought and waterlogging conditions will adversely affect the development of plants (Cai et al., 2019). In addition to being the frontiers of technology, like IoT devices and automated irrigation systems or sensor networks that promote these developments (Magallanes-Quintanar et al., 2022), smart agriculture has advanced significantly in recent years with soil monitoring capabilities and real-time field assessment. Yet, despite these efforts, soil moisture still poses a difficult prediction and characterisation problem because it exhibits strong spatial and temporal variability, complex environmental interactions, and dependence on a wide variety of physical properties and climate parameters.

Conventional soil moisture estimation means, including the hydrological model, the soil water balance model and empirical equations, need local calibration over a large area or are computationally expensive. Furthermore, they fail to consider nonlinear interactions that dominate between the soil and environment (Liu et al., 2020; Li et al., 2022). Remote sensing-derived methods are suitable for large-scale monitoring but tend to have low temporal resolution and cannot be used in real time (Gao et al., 2021). For example, machine learning and deep learning models have been shown to enhance the predictive accuracy of soil moisture and temperature, but are still limited by issues such as a lack of convergence stability in training, poor generalisation between distinct and heterogeneous soil conditions, sensitivity to noisy IoT sensor-based data, and difficulty with modelling long-term temporal dependencies (Karthikeyan et al., 2021). Moreover, most of the currently available IoT-based frameworks are unable to successfully assimilate multiple environmental parameters, such as air temperature, humidity, wind factors, and atmospheric pressure, into a unified predictive model such that performance is inconsistent in time and space (Ekonomou et al., 2007; Fotis et al., 2007), and there is a high level of uncertainty.

In the literature, many methods have been proposed to predict soil moisture, ranging from hydrological and land surface models (Ming et al., 2022), statistical models such as ARMA or empirical water balance equations (Liu et al., 2020), inversion models based on remote sensing (Gao et al., 2021), and machine learning-type or deep learning-based algorithms like SVM, GLM, XGBoost, LSTM, etc. (Granata et al., 2022; Karthikeyan et al., 2021; Nguyen et al., 2022). While these methods have revealed some merits, they are still facing several limitations: physics-based models require parameter tuning for different crops and suffer from computational insufficiency; The remote sensing approaches cannot operate in real time; ML/DL models fail to learn *nonlinear* temporal relationships rather than inherent convergence problems and a mismatch of patterns between trained data and transfer data across different agricultural conditions. Optimisation-based methods (e.g., BES-enhanced SVM) show the benefits of parameter optimisation, but they do not

model multi-source IoT variables jointly or capture the intricate nonlinear hydrological dynamics.

In comparison, the proposed JaNarX model makes use of memory-intensive NARX network architecture and a hybrid JAO optimisation algorithm that leads to a more efficient nonlinear modelling, enhanced convergence dynamics, less localised minima consequences, and more trustworthy IoT-inspired multi-parameter predictions. This model hybridisation, which not only improves the predictability of the JaNarX model but also provides a much stronger alternative to currently available soil moisture prediction methods.

The literature shows a need for a flexible, robust, and data-driven model that can predict nonlinear soil moisture dynamics and that works well by using networked sensor fusion from diverse sources, as well as fulfilling the necessity of keeping its predictions stable under an uncertain environment. NARX is an effective choice for time-series hydrological forecasting because of its memory structure and ability to capture complex temporal dynamics. Nevertheless, NARX models are highly dependent on learning processes, which consist of trial-and-error setting and tuning of parameters. If not properly trained, NARX may return suboptimal models (e.g., converging to local minima and performing poorly in predictions), indicating that efficient optimisation techniques are necessary for improving the training process of NARX.

In order to facilitate these issues, JaNarX, an IoT-based soil moisture prediction model using Jackal Apis Optimisation (JAO) for the optimisation process, is proposed. JAO integrates the cooperation strategies of the Artificial Bee Colony (ABC) algorithm with the chasing behaviour of Coyote Optimisation (CO), and it has good convergence, global search ability, and parameter setting. The motivation behind incorporating JAO into NARX is to help overcome the limitations of the nonlinear model and make it robust under environmental uncertainty, which leads to better prediction accuracy.

The literature gap is systematically covered through the following research questions:

RQ1 : How can the accuracy of soil moisture prediction be improved in a nonlinear, uncertain, and time-varying natural environment?

RQ2 : Whether an IoT-based NARX model trained with JAO can reduce the forecast errors over conventional ML, DL, and non-optimised NARX approaches?

RQ3 : Do predictions considering temporal multi-parameter IoT sensor data (e.g., soil moisture, air temperature, humidity, atmospheric pressure, wind) perform better than those without integrating them?

RQ4 : How much is the convergence improved, model parameters optimised, and local minima avoided by the JAO method in the JaNarX framework?

RQ5 : How well does the proposed JaNarX model perform compared to state of the art models (SVM, LSTM, XGBR, GLM, DT, and LR) and classical NARX in terms of delay-based prediction assessment?

The objective of this study is to propose an IoT-enabled, JAO-tuned NARX model for real-time accurate estimation of SM. Specifically, the study intends to:

RO1 : Develop a combined model, capturing nonlinear hydrological dynamics more precisely

RO2 : Combine multi-source environmental sensor data into an integrated forecasting system (IPS).

RO3 : Use JAO to train NARX and thereby obtain better fitting results and convergence performance and benchmark the proposed approach under different delay patterns versus the state-of-the-art solution.

Several methods for forecasting Soil Moisture have been explored, such as hydrological models (Pal & Olsson, 2004), statistical techniques (Al-Yaari et al., 2016), and empirical water balance equations. Mechanistic hydrological models offer conceptual understanding but are limited by large computational requirements and uncertainty in environmental forcings (Li et al., 2022). They enlarge spatial coverage, although with a compromised temporal resolution and susceptibility to atmospheric noise in remote sensingderived models (Hegazi et al., 2023). Machine learning techniques, such as SVM, Gaussian regression, XGBoost and LSTM achieve better prediction accuracy (Granata et al., 2022; Nguyen et al., 2022), but they still face the challenge of modelling long-term dependency and performance deterioration for noisy or missing sensor data. A BES-optimised SVM (Huang, 2023) is also a good template to refer to, but it still cannot ensure the stable convergence and well-extended generalisation over wide-ranging farming conditions.

Accordingly, there is a serious demand for an optimisation-based predictive framework IoT-supported that can properly represent the nonlinear features of time-series soil moisture and guarantee convergence to potential users in agriculture. While current soil moisture prediction algorithms, such as hydrological models, statistical time-series methods, ML regressors, and DL architectures, can provide good insights into soil water phenomena, they suffer from the common problems in dealing with nonlinear temporal dynamics, heterogeneous environmental inputs, and noisy IoT sensor readings. Machine learning (ML) and DL approaches (e.g., SVM, XGboost, LSTM) frequently suffer from convergence instability and struggle in capturing long-term dependencies (Granata et al., 2022; Karthikeyan et al., 2021), which precludes them from being generalisable over different soil types and climates (Li et al., 2022). While the optimisation-enhanced methods, such as BES-enhanced SVM, have certainly relieved the parameter adjustment focus, there lacks one real time and consistent answer for multi-sensor fusion in IoT systems.

The JaNarX model overcomes these issues by integrating the NARX structure, which has a large memory requirement and therefore poor nonlinear modelling characteristics, but is complemented with the hybrid Jackal Apis Optimisation (JAO) technique that is

capable of bringing better convergence properties and is immune to temporal shifting and local minima effects. This hybrid system enhances prediction performance without sacrificing real-time application.

This paper has numerous original contributions in terms of soil moisture prediction, agriculture for the IoT-driven analytics and model-based optimisation Hybrid JaNarX forecasting model, by combining an IoT-based NARX system with the Jackal Apis Optimisation (JAO) algorithm to deal with nonlinearity as well as uncertainty in the soil moisture dynamics. The JAO algorithm, a hybrid metaheuristic approach that includes both artificial bees foraging behaviour and cooperative coyote hunting strategies, is utilised to enhance the hyperparameter tuning process while preventing convergence to local minima as well as improving model training speed. Integration of multiparameter IoT sensor data, soil humidity, air temperature and humidity, atmospheric pressure, wind speed and direction, gusts in a single predictive model to improve temporal accuracy and environment adaptability. Performance testing of delay-based derivatives for varying population size and time-lags, with JaNarX exhibiting better generalisation stability and predictability when compared to nine state-of-the-art (LR/DT/GLM/SVM/LSTM/XGBR/NARX/CANAX/ANAX) models. Extensive statistical validation (MAE, MSE, RMSE, and R^2) proves that prediction error is reduced as well as the reliability on the real Wazihub Soil Moisture dataset. Demonstrated practical real-time implementation showing that the JaNarX–JAO framework for precision irrigation can be deployed for better water management, and to account for real-time field variabilities in large-scale farms.

LITERATURE REVIEW

Modern agriculture depends critically on effective management of soil moisture, which influences irrigation efficiency, crop yield, and resource economy. Manual sampling and laboratory analysis are among the time-consuming, labour-intensive traditional soil moisture monitoring techniques that lack real-time adaptation. Furthermore, although being extensively used, sensor-based methods often suffer from data loss, misinterpretation, and errors resulting from environmental conditions (Rezk et al., 2021).

Different research has looked at cutting-edge techniques for irrigation optimisation and soil moisture prediction. Using deep learning methods to examine moisture variance (Chandrappa et al., 2023) presented a soil moisture modelling approach enabled by IoT and machine learning. But the method was limited in that it couldn't make real-time irrigation changes. To schedule irrigation plans, Gao et al. (2021) also developed an IoT-based environmental data collecting system connected with Bidirectional Long Short-Term Memory (Bid-LSTM) networks. Although the study needed more improvements in transfer learning to fit various regional conditions, its expected performance was good.

As noted by Ming et al. (2022), indicated that missing value is still a common problem in soil moisture data, and we also found many missing values during the exploratory analysis. While processing these gaps added computational burdens, it also enhanced the general trustworthiness of the dataset. Similarly, Ainiwaer et al. (2020) demonstrated that soil hyperspectral data can be complex, where noise or redundant information could have a negative impact on model predictive capability. Pandey et al. (2021) presented TOTRAM, a soil moisture retrieval method using the land surface temperature (LST) and vegetation indices (VIs), but this technique is computationally demanding and susceptible to different environmental conditions. Similarly, Huang (2023) further enhanced predictive ability by modifying the support vector machine (SVM) with Bald Eagle Search (BES) in tuning for hyperparameters; however, this technique had poor generalisability across agricultural zones.

To break through the bottlenecks of conventional methods, several researchers have introduced machine learning 17-18, deep learning 19-20, and IoT 21-based techniques, which can provide better efficiency and flexibility for soil moisture prediction and irrigation management. For example, Nguyen et al. (2022) adopted a hybrid XGBoost Regression (XGBR) model with the use of GA to select features, which proved to be capable of an accurate prediction of soil moisture variations at different locations around the world. These examples highlight the AI-driven support that precision agriculture could enjoy thanks to such systems, with specific intelligence and guidance for farmers, municipalities, or water management organisations.

Furthermore, emerging as good substitutes for conventional irrigation systems are IoT-driven precision irrigation models. (Abioye et al., 2021) presented an IoT-based model using predictive analytics to improve real-time soil parameter monitoring and optimise irrigation schedules. To guarantee greenhouse applications, though, a greater connection with adaptive control systems was required. Using real-time soil moisture, temperature, and humidity data, Podder et al. (2021) created an IoT-enabled Smart AgroTech system that dynamically changes irrigation. Although it underlined sensor coverage limits that need optimisation, the study confirmed the dependability of IoT-driven agriculture.

Dealing with missing data (Cordeiro et al., 2022) showed how deep learning models coupled with KNN data imputation methods improve soil moisture predictions while reducing discrepancies. Likewise, Bid-LSTM networks, as used by (Gao et al., 2021), have shown promise in long-term irrigation scheduling; nevertheless, more study is required to increase computational efficiency and adaptability over several farming environments. Overcoming a few restrictions of conventional sensor-based approaches (Ainiwaer et al., 2020), also proposed a multi-source remote sensing method integrating hyperspectral and multispectral data to improve the spatial resolution of soil moisture predictions.

By means of analysis of important influencing elements along with topography, climate, soil type, and vegetation, the proposed JaNarX approach seeks to overcome the

shortcomings of past models by refining soil moisture predictions. JaNarX greatly reduces overfitting problems, speeds convergence, and keeps models from being caught in local minima by using the JAO method for optimisation. Moreover, sophisticated methods of data preparation promise the removal of extraneous elements, thereby enhancing the predictive power and dependability of soil moisture estimations. These methods taken together position JaNarX as a highly scalable, flexible, and efficient precision agricultural solution.

This work presents an advanced soil moisture prediction model based on the JaNarX model to overcome these challenges. The model considered IoT-generated soil moisture data, sent to a cloud server for storage and additional processing. Embedded in a controller, a NARX machine learning-based model forecasts soil moisture levels, enabling effective water management and lowering of waste. To optimise the weights of the NARX model, a hybrid optimisation method combining the hunting behaviour of jackals with the communication strategies of bee optimisation is applied: the JAO method. This optimisation reduces overfitting, improves convergence speed, and helps the model not to get caught in local optima. The proposed system guarantees high detection accuracy, enhanced prediction dependability, and a general increase in agricultural water-use efficiency by combining the JAO-optimised JaNarX model.

MATERIAL AND METHODS

The proposed methodology aims to develop a predictive model for soil moisture to optimise irrigation scheduling based on meteorological variations. It comprises four main components, which are: sensor networks, data aggregation, predictive modelling, and decision support. Wireless sensor nodes collect real-time data on soil moisture, temperature, and humidity, which is transmitted to a central server and preprocessed for accuracy, as shown in Figure 1.

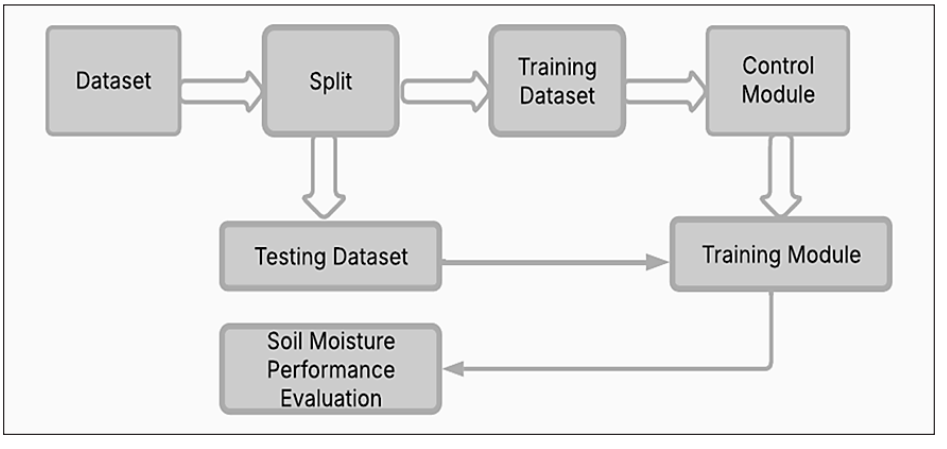


Figure 1. Schematic representation of the proposed soil moisture prediction model

Data Collection

This study utilised the Wazihub soil moisture dataset (Bera et al., 2024), collected over four months from four agricultural plots in Senegal cultivated with peanuts and maize. IoT sensors were deployed in each plot to continuously monitor soil moisture dynamics, enabling precise tracking of hydration levels across different crop environments. To maintain consistent environmental conditions, the plots were positioned within one met of each other. The dataset comprises key parameters, including soil humidity, air humidity, atmospheric pressure, wind speed, air temperature, wind gust, and wind direction-critical variables for comprehensive soil moisture assessment.

IoT-based Soil Moisture Prediction using Jackal Apis Optimised Nonlinear Autoregressive Exogenous (JaNarX) Model

JaNarX is a recurrent dynamic network effective at simulating the nonlinear time series with feedback connections. In addition, the NARX network provides time-delayed feedback on the model's output for the modelling of a nonlinear dynamical system. Considerably, it is intriguing to make use of the NARX neural network's memory function by utilising the historical values of predicted or genuine time series in order to maximise its performance for nonlinear time series prediction. The vector w_r forms the input to the soil moisture prediction, which is forwarded to theNonlinear Autoregressive Exogenous model to predict the moisture present in the soil. JaNarX networks are an effective class of models well-suited for this kind of challenge when modellingnonlinear systems, especially time series. The mathematical formulation of the NARX model is:- of a nonlinear dynamical system. Considerably, it is intriguing to make use of the NARX neural network's memory function by utilising the historical values of predicted or genuine time series in order to maximise its performance for nonlinear time series prediction. The vector w_r forms the input to the soil moisture prediction, which is forwarded to the Nonlinear Autoregressive Exogenous model to predict the moisture present in the soil. JaNarX networks are an effective class of models well-suited for this kind of challenge when modelling nonlinear systems, especially time series. The mathematical formulation of the NARX model is:

$$\hat{z}(t) = f\begin{pmatrix} z(t-1), z(t-2), \ldots . z(t-n_z), \\ w_r(t-1), w_r(t-2), \ldots . w_r(t-n_w) + e(t) \end{pmatrix} \quad [1]$$

The weight of the connection between the i^{th} hidden neuron and the input neuron $w(t-m)$ is indicated by V_{im} a_j is the bias of the i^{th} hidden neuron; f_h is the hidden layer activation function; f_y is the hidden transfer and output function. Further, V_{iq} is the

connection weight between the i^{th} hidden neuron and the output feedback. Combining the output of the hidden layer allows one to get the last prediction as follows:

$$\hat{Z}_j(t) = f_y\left[\sum_{i=1}^{n_h} V_{ji}\, h_i(t) + a_j\right] \quad [2]$$

where n_h is the number of hidden neurons; V_{ij} is the connection weight between the i^{th} hidden neuron and the j^{th} expected output; a_j is the bias of the j^{th} expected output; and f_y is the activation function for the output layer.

Figure 2 shows the Ja NarX network architecture. Increasing hidden neurons helps the model to perform better, so improving learning capacity and generalisation. By changing weights and biases, Jackal Apis Optimisation (JAO) minimises overfitting and increases prediction accuracy, so refining the network.

$$h_i(t) = f_h\left[\sum_{m=0}^{n_w} V_{im}\, w(t-m) + \sum_{q=0}^{n_z} V_{iq}\, z(t-q) + a_i\right] \quad [3]$$

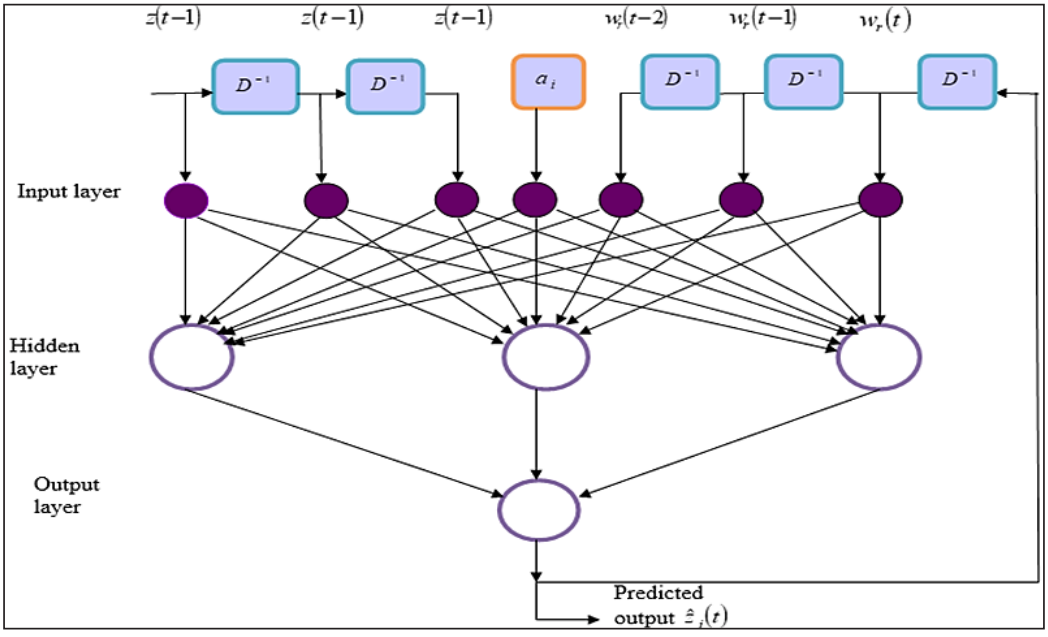


Figure 2. Architecture of the developed JaNarX network

JAO adjusts the hyperparameters of the JaNarX model such that they maximise prediction performance. It hybridises the Artificial Bee Colony (ABC) and Coyote Optimisation Algorithm (COA). Inspired by bee foraging techniques and jackal hunting behaviour, JAO advances quick convergence and effective exploration.

Initialisation of Jackal Coyote Population

Initialising the JAO algorithm as $g_t^{h,\alpha} = [V, \alpha]$ results in a weight and a bias of the JaNarX classifier. In this optimisation, t_e represents the jackal coyote population overall and t_f represents the total number of possible solutions. Based on the social conditions (g) of the jackal coyotes, the jackal apis optimisation is mathematically stated, and these jackal coyotes are considered as the solutions in the search space. As a result, the following equation depicts the social situation of the g^{th} jackal h^{th} coyote pack at that particular moment in time α is given as:

$$g_{tc}^{h,\alpha} = c^{low} + \beta_c\left(c^{up} - c^{low}\right) \quad [4]$$

Here, c stands for the dimension of the jackal coyote's search space, while low and up are used to describe the lower and upper bounds of the search space, respectively. β_c stands for a valid random number that is generated between [0,1] and assigned using uniform probability.

Determine Fitness Function Value

The cost of the system is decided by the fitness solution, which is dependent on the social situation determined by the jackal coyote's behaviour in adapting to its environment. Further, the fitness of the model is based on the lower MSE of the solution and is defined as follows:

$$fit\left(g_t^h\right) = min\left(MSE\left(g_t^h\right)\right) \quad [5]$$

The population of jackal coyotes is initialised as the first phase in the jackal bee optimisation process. According to the following representation, each jackal coyote's social position is initially allocated at random inside the search area.

Exploration Phase

The jackal coyotes are randomly assigned to a group, but occasionally they quit their pack and join another pack or go solitary. The number of members in the pack determines

whether this jackal coyote is expelled, and the possibility that it is expelled from the packs is based on the probabilities listed below:

$$J_t = 0.005\, t_f \qquad [6]$$

There are typically 14 individuals in each pack; the probability given in the equation above should be higher than for the coyote solution $t_f \leq \sqrt{200}$. This exemplifies the coyote's social or communicative behaviour, which facilitates effective coyote interaction and aids in the development of a global search strategy.

Leader Selection Phase

Only one leader, Jackal Coyote (L) that is most suited to the environment is chosen for this optimisation, and this leader jackal coyote is regarded as the most suited to the environment. The leader of the coyotes or the alpha coyote is considered as the best-adapted coyote, and it can be mathematically given as follows: The leader jackal coyote of the h^{th} pack at a given instant α is indicated in a minimisation problem by:

$$G^{t,\alpha} = \left\{ g_t^{h,\alpha} \middle| arg_{t=\{1,2,\ldots tf\}} min\, n\left(g_t^{h,\alpha}\right) \right. \qquad [7]$$

Communicative Phase

The jackal coyote's swarm intelligence demonstrates the ability of the coyote species to share its social conditions with all the packs, illuminating their organisational behaviour and highlighting the importance of each jackal coyote posse in preserving social behaviour. The algorithm considers the life cycle of coyotes, including birth and death. Specifically, the birth of coyotes is a combination of the social behaviour of two parents that are selected randomly within the search space, along with the environmental factor. The pack's whole data is computed, and their social tendency behaviour is represented as:

$$f_c^{h,\alpha} = \begin{cases} U^{h,\alpha}_{\left(\frac{t_f+1}{2}\right),c;} & ;if\, t_f\, isodd \\ U^{h,\alpha}_{\left(\frac{t_f}{2}\right),c} + U^{h,\alpha}_{\left(\frac{t_f+1}{2}\right),c} & ;else \end{cases} \qquad [8]$$

Here, $U^{(h,\alpha)}$ stands for the ranked social state of every jackal coyote t in the pack in the range $[1, g]$ of dimension c . The cultural tendency of a single pack is determined using the average social conditions of all coyotes in that pack.

Balancing Probability

The age of jackal coyotes is determined using two biological features, such as birth and death. The social interactions of the other two jackal coyote wolves and the birth of the additional wolves are provided by:

$$Z_c^{h,\alpha} = \begin{cases} g_{b_{1,c}}^{h,\alpha}; r_c < J_{scatter} \ \ or\ c = c_1 \\ g_{b2}^{h,\alpha}; r_c \geq J_{scatter} \ + J_{associate} \ \ or\ c = c_2 \\ B_c; else \end{cases} \quad [9]$$

In this case, Z stands for the pub of the jackal coyote among the random coyotes b_1 and b_2 in the random dimensions c_1 and c_2. The decision variable's random number is selected, and it is symbolised by the value B_c. In addition, r_c designates a random number in the range [0,1] That was produced using a uniform probability distribution. The dispersion and associative probability assigned to jackal coyotes are represented by, $J_{scatter}\ and\ J_{associate}$, and these probabilities aid in maintaining social communicative behaviour. Following is a mathematical equation that describes the probabilities:

$$J_{scatter} = \frac{1}{g} \quad [10]$$

$$J_{associate} = \left(1 - \frac{1 - J_{scatter}}{2}\right) \quad [11]$$

Exploitation Phase

If a solution does not show any improvement and the number of trials without change exceeds a predefined threshold, known as the boundary, it is considered to have stagnated. In this case, the scout bees will abandon the current solution. After that, the scout bees begin a random search process to explore new potential solutions in the search space. This mechanism helps maintain diversity in the population and prevents the algorithm from getting stuck in local optima:

$$g_{new}^{h,\alpha} = g_{tc}^{h,\alpha} + \delta_{tc}\left(g_{tc}^{h,\alpha} - g_{tc}^{h,\beta}\right) \quad [12]$$

where δ_{tc} is a random number in the range $[-1,1]$

Enhanced Convergence

Coyotes can locate their prey and surround them after recognising their location. The jackal coyote generally directs the hunt. However, we are unsure of the exact location of the ideal (prey) in an abstract search space. To mathematically replicate the coyote's hunting behaviour, we suppose that the jackal coyote has a stronger awareness of the likely location of prey. In this scenario, to improve the attacking strategy, the communication behaviour of the bee is hybridised with the jackal coyote; hence jackal bee communicates with the group members and performs hunting in a cooperative manner, which improves the hunting speed and provides faster convergence. The hybridised equation for the solution update is expressed as follows:

$$g_t^{h+1} = 0.5Z_c^{h,\alpha} + 0.5g_{new}^{h,\alpha} \quad [13]$$

$$g_t^{h+1} = 0.5\left[\begin{cases} g_{b_{1,c}}^{h,\alpha}; r_c < J_{scatter}\ orc = c_1 \\ g_{b2}^{h,\alpha}; r_c \geq J_{scatter} + J_{associate}\ orc = c_2 \\ B_c; else \end{cases}\right] + 0.5\left[g_{tc}^{h,\alpha} + \delta_{tc}\left(g_{tc}^{h,\alpha} - g_{tc}^{h,\beta}\right)\right] \quad [14]$$

In this case, the hunting operation is directed by the coefficient vectors. $\vec{B}$ and $\vec{P}$, the determined vector $\vec{M}$ and the vector of the coyote's positions $\vec{Q}$, which are used to construct a new position for the jackal coyote. Another way to state it is, μ estimates the locations of the prey and the other wolves before updating their new locations arbitrarily close to the prey.

The probability of successfully obtaining food is determined by the fading parameter. u, which is a positive integer number, the distance t between the bee and the other colony members, the targeted communication range, g denotes the neighbour bee node.

Termination

The global solution is declared through the final updated position of the solution, and the verification process ends when the iteration reaches the condition $(t > t_{max}())$.

Algorithm: Pseudo Code for Proposed Jackal Apis Optimisation

S.NO	Pseudo Code
1	Input: $g = (g_1, g_2, \ldots g_E)$
	Output: g_t^{h+1}
2	Determine fitness: $fit\left(g_t^h\right) = min(\,MSE(g_t^h))$
3	Exploration phase: $J_t = 0.005 t_f$
4	Selection of leader: $G^{t,\alpha} = \left\{ g_t^{h,\alpha} \middle\vert ar g_{t=\{1,2,\ldots tf\}} \, min\, n\left(g_t^{h,\alpha}\right)\right.$
5	Communicative behaviour: $f_c^{h,\alpha} = \begin{cases} U^{h,\alpha}_{\left(\frac{t_f+1}{2}\right),c;}; if\, t_f\, is odd \\ U^{h,\alpha}_{\left(\frac{t_f}{2}\right),c}; else \end{cases}$
6	Balancing probability: $Z_c^{h,\alpha} = \begin{cases} g_{b_1,c}^{h,\alpha}; r_c < J_{scatter}\ orc = c_1 \\ g_{b2}^{h,\alpha}; r_c \geq J_{scatter} + J_{associate}\ orc = c_2 \\ B_c; else \end{cases}$
7	Probability: $J_{associate} = \left(1 - \frac{1 - J_{scatter}}{2}\right)$
	Exploitation Phase: $g_{new}^{h,\alpha} = g_{tc}^{h,\alpha} + \delta_{tc}\left(g_{tc}^{h,\alpha} - g_{tc}^{h,\beta}\right)$
8	Solution update with Enhanced convergence: $g_t^{h+1} = 0.5\left[\begin{cases} g_{b_1,c}^{h,\alpha}; r_c < J_{scatter}\ orc = c_1 \\ g_{b2}^{h,\alpha}; r_c \geq J_{scatter} + J_{associate}\ orc = c_2 \\ B_c; else \end{cases}\right] + 0.5\left[g_{tc}^{h,\alpha} + \delta_{tc}\left(g_{tc}^{h,\alpha} - g_{tc}^{h,\beta}\right)\right]$
9	Termination

RESULTS

Model Evaluation Overview

The IoT-SM-based JaNarX model was evaluated using the Wazihub Soil Moisture Database, with performance assessed through MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and R^2. These metrics collectively capture the magnitude of prediction errors, their variability, and the correlation between predicted and observed values. The results indicate that JaNarX provides reliable predictions with high accuracy and stability across different parameter settings.

Delay Analysis

Delay profile analysis was conducted with a delay of 25 and population sizes ranging from 10 to 50, shown in Figure 3. The results show that increasing population size significantly enhances model performance. Specifically, MAE decreased from 2.87 to 1.70, and MSE decreased from 8.22 to 2.89, reflecting a substantial reduction in prediction errors. Similar improvements were observed for RMSE, while R^2 values remained consistently around 0.93, indicating a strong correlation between predicted and observed soil moisture and confirming the stability of the JaNarX model.

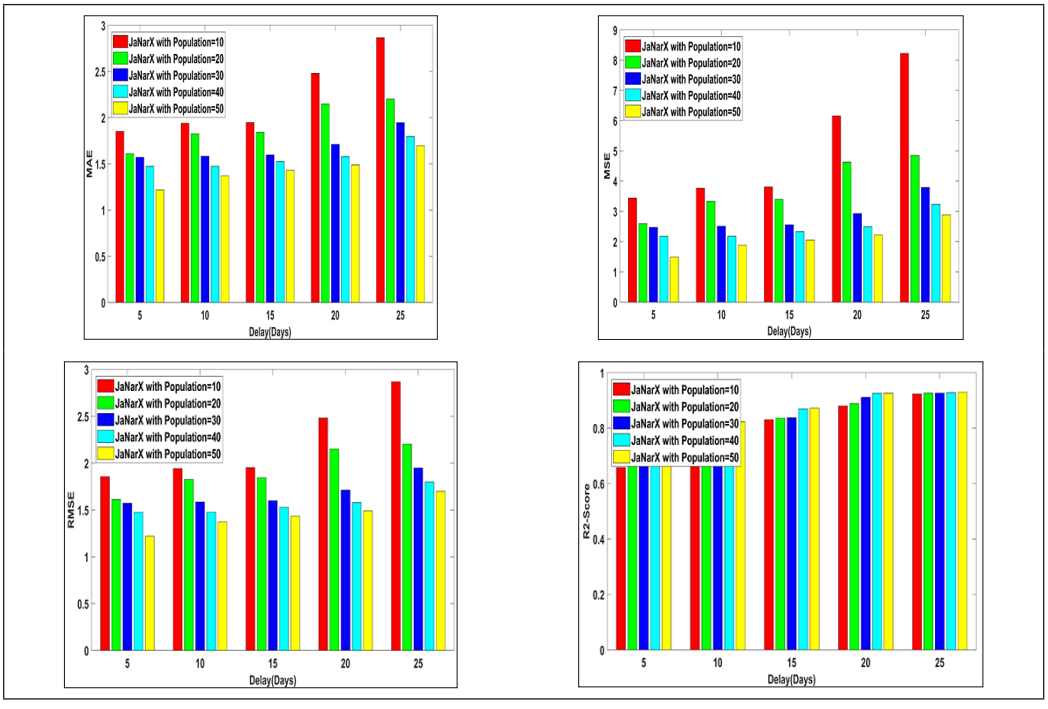


Figure 3. Performance analysis of JaNarX using the Wazihub soil moisture database with MAE, MSE, RMSE, and R^2 metrics

Statistical Evaluation

Statistical analysis included computation of means, variances, and best-case performance (Table 1). The lower mean error and variance observed for JaNarX suggest greater reliability and minimal fluctuation under test conditions compared to benchmark models. These results validate that JaNarX generates soil moisture estimates that are more stable and accurate than competing methods.

Table 1
Statistical summary of Wazihub soil moisture database performance metrics

Models	Best	Mean	Variance	Best	Mean	Variance	Best	Mean	Variance
LR (Rani et al., 2022)	6.99	4.56	3.91	60.35	30.52	587.27	7.77	5.07	4.82
DT (Nikjoo et al., 1997)	5.70	3.40	1.34	40.18	15.90	147.62	6.34	3.77	1.65
XGBR (Nguyen et al., 2022)	4.33	3.01	0.45	25.52	12.22	44.62	5.05	3.39	0.70
GLM (Pekár & Brabec, 2018)	2.96	2.62	0.05	10.86	8.54	2.07	3.29	2.91	0.06
SVM (Huang, 2023)	2.84	2.36	0.09	10.00	6.99	3.16	3.16	2.62	0.11
LSTM (Datta & Faroughi, 2023)	2.84	2.44	0.07	10.00	7.47	2.74	3.16	2.72	0.09
NARX (Siegelmann et al., 1997)	2.84	2.27	0.11	10.00	6.51	3.81	3.16	2.53	0.13
CNaX (Lin et al., 1996)	2.84	2.15	0.16	10.00	5.90	5.29	3.16	2.39	0.20
AnaX (Omkar et al., 2020)	2.77	2.07	0.16	9.51	5.51	4.82	3.08	2.31	0.19
JaNarX	1.53	1.30	0.02	2.89	2.11	0.21	1.70	1.44	0.02

Loss Curve Behaviour

The training loss curve (Figure 4) demonstrates a smooth and continuous decrease in error over epochs, indicating effective learning behaviour. This validates the capability of the JAO algorithm to optimise the NARX network efficiently. The declining loss curve also confirms that the model successfully captures temporal dependencies and reduces uncertainty in predictions.

Comparative Evaluation

Comparative performance under a delay of 25 is shown in Figure 5, where JaNarX is compared against LR, DT, XGBR, GLM, SVM-LSTM, NARX, CNAX, and ANAX

models. JaNarX achieves the lowest error metrics with MAE = 1.53, MSE = 2.88, and RMSE = 1.69, outperforming all other models by a significant margin. Additionally, R^2 = 0.93, the highest among the compared models, demonstrates superior ability to capture the nonlinear behaviour of soil moisture dynamics and stronger predictive skill. These results confirm that JaNarX improves both generalisation and accuracy compared to classical ML, DL, and NARX-based approaches.

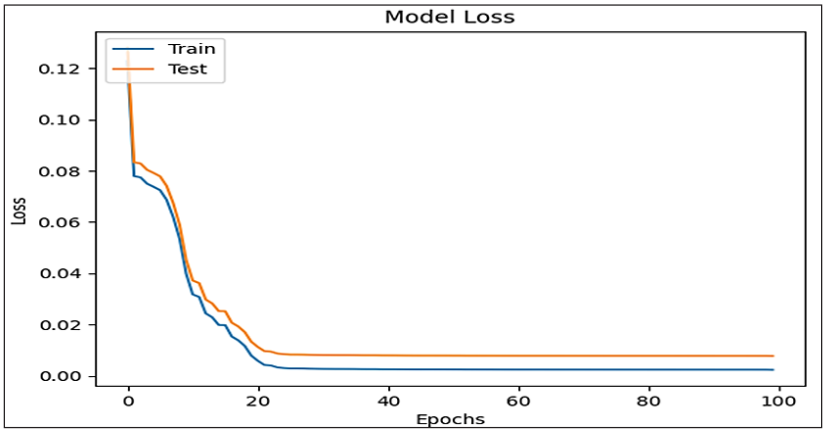


Figure 4. Training loss curve of JaNarX on the Wazihub soil moisture database

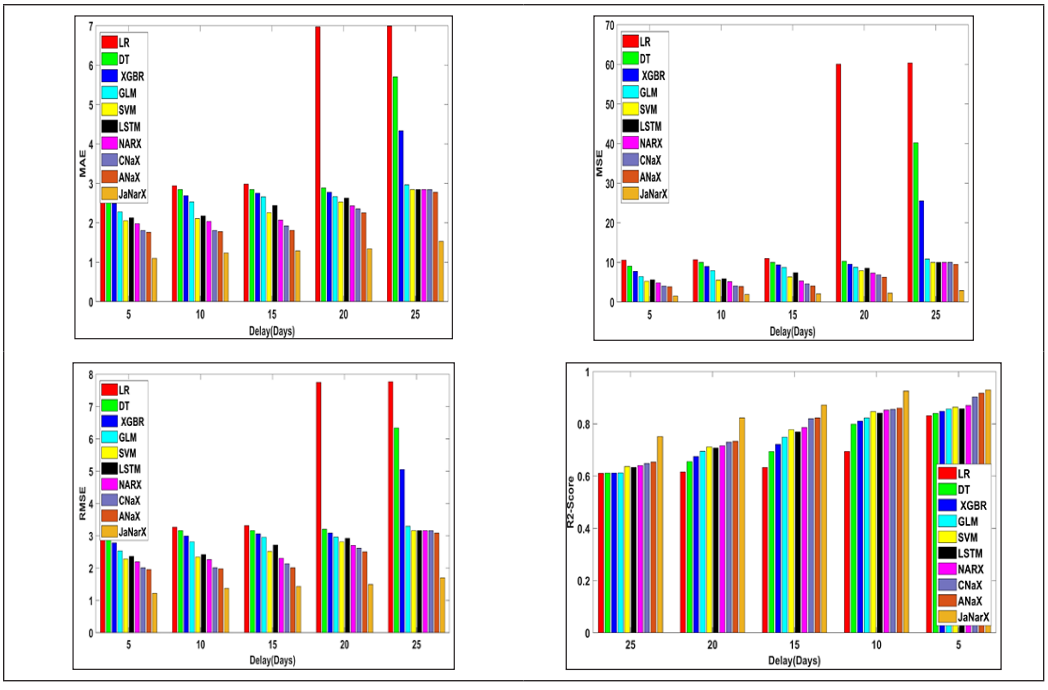


Figure 5. Comparative evaluation of JaNarX and benchmark models on the Wazihub soil moisture database with MAE, MSE, RMSE, and R^2 metrics

DISCUSSION

The results shown in this study indicate that the JaNarX model significantly increases the quality of soil moisture forecasts. Through multi-parameter IoT sensor data fusion, the model can effectively reflect various environmental effects. JaNarX being used with the NARX structure that has a large amount of memory, can learn complex temporal dependencies that traditional models cannot consider. Additionally, the consistently high R^2, together with decreases in MAE, MSE and RMSE for all training population sizes, again underlines the model's great generalisation ability across diverse scenarios.

The faster behaviour of the loss descent curve also emphasises the convergence-enhancing effectiveness of the JAO algorithm. Smooth and reasonable convergence of training loss verifies that JAO can avoid local minima effectively and allow for a stable optimisation. When compared to traditional hydrological models and statistical predictors, as well as deep learning methods, JaNarX outperforms for robustness and reliability, in particular given noisy IoT sensor inputs and nonlinear temporal trends in soil moisture.

In summary, quantitative comparisons demonstrate that JaNarX achieves better performance than previous methods in term of accuracy and robustness. Earlier models like LR and DT suffer from high prediction errors because they cannot capture nonlinear interactions. Some of the state-of-the-art deep learning-based models, like LTSM, offer superior temporal learning, but their convergence is often not guaranteed, and convergence to noise is a challenge, hence delivering far inferior performance than JaNarX. It is observed that, although CNAX and ANaX are more complex models in terms of model structure, they do not lead to much of a gain in error reduction or correlation.

These results also have potential practical applications for precision agriculture. The improved predictability of JaNarX may aid real-time irrigation scheduling and soil-water management, helping to reduce the possibility of over- or under-application. Moreover, the stability of the model under differential context supports its integrability in a landscape-scale, real-time decision support system.

CONCLUSION

To tackle the long-standing issues of soil moisture prediction, this article has provided an innovative JaNarX model for the NARX framework in the IoT domain, empowered by the JAO algorithm that is particularly designed to minimise the impacts of uncertainty and non-linear perturbations and temporally varying responses. Multi-parameter time-series data was utilised effectively to decrease the ambiguity between input and output data values and represent dynamic hydraulic features for practical agricultural use. The inclusion of climatic variables like soil humidity, air temperature and humidity, atmospheric pressure and wind has allowed JaNarX to provide a complete picture of the variability in soil moisture throughout agricultural regions.

Because of the continuous monitoring through IoT sensors, such a model was able to capture seasonal patterns and aid in data-driven decisions on when to irrigate, how much

water should be applied and how much efficiency will result from it. Due to the adaptive parameter search of JAO and the nonlinear structure of NARX, it allowed for faster convergence, decreased training loss and more stable predictions. Validation on Wazihub Soil Moisture Database showed it to outperform versus benchmark models (MAE = 1.53, MSE = 2.89, RMSE = 1.70).

JaNarX has some limitations despite its predictive power. The model is, however, dependent on the temporal continuity of environmental data, and its generalisation could suffer in the case of sudden or extreme events, which are not aligned to learned temporal patterns, like abrupt rainfall. The short-term reliability of the forecast can also be decreased by noise spikes or missing sensor measurements. Furthermore, predictions based on IoT devices developed in a particular geographic location can have reduced utility in areas with different soil or climate and would need to be recalibrated or retrained.

Possible future directions of the JaNarX framework could be hybrid deep learning architectures that merge NARX's temporal modelling with the feature extraction capability of LSTM, GRU or transformer-based attention mechanisms. These strategies could improve stress tolerance to rapid changes in the environment and field adaptation. Data assimilation or physics-based primitives can also further enhance robustness and promote immediate reaction to malfunctions of sensors. Generalising these findings in different climate, land types, and soil profiles will lend to application in precision agriculture/smart farming ecosystems.

ACKNOWLEDGEMENT

Experiments in this paper are performed on data from the Wazihub Soil Moisture Prediction Challenge, by Zindi. The authors thank Wazihub, Zindi, Microsoft and Université Gaston Berger (Senegal) for the access to this Soil Moisture dataset used in this study. The Wazihub initiative is funded by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 780229, which allowed making this dataset publicly available for research purposes.

LIST OF ABBREVIATIONS

XGBR	:	XGBoost regressor
GLM	:	Generalised linear model
SVM-LSTM	:	Support vector machine- long short-term memory
NARX	:	Nonlinear AutoRegressive with eXogenous inputs
MAE	:	Mean absolute error
MSE	:	Mean square error
RMSE	:	Root mean squared error
ABC	:	Artificial bee colony

REFERENCES

Abioye, E. A., Mohd Ali, M. A., Ahmad, M. S., *et al.* (2021). IoT-based monitoring and data-driven modelling of drip irrigation system for mustard leaf cultivation experiment. *Information Processing in Agriculture, 8*(2), 270-283. https://doi.org/10.1016/j.inpa.2020.06.002

Ainiwaer, M., Ding, J., Kasim, N., Wang, J., & Wang, J. (2020). Regional scale soil moisture content estimation based on multi-source remote sensing parameters. *International Journal of Remote Sensing, 41*(9), 3346-3367. https://doi.org/10.1080/01431161.2019.1701723

Bera, S., Dey, T., Mukherjee, A., & De, D. (2024). FLAG: Federated learning for sustainable irrigation in agriculture 5.0. *IEEE Transactions on Consumer Electronics*. Advance online publication. https://doi.org/10.1109/TCE.2024.3370373

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., & Xue, X. (2019). Research on soil moisture prediction model based on deep learning. *PLOS ONE, 14*(4), Article e0214508. https://doi.org/10.1371/journal.pone.0214508

Chandrappa, V. Y., Ray, B., Ashwatha, N., & Shresth, P. (2023). Spatiotemporal modelling to predict soil moisture for sustainable smart irrigation. *Internet of Things, 21*, Article 100671. https://doi.org/10.1016/j.iot.2022.100671

Cordeiro, M., Rosário, D., Gomes, D., *et al.* (2022). Towards smart farming: Fog-enabled intelligent irrigation system using deep neural networks. *Future Generation Computer Systems, 129*, 115-124. https://doi.org/10.1016/j.future.2021.11.029

Datta, P., & Faroughi, S. A. (2023). A multihead LSTM technique for prognostic prediction of soil moisture. *Geoderma, 433*, Article 116452. https://doi.org/10.1016/j.geoderma.2023.116452

Ekonomou, L., Fotis, G., Maris, T., & Liatsis, P. (2007). Estimation of the electromagnetic field radiating by electrostatic discharges using artificial neural networks. *Simulation Modelling Practice and Theory, 15*(9), 1089-1102. https://doi.org/10.1016/j.simpat.2007.07.003

Fotis, G., Ekonomou, L., Maris, T., & Liatsis, P. (2007). Development of an artificial neural network software tool for the assessment of the electromagnetic field radiating by electrostatic discharges. *IET Science, Measurement & Technology, 1*(5), 261-269. https://doi.org/10.1049/iet-smt:20060137

Gao, P., Li, H., Wang, Z., Zhao, Y., Sun, W., & Zhang, Y. (2021). Improved soil moisture and electrical conductivity prediction of citrus orchards based on IoT using deep bidirectional LSTM. *Agriculture, 11*(7), Article 635.https://doi.org/10.3390/agriculture11070635

Granata, F., Di Nunno, F., Najafzadeh, M., & Demir, I. (2022). A stacked machine learning algorithm for multi-step ahead prediction of soil moisture. *Hydrology, 10*(1), Article 1. https://doi.org/10.3390/hydrology10010001

Hegazi, E. H., Samak, A. A., Yang, L., Huang, R., & Huang, J. (2023). Prediction of soil moisture content from Sentinel-2 images using convolutional neural network (CNN). *Agronomy, 13*(3), Article 656. https://doi.org/10.3390/agronomy13030656

Huang, Y. (2023). Improved SVM-based soil-moisture-content prediction model for tea plantation. *Plants, 12*(12), Article 2309. https://doi.org/10.3390/plants12122309

Karthikeyan, L., & Mishra, A. K. (2021). Multi-layer high-resolution soil moisture estimation using machine learning over the United States. *Remote Sensing of Environment, 266*, Article 112706.https://doi.org/10.1016/j.rse.2021.112706

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., & Yu, F. (2022). An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma, 409*, Article 115651. https://doi.org/10.1016/j.geoderma.2021.115651

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks, 7*(6), 1329-1338. https://doi.org/10.1109/72.548162

Liu, M., Huang, C., Wang, L., Zhang, Y., & Luo, X. (2020). Short-term soil moisture forecasting via Gaussian process regression with sample selection. *Water, 12*(11), Article 3085. https://doi.org/10.3390/w12113085

Liu, Y., Mei, L., & Ooi, S. K. (2014). Prediction of soil moisture based on extreme learning machine for an apple orchard. In *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 400-404). IEEE. https://doi.org/10.1109/CCIS.2014.7175760

Magallanes-Quintanar, R., Galván-Tejada, C. E., Galván-Tejada, J. I., Méndez-Gallegos, S. d. J., García-Domínguez, A., & Gamboa-Rosales, H. (2022). NARX neural network models for prediction of the standardised precipitation index in Central Mexico. *Atmosphere, 13*(8), Article 1254.https://doi.org/10.3390/atmos13081254

Ming, W., Zhang, Y., Chen, L., He, L., & Liu, Z. (2022). A hybrid triple collocation-deep learning approach for improving soil moisture estimation from satellite and model-based data. *Remote Sensing, 14*(7), Article 1744.https://doi.org/10.3390/rs14071744

Nguyen, T. T., Do, H. H., Nguyen, D. H., Tran, Q. H., & Pham, T. T. (2022). A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithms. *Science of the Total Environment, 833*, Article 155066. https://doi.org/10.1016/j.scitotenv.2022.155066

Nikjoo, P. O. N., Goodhead, D. T., & Terrissol, H. M. (1997). Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. *International Journal of Radiation Biology, 71*(5), 467-483.https://doi.org/10.1080/095530097143509

Omkar, S., Mudigere, D., Senthilnath, J., & Kumar, M. V. (2020). Identification of helicopter dynamics based on flight data using nature-inspired techniques. In *Deep learning and neural networks: Concepts, methodologies, tools, and applications* (pp. 257-273). IGI Global. https://doi.org/10.4018/978-1-7998-0945-6.ch012

Pandey, R., Goswami, S., Sarup, J., & Matin, S. (2021). The thermal–optical trapezoid model-based soil moisture estimation using Landsat-8 data. *Modelling Earth Systems and Environment, 7*(2), 1029-1037. https://doi.org/10.1007/s40808-020-00975-8

Pekár, S., & Brabec, M. (2018). Generalised estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology, 124*(2), 86-93.https://doi.org/10.1111/eth.12709

Podder, A. K., Chaki, R., Ghosal, R., Chakraborty, S., & Ghosh, S. (2021). IoT-based smart agrotech system for verification of urban farming parameters. *Microprocessors and Microsystems, 82*, Article 104025. https://doi.org/10.1016/j.micpro.2021.104025

Rani, A., Kumar, N., Kumar, J., & Sinha, N. K. (2022). Machine learning for soil moisture assessment. In *Deep learning for sustainable agriculture* (pp. 143-168). Elsevier. https://doi.org/10.1016/B978-0-323-85214-2.00001-X

Rezk, N. G., Hemdan, E. E.-D., Attia, A.-F., El-Sayed, A., & El-Rashidy, M. A. (2021). An efficient IoT-based smart farming system using machine learning algorithms. *Multimedia Tools and Applications, 80*, 773-797.https://doi.org/10.1007/s11042-020-09740-6

Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 27*(2), 208-215. https://doi.org/10.1109/3477.558801

Pertanika J. Sci. & Technol. 34 (2): 743 - 762 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Interactive Location-based Augmented Reality Virtual Chatbot

Lai Yee Jen, Goh Eg Su*, and Johanna Ahmad

Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

ABSTRACT

Large campus's navigation becomes a crucial issue nowadays due to its layout complexity and digital services' diversities. Addressing this challenge, this research focusses on Universiti Teknologi Malaysia and presents a novel campus navigation solution that incorporates location-based Augmented Reality with an Artificial Intelligence chatbot. In contrast to previous AR navigation systems, which rely on static visual markers, and chatbot systems, which provide text-based support, the proposed system links real-time AR guidance with intelligent and context-sensitive handling of inquiries. This dual functionality capability sets the solution apart because users can follow AR-driven directions, ask questions, and receive instant information about campus facilities. The prototype was developed using an Agile-inspired iterative methodology that allows for continuous refinement based on user feedback. Evaluation results show a high degree of user satisfaction: 95% of respondents agreed that the system enhances the clarity of wayfinding, while 70% reported that they feel more confident in navigating a campus independently. The study emphasises the practical impact of using combined AR and AI technologies for campus digitisation by showing that the system can be effective in improving navigation efficiency, reducing reliance on physical signage and human assistance, and making the process more engaging and interactive for new students and visitors. In this work, the authors provide a novel model of a hybrid AR-chatbot navigation that can be used in future initiatives on smart campus development and scalable digital support systems in higher education.

Keywords: Augmented reality, location-based, navigation, virtual chatbot

ARTICLE INFO

Article history:
Received: 23 October 2025
Accepted: 12 December 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.07

E-mail addresses:
laiyeejen@ymail.com (Lai Yee Jen)
eg.su@utm.my (Goh Eg Su)
johanna@utm.my (Johanna binti Ahmad)
* Corresponding author

INTRODUCTION

Navigation ability is an innate, essential, and crucial cognitive sense possessed by everyone. It is the ability to retain the sense of direction and location while roaming around in the surroundings.

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Spatial navigation can be dependent on internal representations obtained from sensory input or experience as well as externalised representations, including maps, diagrams, or signs (Cheng et al., 2024). Navigation to a new, unexplored area is a constant challenge. It necessitates an innovative solution to ease users into unpacking the potential of the new area, specifically the new technologies such as Augmented Reality (AR).

However, based on the previous studies (Naqvi et. al., 2024; Nguyen et al., 2021), navigation is not a focus with AR technology. Previous studies showed that AR is primarily involved in providing immersive and natural interactions in other aspects, except for navigation. Hence, the potential of AR for real-time navigation remains underexplored (Ling et al., 2024).

In this context, educational institutions were chosen as the main study cases due to the expansiveness of their campuses. This study was focussed on one institution, Universiti Teknologi Malaysia (UTM), specifically building N28 in the Johor Bahru UTM campus. It was reported that many people visiting the institution, including students, face difficulties in navigating their way around it. It is challenging for them, particularly visitors and new students, to navigate and obtain information about the locations of rooms where activities are held. This is due to their unfamiliarity with the institution's layout, which makes it difficult for them to get around due to the size of the campus and academic buildings. (Yashaswini et al., 2020).

Hence, the goal of this research is to develop a research prototype that employs a virtual guide with innovative AR interaction to boost the user's interest towards the campus navigation process. The research aim is to expose users to an AR-guided location-based chatbot to enhance their navigation experiences. To achieve this aim, the research objectives include studying the related research issues based on previous works, designing and developing the proposed navigation solution integrated with AR technology and an AI Chatbot to enhance the user experience, and lastly, evaluating the research prototype for further improvement. The proposed solution is expected to improve the navigation experience gained by the users through the integration of AR technology and a virtual Chatbot.

LITERATURE REVIEW

This literature review highlights and discusses technologies and studies relevant to AR and AI chatbots.

Augmented Reality (AR) Overview and Applications

Augmented Reality (AR) is a technology that superimposes digital objects or information on the physical world to improve the perception and interaction of the users with the real-world environments. It is realised using devices like smartphones, AR glasses, or head-

mounted displays that integrate cameras and sensors to track the surrounding environment and overlay digital content. This definition is like Su et al. (2018), which stated that AR is both interactive and acknowledgeable in three-dimensional (3D), combining virtual and real-world elements. As shown in Figure 1, merging real and virtual elements, aligning the real and virtual objects, and offering dynamic interactions between the real and virtual are the three (3) vital requirements that are satisfied by AR (*Dargan et al., 2022)*. Figure 1 illustrates the AR process, which allows users to view both virtual and real objects at the same time using a single interface. A blended interactive experience is produced, for instance, when a user views a real tree and a virtual giraffe superimposed in the same area.

AR uses cutting-edge localisation and sensing technologies to accomplish the integration of virtual content with the real-world environment. AR determines the precise location and orientation of the mobile device using a digital compass, accelerometer, GPS, and additional sensors (Cheng et al., 2024). These hardware elements enable the AR system to do a precise mapping of objects in the real world, spatially aligning them with virtual objects, and tracking consistency as the user moves. An AR system can be categorised into marker-based and markerless AR systems in general. In marker-based AR systems, visual cues such as physical AR printed markers, real-world existences like carpet, clothing, drawings are used to track, while for markerless AR systems, GPS, Simultaneous Localisation and Mapping (SLAM), or depth sensing are used to track and calibrate both virtual and real-world contents.

The technological advances of the recent past have enhanced the use of AR in education, navigation, and healthcare because visual information has become both context-sensitive and interactive. Spatial recognition applications, which include high-quality tracking

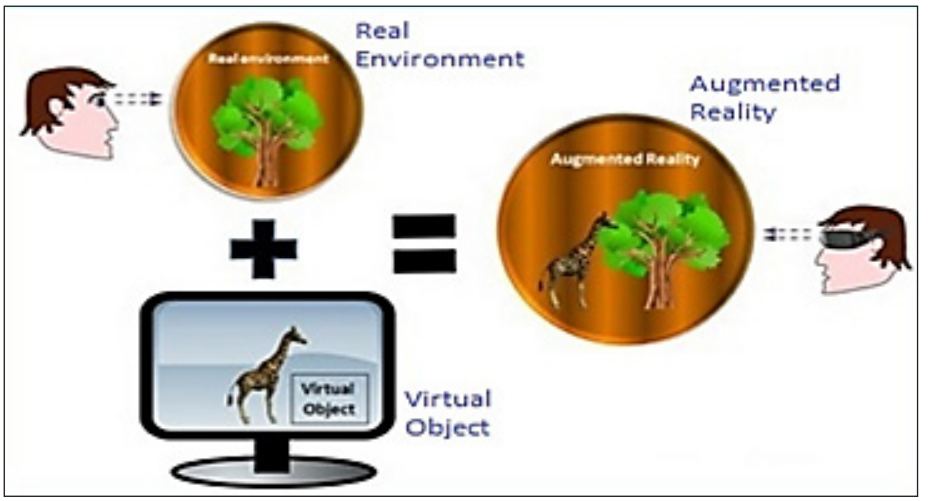


Figure 1. Vision of augmented reality (Dargan et al., 2022)

precision and 3D visualisation, have been developed with the assistance of built-in platforms like ARCore (Google) and ARKit (Apple) (Billinghurst et al., 2018). Research underscores the fact that AR is an excellent method to enhance spatial orientation, user engagement, and experience by integrating computer-based data and real-world spaces (Cheng et al., 2017; Dargan et al., 2022), which may improve users' immersive experiences.

AR encourages experiential learning and improves conceptual knowledge in learning environments. Zouhri and Mallahi (2024) revealed that AR, coupled with intelligent algorithmic learning, results in an interactive learning experience that improves student engagement. In the same way, AR-based interfaces in navigation systems like campus guides have been found to enhance location awareness and decrease the cognitive load that would be formed by conventional map interfaces (Lokhande et al., 2024). Nevertheless, AR navigation is still limited due to environmental factors, compatibility with hardware, and CPU demands (Garg et al., 2024).

Artificial Intelligence (AI) and Chatbots in Education and Guidance Systems

Artificial Intelligence (AI) technologies, particularly chatbots, have revolutionised human-computer interaction (HCI) as they allow natural, conversational interaction. AI chatbots are based on Natural Language Processing (NLP) and Machine Learning (ML) algorithms to analyse user inputs and deliver context-sensitive replies, as well as learning through experience (Sonawane et al., 2023). Such systems are common in learning systems, business support systems, and health care applications to create personalised support.

AI chatbots play roles in adaptive learning and have real-time feedback in higher education. As stated in the systematic reviews conducted by Labadze et al. (2023), the use of AI chatbots improves student engagement and accessibility by providing tutoring and information search based on personalisation. The use of conversational AI such as ChatGPT has already demonstrated good outcomes in encouraging critical thinking and self-directed study, yet there are ethical and accuracy problems (Dempere et al., 2023; Leelavathi & Surendhranatha, 2024).

AI has also been integrated into smart campus systems, where chatbots serve as virtual assistants that provide information on facilities on campus, directions, and administrative procedures. Additionally, the integration of generative AI into the educational setting has made the experience of students and instructors more dynamic and personalised through the involvement of digital platforms (Chukwuere, 2024).

Despite these developments, the state-of-the-art chatbot systems are usually text-based or voice-based, which do not emphasise spatial context utilisation. This restricts AR's potential capabilities in tasks such as real-world awareness or navigation, which leads to the possibility of combining AI with AR to offer a more immersive and context-sensitive experience.

Integration of AR and AI for Context-aware Interaction

The integration of AR and AI has become a paradigm shift that leads to increased interaction between users and situational cognition. AI contributes capabilities such as perception, decision-making, and personalisation in bringing the immersive visualisation and environmental perspective. These technologies combined result in a context-aware system that promises visual and conversational guidance.

Recent reports highlight that the introduction of AI chatbots with AR can transform learning and navigation. For example, Srilakshmi (2024) discovered that AR and AI chatbots enhanced the conceptual learning of geography among learners because they visualised the information and obtained feedback in real time. Similarly, Garg et al. (2024) established that AR-AI fusion can augment learning environments by increasing interactivity, motivation, and retention.

Storey and Wagner (2024) defined that AI-AR integration is an essential step to adaptive learning ecosystems, which makes it possible to create a personalised experience and transform it dynamically based on the needs of the user. Similarly, Naqvi et al. (2024) emphasised the role of AI-powered AR systems in clinical education, as they can create a simulated environment of a medical setting to enhance the involvement of the learners in the process and their levels of retention. Nevertheless, issues of privacy of information, ethical conduct of AI, and the optimisation of the computational needs of these integrated systems are still a challenge.

Even though the integration of AR-AI has become a trend in the field of education and healthcare, it has not been fully studied in the field of navigation. Current AR navigation systems tend to take up set visual objects without conversational intelligence, whereas chatbot applications are blind to the environment. Therefore, a hybrid of AR spatial features and AI conversational reasoning provides a new direction for intelligent, location-based navigation architecture.

To synthesise existing research on AR, AI, and chatbot applications across educational and navigation domains, Table 1 summarises ten (10) representative studies published between 2021 and 2024. These studies collectively demonstrate the growing interest in combining immersive visualisation with intelligent interaction. However, as shown in Table 1, most works remain domain-specific, lacking full integration with context-aware systems for location-based navigation.

Research Gap and Study Motivation

Based on the literature review, as summarised in Table 1, while AR and AI technologies have shown a high rate of development, the integration of AR and AI into real-time and location-based navigation is still minimal. Previous AR studies were dominated by the visualisation and immersion, whereas AI chatbots have been concerned with linguistic and cognitive assistance.

Table 1
Prior studies on AR, AI, and chatbot integration

Author (Year)	Focus	Technology Used
Garg et al., 2024	AR-AI synergy for immersive learning	AI chatbots, AR mobile application
Zouhri & Mallahi, 2024	AI and AR for interactive teaching	AI recognition, 3D AR objects
Naqvi et al., 2024	AI-enhanced AR for medical education	ARCore platform, AI model for anatomy
Storey & Wagner, 2024	Integration of AI for adult learning	Generative AI, adaptive content
Dempere et al., 2023	ChatGPT and AI applications in higher education	ChatGPT, GPT-based systems
Srinivasan et al., 2023	AI chatbots in nursing education	AI dialogue systems
Srilakshmi, 2024	AR and chatbot integration for geography education	Unity3D platform, AI chatbot
Labadze et al., 2023	Systematic review of AI chatbots in education	AI chatbots, NLP
Mageira et al., 2022	AI chatbot for language learning (CLIL)	AI chatbot, NLP, multimodal interaction
Rubio-Sandoval et al., 2021	Indoor navigation using AR and Semantic Web	AR mobile application, ontology-based model
Nguyen et al., 2021	AI chatbot for university admissions	Rasa framework, NLP, deep learning

There are still limited studies that have tried to integrate these features into a context-aware adaptive navigation tool.

Previous studies have largely explored the use of AR as either a fixed, stationary navigation tool or as a text-based chatbot-based aid (Srinivasan et al., 2023). As a result, a clear research gap in coming up with systems that integrate spatial awareness, visual interaction, and conversational intelligence to provide the user with an engaging experience is evident.

Therefore, this paper fills in the noticeable gap with the proposed Interactive Location-based Augmented Reality Virtual Chatbot system. The proposed system combines ARCore-based spatial visualisation and an AI-based chatbot engine, allowing users to get real-time, customised, and location-based directions in both visual and verbal ways. The proposed solution extends the development of the ever-expanding area of smart campus innovation to fill in the gap between virtual assistance and physical navigation, providing an easy-to-use, engaging, and more accessible navigation tool.

METHODOLOGY

The agile methodology, also known as the agile model, was applied in this research. Based on the selected methodology, the research work was divided into smaller iterations or

pieces without directly including long-term planning. As described by Kumar and Bhatia (2012), the research requirements and scope should be established at the beginning of the development phase. In line with the expansion of Agile methodologies over the years (Lazorenko & Krasnenko, 2020), it emphasises flexibility and users' involvement with the prototyping process, which is in line with the achievement of the research objectives. Users' feedback is required for continuous enhancement and improvement of the functionalities, such as AR and chatbot interactions, aligning more with the Agile methodology's workflow compared to other linear models, such as Waterfall (Maharao, 2022). Furthermore, Agile methodology was selected to be adapted in this research to ensure that each phase contributes directly and progressively to the research's objectives, which finally lead to a highly usable and responsive AR-based campus navigation prototype. The conceptual workflow of the research prototype is shown in Figure 2 and explained explicitly in the following subsections.

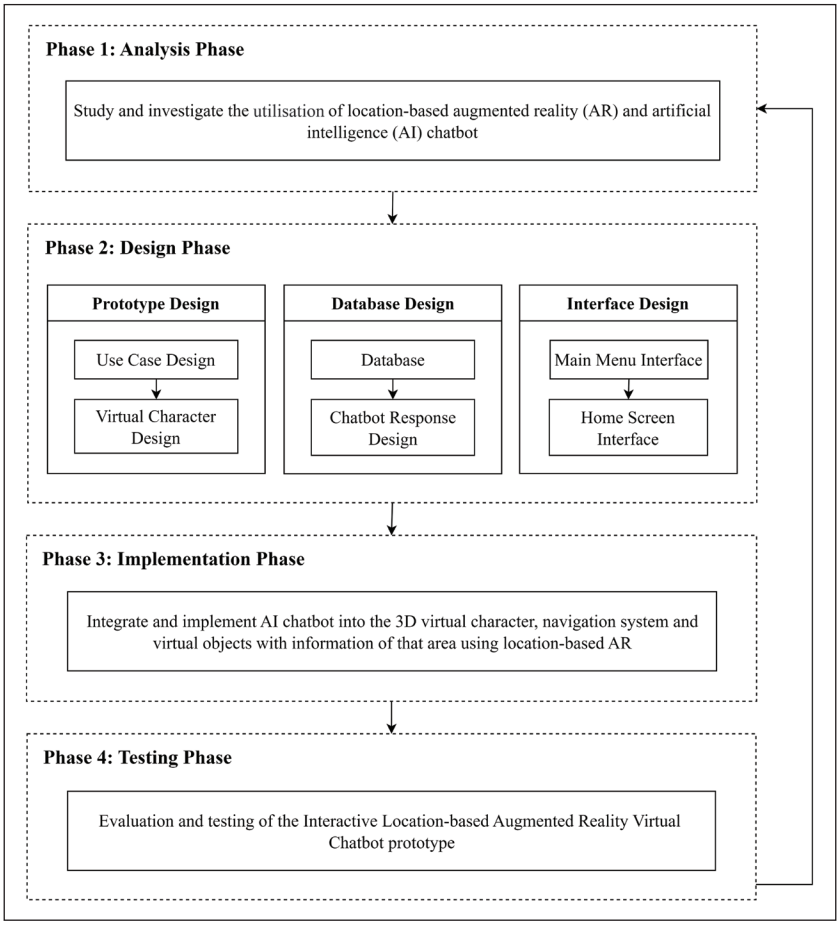


Figure 2. Research methodology

Analysis Phase

The initial phase of the research involved reviewing and gathering knowledge on location-based AR and AI chatbots. This phase consisted of studying key technologies, methods, and literature related to AR fundamentals, such as definitions, display options, types, and frameworks. Similarly, research into AI basics and applications, particularly chatbots, was conducted. Comparative analysis of existing studies is also undertaken to identify the research gaps generally and the research requirements specifically. This analysis helped to align the system requirements with the research objectives, especially identifying the problem and constraints of real-time AR navigation and AI chatbot in existing solutions.

Design Phase

In this phase, a research prototype was designed. This study constructed a use case diagram to identify the system narration, and the 3D virtual content was created for user interaction. Database design was conducted through entity relationship diagrams in storing the chatbot's instruction prompts about the real-world building as the navigated physical content and saving user feedback for future improvements. Extensively, interface design is conducted, consisting of the graphical user interface (GUI), which provides the functionalities to the user to launch the prototype and communicate as well as interact with the virtual chatbot for its intended purposes. Ensuring that the design achieves the research goals, such as intuitive user interaction and ease of navigation, is essential in this phase.

Implementation Phase

In this phase, the research prototype was employed with its essential functionalities in iterative cycles. The implementation tasks were divided into several specific functionality components, such as AR plane detection, AI chatbot, and navigation system. Each component integrated the features designed earlier for pilot evaluation on the component's performance, as well as adjusting based on the usability requirements and issues found. This approach ensured the seamless interaction between the technologies, allowing the research prototype to gradually progress towards accomplishing the research goal of creating a 3D AR virtual navigation guide with the ability to converse with the user.

Testing Phase

This is the final stage in the research methodology. In this phase, the research prototype's usability was evaluated based on the users' feedback and comments. Black box testing was used to identify usability issues without knowledge of the internal structure, focussing on system functionality. Users select a function, input a value, and verify if the system produces the desired results, following the standard suggested by Nidhra and Dondeti (2012) to ensure that the research prototype is aligned with the research objectives. Additionally, the user

feedback collected during the testing continually contributed to refining and improving the research prototype.

DESIGN PHASE

The use case diagram illustrating the prototype design is provided in Figure 3. Based on Figure 3, the user can control the prototype through the device to start or terminate the prototype and carry out user control through the provided GUI. The camera on the user's handheld device is used to capture the surroundings. By enabling the AR virtual chatbot, the prototype would prompt the user to tap on the detected surface to specify the location where the AR 3D virtual chatbot character exists on screen to assist the user.

The user can interact with the AR virtual character by asking questions regarding navigation and location. The prototype analyses the intention of the conversation or question asked by the user and provides responses and animated actions after extracting the respective information from the system-integrated database.

In addition, the prototype implements the location-based AR to display the precise location of AR virtual objects in a specific area. The AR virtual object displays the information of that area to broaden the knowledge of the user regarding the area. The user may capture the specific spot where the AR virtual object is embedded to display the AR virtual object. The user can also provide feedback about the research prototype through the GUI of the prototype for future suggestions or improvements.

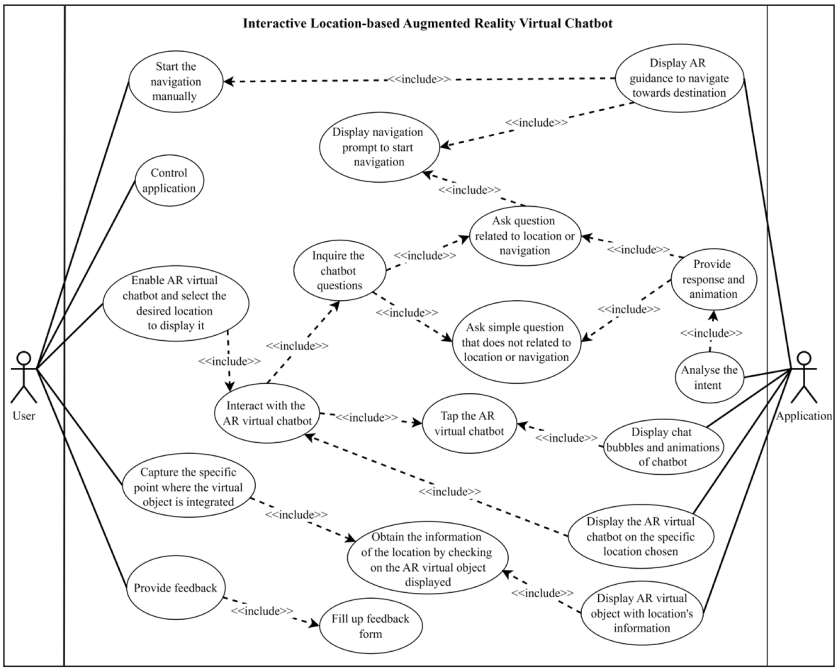


Figure 3. Use case diagram

IMPLEMENTATION PHASE

This section describes the implementation of the AR plane detection, chatbot, database handling, navigation, and 3D virtual objects with information.

Implementation of AR Plane Detection

When the AR virtual chatbot function is enabled, and the user taps on a detected surface, a 3D virtual chatbot character is placed at the selected location. This process requires AR plane detection, which uses the device's camera to identify flat surfaces using the ray-casting method, as shown in Figure 4.

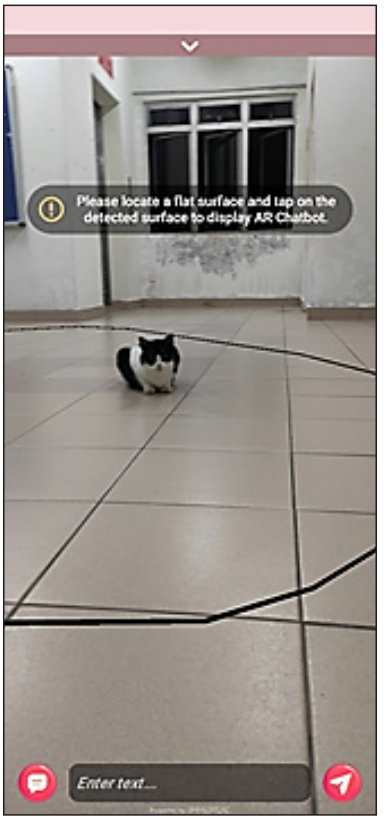
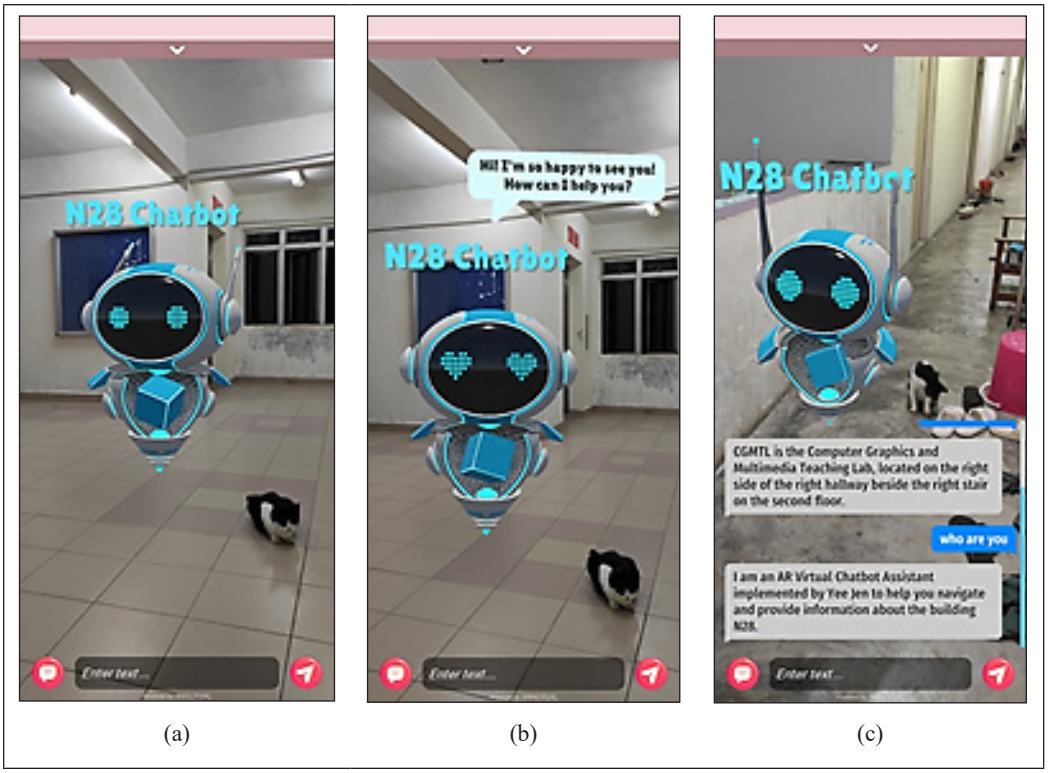


Figure 4. AR plane generated after detecting the plane

Implementation of Chatbot

The AR virtual chatbot is instantiated by detecting user touch input on the screen. Using the AR Raycast Manager, which is one of the built-in plugins provided in the prototype development platform. An intersection points between the touch raycast and the AR plane is created to instantiate the chatbot at the intersection point with an adjusted height and the camera-facing pose as shown in Figure 5a. After instantiation, the spawning of further chatbots was terminated. When the AR virtual chatbot is instantiated, users can interact with it, such as displaying an AR chat bubble with different messages and animations, as shown in Figure 5b.

The chatbot's responses are implemented using the OpenAI Generative Pre-Trained Transformer 4o (GPT-4o) API integrated with an NLP model. The instruction prompts are appended before user messages and sent to the chatbot. Responses are generated and returned containing information such as text, destination name, destination floor, and animation name mapped in the system's database. To trigger the animated content, responses are appended to the chat history to map animations that align with the corresponding responses. User messages chatbot responses are differentiated using a blue bubble and grey text, respectively, as shown in Figure 5c.



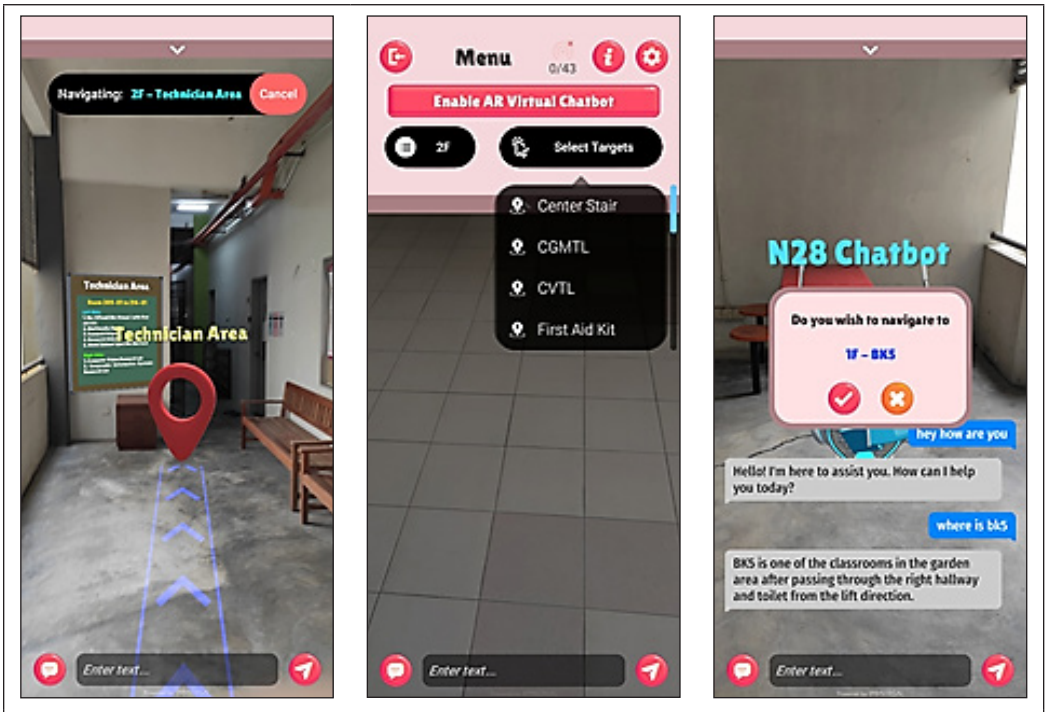
(a) (b) (c)

Figure 5. (a) Instantiation of AR virtual chatbot after tapping on detected plane; (b) Interaction of AR virtual chatbot after receiving input touch; (c) Chat bubbles generated from responses

Implementation of the Navigation System

Real-world environment is treated as a "marker" for AR experiences linked to real-world locations. The environmental feature points are captured and saved as point clouds, allowing the prototype to localise using this data. A floor map aligned with the point cloud facilitates creating a navigation mesh surface registered with 3D waypoint objects; each marked as a navigation target and placed on the floor map as shown in Figure 6a. By doing so, the system can generate AR navigation paths to guide users towards their desired destination.

The prototype can initialise navigation in two ways: manually through the provided menu interface or via chatbot interaction. As illustrated in Figure 6b, the menu provides an interface that allows users to select a floor level and choose the destinations of that floor from the dropdown menu to start navigation. Alternatively, users can ask the chatbot for directions, and it will provide destination details and a prompt to start navigation, as shown in Figure 6c. For destinations with similar names, like staircases, the closest one is chosen. A notification prompt appears upon reaching the destination.



(a) (b) (c)

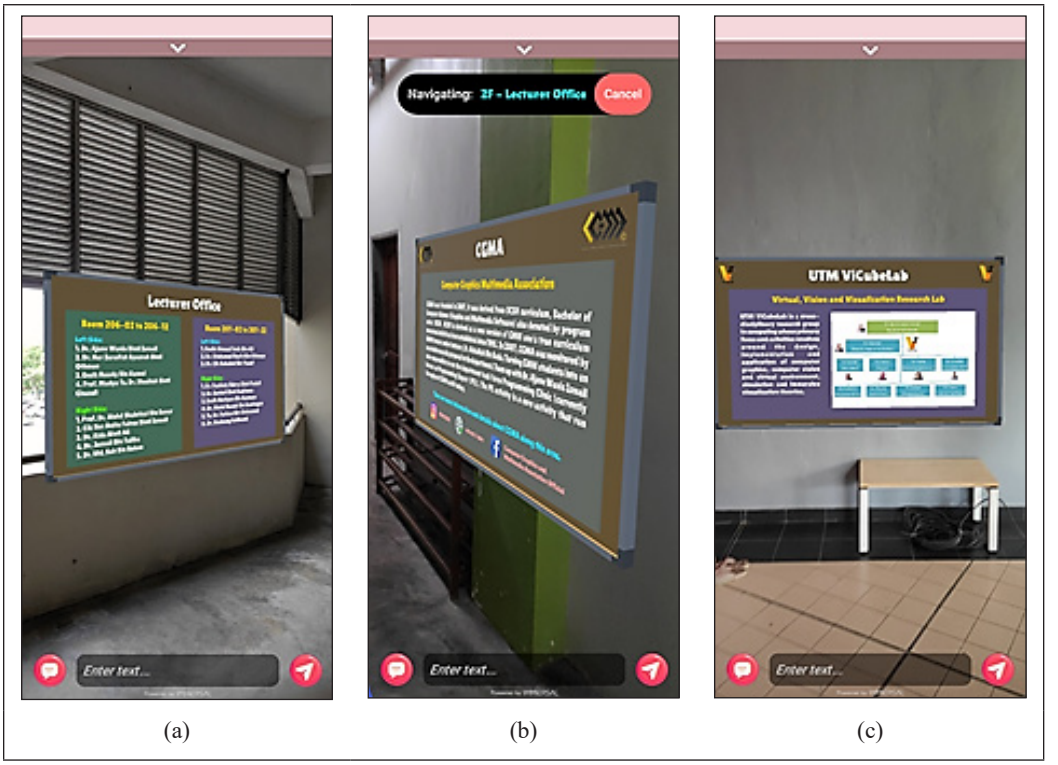
Figure 6. (a) Navigation path towards a 3D waypoint navigation target object; (b) Partial of available destination in dropdown menu for second floor; (c) Navigation prompt displayed after mentioning a destination

Implementation of AR 3D Virtual Objects

The environment contains various AR 3D virtual objects that display information related to specific destinations, immersing the user in the AR environment. These virtual objects appear in the environment when the user is detected to be close enough to the destination using the camera's collider, as illustrated in Figure 7. Based on Figure 7a, the AR 3D virtual object displays the information regarding the available office rooms of the lecturers in that area. In Figure 7b and Figure 7c, the information and knowledge regarding the laboratories are displayed in the 3D virtual object superimposed in their respective areas.

Implementation of Database

The prototype implements Firebase Realtime Database to store instruction prompts for chatbot responses and user feedback. Instruction prompts are retrieved when the prototype starts. These instruction prompts are combined before being sent to the chatbot. The user's feedback is also saved in the database for further prototype improvement purposes.



(a) (b) (c)

Figure 7. AR 3D virtual object shown near: (a) lecturer's office; (b) particular laboratory hallway; (c) entrance of a particular laboratory

RESULTS AND DISCUSSION

This section discusses the evaluation and testing of the research prototype, including the experimental setup and the input obtained from user testing. The user testing experimental setup requires a smartphone with the prototype installed and is conducted in building N28 of Universiti Teknologi Malaysia (UTM), specifically on the first and second floors. The user must hold the device upright to capture the surroundings to enable localisation. They also need to stay approximately 3 metres away from the environment to allow the prototype to recognise the captured images, as shown in Figure 8.

20 respondents from the UTM community were involved in user testing using the black box testing method. Black box testing was used to identify usability issues without knowledge of the internal structure, focussing on system functionality (Nidhra & Dondeti, 2012). Respondents were asked to fill in the questionnaires before and after the user testing to obtain the results. Before using testing, 75% of respondents have prior experience with augmented reality (AR), while 25% do not. All 20 respondents have interacted with chatbots, specifically with ChatGPT (100%), Copilot (75%), and Gemini (45%), but none with Perplexity.ai. Besides, all respondents have experience in using navigation

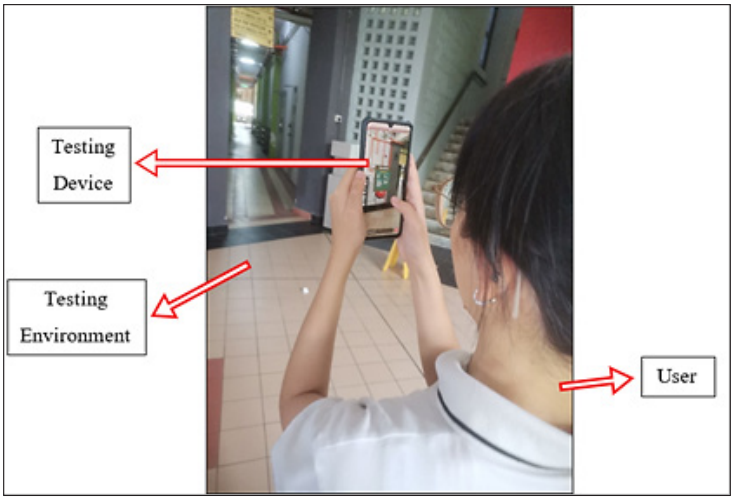


Figure 8. Experimental setup

applications, with a high percentage of 95% in using Waze and Google Maps. Notably, 90% of respondents have basically zero experience in using applications incorporating AR, chatbots, and navigation technology.

Post-test Questionnaire: Functional Requirements Test

The result of this post-test questionnaire demonstrates that this research prototype has successfully achieved and fulfilled its functional requirements. 90% and 85% of the respondents strongly agree that the research prototype has offered a consistent user interface, and the research prototype could capture the surroundings using the device's camera, respectively. The small proportion of users who did not strongly agree may have experienced minor device or interface transition issues, demonstrating potential for further improvement. 70% of the respondents strongly agreed that the AR 3D virtual chatbot was functioning correctly, and 55% of them confirmed that the virtual chatbot can be observed from any angle. Some users did not explore the chatbot from every angle since they interacted primarily with the features relevant to their current needs. Users often restrict their exploration to what appears to be the easiest when they face a time constraint or are unsure about the interface.

Additionally, 90% of the respondents strongly agreed that the chatbot interacts with effective English and handles the inquiries well, regardless of location and non-location inquiries. This shows that users have a high level of confidence in the chatbot's capability to understand a variety of requests and its clear language, which is important for building trust and sustained engagement. Navigation features like the chatbot provided accurate navigation guidance were strongly agreed upon by 70% of the respondents. It shows that

most users considered the navigation support reliable and helpful, suggesting that the chatbot's navigation and direction responses are operating properly. Lastly, 85% of the respondents strongly agreed that the AR virtual objects in the research prototype were displayed correctly and provided useful information about that region. The results show the reliability of the technical aspects of the AR features, which successfully provide users with new knowledge about the environment through the AR features. Additionally, the minor percentage might be due to the rendering problems or lack of familiarity with AR features, which did not strongly approve, suggesting room for improvement to enhance the users' experience.

Overall, the results from this functional requirement test indicated that the primary functional elements, such as AR visualisation, chatbot interaction, and navigation, have performed reliably well under real conditions. These high ratings also illustrated that the research prototype is very user-friendly, despite most of the users having no prior experience or exposure to applications combining these technologies. However, the lower percentage of observing the virtual chatbot from any angle implies that inspection from different perspectives was not as naturally prompted for navigation usage.

Post-test Questionnaire: System Usability Scale (SUS)

The System Usability Scale (SUS) is a commonly used assessment tool for evaluating a system's usability because of its simplicity and effectiveness. This scale provides precise and rapid usability assessments with a result from 0 to 100 (Brooke, 2013). The results showed that the users had full satisfaction with the research prototype; 100% of them agreed that they would like to use the prototype frequently. 95% of the respondents found that the prototype is straightforward and it is not unnecessarily complex, but instead, easy to use.

Furthermore, 70% of the respondents strongly disagreed that they required any technical support while using the prototype. This shows that the prototype is highly user-friendly. All the respondents agreed that the functions of the prototype are well integrated. Consistency was noted, with 55% strongly disagreeing and 45% disagreeing that there is too much inconsistency. Additionally, 80% strongly agreed, and 20% agreed that most users could quickly learn to use the prototype; 70% strongly disagreed that the prototype was challenging to use, and users needed to learn a lot of things before being able to use the prototype; and lastly, 60% strongly agreed and 40% agreed that they felt confident using it.

SUS provides a single composite score indicating overall system usability. The questionnaire uses a 5-point scale, with scores processed to range from 0 to 4. For odd-numbered questions, the score is the scale position minus one, and for even-numbered questions, it is five minus the scale position. The total score is then multiplied by 2.5 to yield a final SUS score between 0 and 100. According to Sauro (2011), the average SUS score is 68, with scores above 68 considered above average and indicating good usability. The final score is calculated by averaging the scores of 20 respondents.

Based on Table 2, the final SUS score is 90.25, which is significantly higher than the average score of SUS 68. When compared to the acceptable SUS benchmark of 68, which represents average usability across most systems, the score of 90.25 shows that the prototype has 'Excellent' usability Lewis and Sauro (2018). It proves that the prototype is easy to use and provides users with a comfortable and very positive experience when interacting with it, and that the prototype performs beyond the standard usability requirements. Therefore, it can be concluded that the research prototype of the Interactive Location-based AR Virtual Chatbot is exceptionally usable and offers an excellent user experience.

Table 2
Result of system usability scale (SUS)

User	Question										SUS Score
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
1	3	3	3	4	4	3	4	3	4	4	87.50
2	3	3	4	4	3	3	4	4	3	4	87.50
3	3	2	3	3	3	3	3	3	4	3	75.00
4	3	4	4	3	4	4	4	2	4	4	90.00
5	4	4	4	4	4	4	4	4	4	3	97.50
6	4	3	3	3	4	4	3	3	3	4	85.00
7	4	3	3	4	4	4	4	4	4	3	92.50
8	3	3	4	3	3	4	4	4	3	4	87.50
9	3	3	3	4	4	4	4	4	4	4	92.50
10	3	4	4	4	4	4	4	3	4	4	95.00
11	3	4	4	4	3	4	4	4	3	4	92.50
12	3	4	4	4	4	3	4	4	4	4	95.00
13	4	3	3	4	4	3	4	4	4	3	90.00
14	3	3	4	4	3	3	4	4	3	3	85.00
15	4	3	3	3	4	4	3	4	3	3	85.00
16	4	3	4	3	4	4	4	4	4	4	95.00
17	4	4	4	4	3	3	4	4	3	4	92.50
18	4	4	4	4	4	4	4	4	4	4	100.00
19	4	3	4	4	4	3	3	4	3	4	90.00
20	4	3	3	4	4	3	4	3	4	4	90.00
									Mean SUS Score		90.25

User Acceptance Testing

This section demonstrated the results of the user acceptance testing that was carried out with black box testing. This testing also involved 20 respondents interacting with the research prototype without any prior knowledge or instructions. According to the result, all 20 respondents have accomplished main functionalities such as superimposing the AR chatbot in the surroundings, interacting and asking questions with it, as well as starting navigation. However, some actions were less frequently completed: 5 respondents did not read the user guide, over half did not inspect the AR chatbot by walking around it, and a few did not tap the chatbot for animations. Additionally, approximately 50% did not start navigation manually, and some did not follow navigation guidance to completion.

These trends show that the essential functionalities were intuitive and naturally performed, whereas the optional features lacked adequate visibility and prompting. For example, the lack of engagement with the optional features, such as inspecting the AR chatbot model from different perspectives and its animation, indicates that the users may not recognise the value of these features in supporting the navigation. This can lead to some possible design improvements, such as adding some subtle hints and indications or progressive demonstration of the features. This may encourage the users to explore and play around the interface or features and enhance their overall experience. Table 3 shows the partial result of the user acceptance testing mentioned.

Table 3
Partial result of user acceptance testing

No	Action	Expected	Question									
			1	2	3	4	5	6	7	8	9	10
1	Enter prototype	The user taps the Start button in main menu.	√	√	√	√	√	√	√	√	√	√
2	Follow tutorial	The user reads all instructions in the tutorial.	√	√	√	√	√	√	√	√	√	√
3	Open menu	The user taps on top bar to open the menu.	√	√	√	√	√	√	√	√	√	√
4	Close menu	The user taps on area outside of menu or top bar arrow to close the menu.	√	√	√	√	√	√	√	√	√	√
5	Check user guide	The user taps on the user guide button to display it and reads it.	√	√	√	√	√	x	√	√	√	√
6	Enable AR virtual chatbot	The user taps on the button to enable AR virtual chatbot.	√	√	√	√	√	√	√	√	√	√
7	Superimpose AR virtual chatbot	The user taps on the detected surface to superimpose the AR virtual chatbot.	√	√	√	√	√	√	√	√	√	√
8	Observe AR virtual chatbot	The user walks around to observe the AR virtual chatbot from any angle.	x	x	x	√	√	x	x	√	x	x

CONCLUSION

This research developed a prototype consisting of an interactive location-based AR virtual chatbot aimed at enhancing user experience by eliminating the need for human guides. The main objectives of this research have been achieved, as the AR virtual chatbot has performed its function to assist users with navigation and enquiries. This research helps address gaps in AR-based navigation research in making an endeavour to reach the unexplored knowledge of the AR application potential. Through the case study in this research, it opens a new opportunity for further research in studying the different interactions, such as tangible and multimodal interaction probabilities in enhancing the navigation experiences among the users derived with the assistance of an AI intelligent system. There are several limitations found from this research, including the occasional inaccuracies in the responses provided by the virtual chatbot due to its AI technology that can make mistakes, and its AR navigation due to GPS constraints. Besides that, the compatibility that is restricted to Android smartphones that come with features that support AR, text-only interactions with the chatbot, and finally, the need for a stable internet connection are also limitations discovered from this research. However, the intention of the research encroaches these limitations, which attempts to enhance the existing navigation system with the emerging technologies such as AR and AI. It not only solves the navigation system's weaknesses, which were limited to providing users with location data, but it also enhances the content with engaging AI chatbots and insightful augmented reality visualisations. On the other hand, various additional features can be introduced to increase the prototype's usability further, such as expanding the navigation coverage to include multiple facilities across the campus. In addition, providing various language options for native users and combining voice-based interaction with the chatbot may dramatically improve user engagement and overall efficacy inside the campus environment.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2022/ICT10/UTM/02/1).

REFERENCES

Anireddy Srilakshmi. (2024). Integrating augmented reality and artificial intelligence to enhance the spatial ability skills in the field of geography. *International Journal of Intelligent Systems and Applications in Engineering, 12*(4), Article 201. https://ijisae.org/index.php/IJISAE/article/view/6194

Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. *Foundations and Trends in Human-Computer Interaction, 8*(2-3), 73-272. https://doi.org/10.1561/1100000049

Brooke, J. (2013). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189-194). Taylor & Francis.

Cheng, C. L. Y., Su, G. E., Ahmad, J. B., & Sutikno, T. (2024). Game-based augmented reality learning of Sarawak history in enhancing cultural heritage preservation. *Indonesian Journal of Electrical Engineering and Computer Science, 34*(3), 1718-1729. https://doi.org/10.11591/ijeecs.v34.i3.pp1718-1729

Cheng, J. C. P., Chen, K., & Chen, W. (2017, July 4-7). Comparison of marker-based AR and markerless AR: A case study on indoor decoration system. In *Proceedings of the Joint Conference on Computing in Construction (JC3), Volume I* (pp. 483-490). Heraklion, Greece. https://doi.org/10.24928/JC3-2017/0231

Chukwuere, J. E. (2024). The future of generative AI chatbots in higher education. *arXiv*. https://doi.org/10.32388/UE841K

Dargan, S., Bansal, S., Kumar, M., Mittal, A., & Kumar, K. (2022). Augmented reality: A comprehensive review. *Archives of Computational Methods in Engineering, 30*(2), 1057-1080. https://doi.org/10.1007/s11831-022-09831-7

Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education, 8*, Article 1206936. https://doi.org/10.3389/feduc.2023.1206936

Garg, N., Kaur, A., & Dutta, R. (2024). Immersive learning: Evaluating the effectiveness of AI chatbots and AR in education. In *Proceedings of the 3rd IEEE Delhi Section flagship conference* (pp. 1-7). https://doi.org/10.1109/DELCON64804.2024.10866724

Kumar, G., & Bhatia, P. K. (2012). Impact of agile methodology on software development process. *International Journal of Computer Technology and Electronics Engineering (IJCTEE), 2*(4), 46-50.

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education, 20*(1), Article 42. https://doi.org/10.1186/s41239-023-00426-1

Lazorenko, L., & Krasnenko, O. (2020). Applying agile learning to teaching English for specific purposes. *International Journal of Learning, Teaching and Educational Research, 19*(9), 238-258. https://doi.org/10.26803/ijlter.19.9.13

Leelavathi, R., & Surendhranatha, R. C. (2024). ChatGPT in the classroom: Navigating the generative AI wave in management education. *Journal of Research in Innovative Teaching & Learning*. Advance online publication. https://doi.org/10.1108/JRIT-01-2024-0017

Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the System Usability Scale. *Journal of Usability Studies, 13*(3), 158-167.

Ling, C. L., Su, G. E., & Yong, P. Y. (2024). Multi-user interaction for living room furniture layout design in handheld augmented reality. *ASEAN Engineering Journal, 14*(3), 161-167. https://doi.org/10.11113/aej.v14.21372

Lokhande, J., Bhuvad, A. S., Bendre, A., Bath, R., Bodhankar, P., Rao, M., & Idhate, S. (2024). Campus navigation using augmented reality and virtual reality. *International Journal for Multidisciplinary Research, 6*(3), Article 19504. https://doi.org/10.36948/ijfmr.2024.v06i03.19504

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences, 12*(7), Article 3239.https://doi.org/10.3390/app12073239

Maharao, C. S. (2022). A comparative analysis of agile, waterfall, and hybrid methodologies in software project success. *ShodhKosh: Journal of Visual and Performing Arts, 3*(2), 917-926.https://doi.org/10.29121/shodhkosh.v3.i2.2022.3396

Naqvi, W. M., Sundus, H., Mishra, G., Muthukrishnan, R., & Kandakurti, P. K. (2024). AI in medical education curriculum: The future of healthcare learning. *European Journal of Therapeutics, 30*(2), e23-e25.https://doi.org/10.58600/eurjther1995

Nguyen, T. T., Le, A. D., Hoang, H. T., & Nguyen, T. (2021). NEU-chatbot: Chatbot for admission of National Economics University. *Computers and Education: Artificial Intelligence, 2*, Article 100036.https://doi.org/10.1016/j.caeai.2021.100036

Nidhra, S., & Dondeti, J. (2012). Black box and white box testing techniques: A literature review. *International Journal of Embedded Systems and Applications (IJESA), 2*(2), 29-50. https://doi.org/10.5121/ijesa.2012.2204

Rubio-Sandoval, J. I., Martinez-Rodriguez, J. L., Lopez-Arevalo, I., Rios-Alvarado, A. B., Rodriguez-Rodriguez, A. J., & Vargas-Requena, D. T. (2021). An indoor navigation methodology for mobile devices by integrating augmented reality and semantic web. *Sensors, 21*(16), Article 5435. https://doi.org/10.3390/s21165435

Sauro, J. (2011, February 3). Measuring usability with the System Usability Scale (SUS). *MeasuringU*.https://measuringu.com/sus/

Sonawane, P., Kunekar, P., Bhatele, P. R., Sonkusare, S., Sonar, R., Bisen, S., & Sonar, K. (2023). ConvoMind: A user-friendly AI chatbot. *International Journal for Research in Applied Science and Engineering Technology, 11*(12), 1-5. https://doi.org/10.22214/ijraset.2023.57180

Srinivasan, M., Venugopal, A., Venkatesan, L., & Kumar, R. (2023). Navigating the pedagogical landscape: Exploring the implications of AI and chatbots in nursing education. *JMIR Nursing, 7*, Article e52105. https://doi.org/10.2196/52105

Storey, V. A., & Wagner, A. (2024). Integrating artificial intelligence (AI) into adult education. *International Journal of Adult Education and Technology, 15*(1), 1-15. https://doi.org/10.4018/IJAET.345921

Su, G. E., Sunar, M. S., Andias, R., & Ismail, A. W. (2018). An inertial device-based user interaction with occlusion-free object handling in a handheld augmented reality. *International Journal of Integrated Engineering, 10*(6), 159-168. https://doi.org/10.30880/ijie.2018.10.06.023

Yashaswini, R., Mohith Prajwal, M., Nesar, H., Aishwarya, D. N., & Soumya, B. S. (2020). Campus navigation using augmented reality. *International Journal of Engineering Technology Research & Management (IJETRM), 4*(8), 130-136.

Zouhri, A., & Mallahi, M. E. (2024). Improving teaching using artificial intelligence and augmented reality. *Journal of Automation, Mobile Robotics & Intelligent Systems*, 57-61. https://doi.org/10.14313/JAMRIS/2-2024/13

Pertanika J. Sci. & Technol. 34 (2): 763 - 786 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Ensemble Docking and Molecular Dynamics Pattern Similarity Analysis for the Discovery of Dengue Virus RNA-Dependent RNA Polymerase Inhibitor

Amaal Mohammed Salih Nasr[1], Mohd Basyaruddin Abdul Rahman[1], Muhammad Alif Mohammad Latif[1,2], Faridah Abas[3], Siti Munirah Mohd Faudzi[1,4], and Bimo A. Tejo[1]*

[1]Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Centre for Foundation Studies in Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[3]Department of Food Science, Faculty of Food Science and Technology,Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[4]Natural Medicine and Product Research Laboratory, Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

Dengue fever, transmitted by Aedes mosquitoes, remains a major public health concern across tropical and subtropical regions. The urgent need for effective antiviral agents has drawn significant attention to the NS5 RNA-dependent RNA polymerase (RdRp), a key enzyme required for viral replication. This study aimed to identify promising RdRp inhibitors using ensemble docking, molecular dynamics (MD) simulations, and MD pattern similarity analysis. A 150-ns MD simulation of both the NS5 RdRp apoprotein and its ligand-bound complex was conducted to assess conformational stability and dynamic behaviour. Ligand binding reduced overall protein flexibility, indicating stabilising interactions suggestive of inhibitory activity. Structural and dynamic features were examined through RMSF, SASA, Rg, and dynamic cross-correlation matrix (DCCM) analyses, which collectively revealed distinct residue correlation patterns upon ligand binding. Clustering of the MD trajectories generated 225 representative conformations, 138 from the apoprotein and 87 from the complex used for ensemble docking of 328 curcuminoid compounds. ADMET analysis further showed that compounds 1CJS and 90 possessed favourable absorption and metabolic properties, with low predicted toxicity. Among all screened ligands, compound 90 demonstrated the strongest binding affinity.

ARTICLE INFO

Article history:
Received: 03 September 2024
Accepted: 10 December 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.08

E-mail addresses:
gs59663@student.upm.edu.my (Amaal Mohammed Salih Nasr)
basya@upm.edu.my (Mohd Basyaruddin Abdul Rahman)
aliflatif@upm.edu.my (Muhammad Alif Mohammad Latif)
faridah_abas@upm.edu.my (Faridah Abas)
sitimunirah@upm.edu.my (Siti Munirah Mohd Faudzi)
bimo.tejo@upm.edu.my (Bimo A.Tejo)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Subsequent MD simulations confirmed its stable binding mode and favourable dynamic profile. MD pattern similarity analysis also showed that compound 90 closely mimics the structural and dynamic behaviour of the native ligand GTP, supporting its potential as a potent RdRp inhibitor. Overall, these findings highlight compound 90 as a promising curcuminoid derivative for further development of antiviral agents against the dengue virus.

Keywords: Drug resistance, dengue, molecular docking, molecular dynamics, RdRp, structure-based drug discovery, virtual screening

INTRODUCTION

Dengue virus (DENV), a Flavivirus genus member, poses a major global health threat with four serotypes (DENV-1 to DENV-4) sharing 60%-70% amino acid identity. Each serotype includes distinct genotypes with less than 3% sequence variation (Weaver & Vasilakis, 2009). The virus carries a ~11 kb positive-sense, single-stranded RNA genome that encodes a single open reading frame (ORF) flanked by 5' and 3' untranslated regions, as shown in Figure 1 (Khetarpal & Khanna, 2016). This ORF translates into 10 proteins: three structural (C, E, prM) and seven non-structural (NS1-NS5). Structural proteins play a crucial role in the virus's assembly, maturation, and entry into the host cell by forming the capsid and viral envelope (Qureshi, 2020). In contrast, non-structural proteins are essential for sustaining the viral life cycle, as they facilitate RNA replication, enzymatic processing, and immune evasion. Additionally, non-structural proteins can remodel specific host cellular pathways to enhance viral infection. These coordinated functions are vital for the successful replication of the dengue virus.

Due to their role in RNA virus replication, RNA-dependent RNA polymerases (RdRps) serve as crucial anti-viral drug targets (Singh et al., 2020). They generate viral

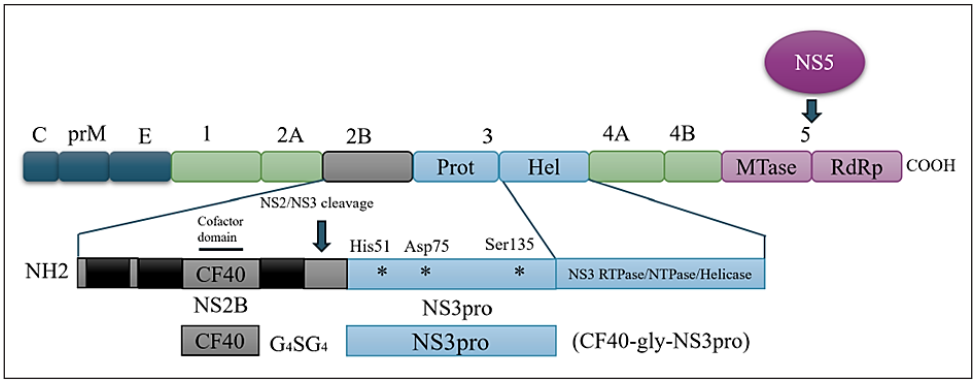


Figure 1. Schematic representation of the genome of DENV, consisting of structural (capsid (C); membrane (M); envelope (E); and non-structural (NS1-NS5) proteins

RNA using an RNA template, catalysed by conserved catalytic residues such as Trp795, Thr794, Ser710, and Arg729, directly interacting with GTP Figure 2 (Yap et al., 2007). These residues are essential for catalysis and substrate recognition. Lim et al. (2018) also identified an allosteric "N-pocket" at the thumb/palm interface, in addition to the active site, using fragment-based crystallographic screening. Fragment-based optimisation of portions of the molecules that address these regions led to potent inhibitors of biphenyl acetic acid. This two-site mechanism offers more options for developing antiviral therapy and further explores the space of RdRp druggability.

Currently, no specific antiviral drugs have been approved for the treatment of dengue. The current management of dengue primarily involves supportive care, such as hydration and fever management, rather than targeting the virus directly. This gap underscores the urgent need for effective antiviral agents that can directly inhibit the replication of DENV.

Due to their ability to target various biological processes without side effects, curcuminoids have attracted therapeutic interest in treating immune, metabolic, and cancer diseases (Mahmood et al., 2015; Siviero et al., 2015). These curcuminoids have demonstrated antibacterial (Prlina et al., 1958), antimalarial (Prlina et al., 1958), diuretic (Douglas & Rubin, 1961), antirheumatic (Markham, 2017) and antiretroviral activities (Adkins & Faulds, 1998). The term "curcuminoids" is the appropriate choice because the context suggests a broader examination of the therapeutic potential of turmeric-derived compounds rather than focusing solely on curcumin.

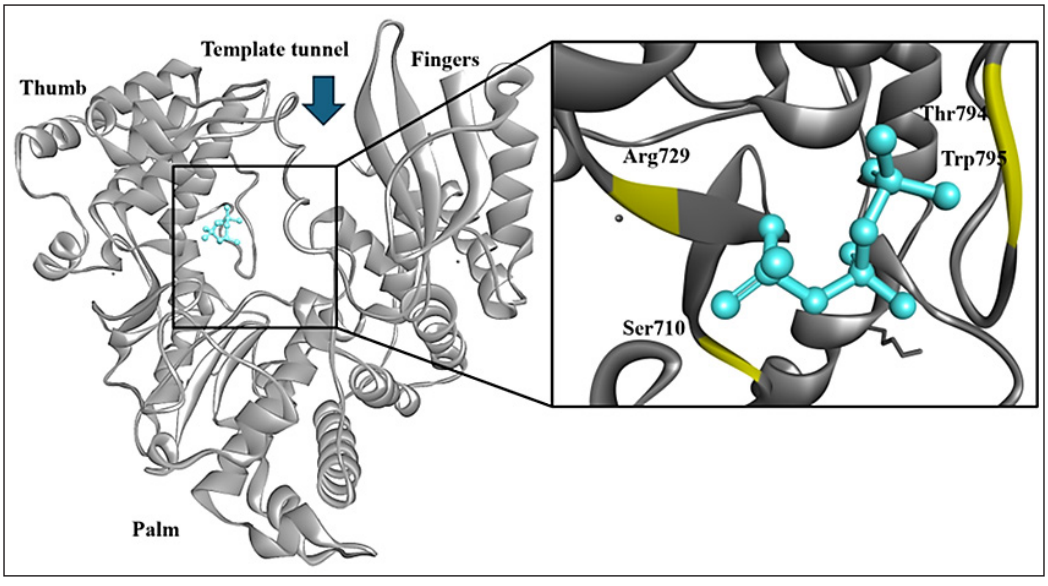


Figure 2. The RdRp enzyme's essential elements and active site

Curcuminoids are increasingly recognised as promising inhibitors of dengue virus, as they can target both viral proteins and host pathways essential for viral replication. Several studies show that curcumin and its derivatives inhibit the dengue NS2B-NS3 protease by binding to an allosteric pocket and disrupting the β-hairpin structure required for enzymatic activation, ultimately blocking viral polyprotein processing (Lim et al., 2020). Curcuminoid analogues have also demonstrated stronger antiviral activity than curcumin by suppressing lipid-biosynthesis genes such as ACC and FAS and by interfering with actin polymerisation, a key process for viral entry and intracellular movement (Balasubramanian et al., 2019). Curcumin additionally reduces DENV-2 plaque formation at non-toxic levels, supporting its potential as a safe antiviral scaffold (Padilla-S et al., 2014). The potential of curcuminoids as dengue RdRp inhibitors remains largely unexplored.

This study uniquely integrates ensemble docking with MD pattern similarity analysis to identify dengue RdRp inhibitors from our curcuminoid collection. By screening 328 curcuminoid compounds against multiple RdRp conformations from a 150-ns MD simulation, we accounted for the protein's dynamic flexibility. Top ligands were evaluated using RMSD, RMSF, Rg, and DCCM to assess their dynamic similarity to the native ligand GTP. This dual approach enhanced the accuracy of identifying strong binders with GTP-like behaviour. Arg729 and Trp795 emerged as key anchors in stabilising lead compounds. Additionally, the application of curcuminoids created a new chemical area for antiviral targeting.

METHODOLOGY

Target Protein and Ligand Preparation

The three-dimensional crystallised structure of DENV RdRp complexed with GTP (PDB ID: 2J7U) was taken from the Protein Data Bank (Berman et al., 2000) and processed using Play-Molecule Protein-Prepare web service to have it ready for molecular dynamics (MD) simulations (Martínez-Rosell et al., 2017). The web platform facilitates the automation of the addition of hydrogen atoms, removal of crystallographic water molecules, a nd adjustment of the protonation state, to ensure chemical accuracy. It also clears up typical structure problems, such as the disulfide bond recognition and the absence of the side chain, adjusts disulfide bonds, pH, and the concentration of ions in a customisable way. Through the optimisation of the structure, the model becomes applicable to the accurate simulation of the behaviour of small molecules and, more in-depth analysis of the interaction of ligands can be undertaken as well as creating high-quality conformational ensembles.

Molecular Dynamics Simulations

The RdRp was simulated in an apoprotein and GTP-bound state using AMBER14 force field (Maier et al., 2015) implemented in YASARA Structure (v21.12.19) (Krieger & Vriend, 2015).

The simulations used the Particle Mesh Ewald technique of accurate long-range electrostatics and periodic boundary conditions at 300 K and 0.9% NaCl. Solvation was done in a cubic simulation cell extended 20 Å beyond the protein surface. Temperature in the model was controlled with the help of a Berendsen thermostat to maintain system stability throughout the simulation.

The protein complex underwent a thorough preprocessing scenario prior to the simulation, during which fixing of energy by the use of steepest decline algorithm on 5000 cycles and the cleaning of structures (Brooks et al., 1983) were used in attaining stable conformation and broadening saddles (Hassan et al., 2022). A 1.25 fs time step was used to model the simulations, with a 100 ps interval used to record the trajectories. Analyses were conducted by utilising the default macro md_analyse.mcr to verify the system's structural stability and dynamic characteristics.

Generation of Ensemble Structure

The ensemble structures were constructed by clustering protein conformations by backbone RMSD with a cutoff of 2.0 Å using the YASARA macro md_analyse.mcr. This generated 225 clusters, with 87 deriving from the RdRp complex and 138 from the apoprotein. In the case of ensemble docking, every cluster was identified by representative structures and saved in PDB format. The 150-ns production MD simulations were performed in duplicate, aimed at determining the structural changes induced by the ligand, and to represent conformational diversity of both unbound and bound forms. The dynamic stability of proteins and ligand binding flexibility were comprehensively understood due to this dual strategy.

Ligand Preparation

OpenBabel (O'Boyle et al., 2011) was used to convert ligand structures into SDF format. The ligands were subjected to a 20-step conjugate gradient minimisation after a 200-step steepest descent minimisation with a step size of 0.02 using YASARA Structure (Krieger & Vriend, 2015).

Protein-ligand Molecular Docking

This study tested 328 curcuminoid compounds from the Universiti Putra Malaysia (UPM) natural product collection against the dengue virus RdRp. After ligand preparation, 225 representative RdRp conformations produced by clustering analysis and molecular dynamics simulations were used for ensemble docking. Docking grids were established using known active site coordinates, and these conformers, which represented apoprotein and ligand-bound states, were aligned and superimposed to preserve uniform orientation.

Using the Lead Finder engine in Standard Precision mode, which is tuned for quick, high-throughput screening, virtual screening was carried out in Cresset Flare (v5.0.0). The approach was designed to balance speed and accuracy in order to evaluate big libraries efficiently. An Intel® Core™ i5-11300H CPU, 16 GB of RAM, and a 64-bit operating system was used for all calculations. For every ligand, docking results yielded virtual screening (VS) scores and binding free energies (ΔG). To evaluate their potential as RdRp inhibitors, top-scoring compounds were chosen for additional examination of binding interactions, stability, and MD-based dynamic similarity to GTP (Stroganov et al., 2009).

Evaluation of Drug-like Properties using ADME Analysis

After completing the docking study, we analysed the selected compounds using the SwissADME web tool (https://www.swissadme.ch/) to evaluate their drug-like properties. This allowed us to check their stability, solubility, absorption, metabolism, and potential toxicity. The compounds' chemical structures were first converted into SMILES format using Open Babel, which is required for the analysis. This step helped us understand how the compounds might behave in the human body and identify the most promising candidates for further study.

Molecular Dynamics Simulations of Selected Curcuminoids

After ensemble docking, molecular dynamics (MD) simulation was used to determine the stability of high-ranking ligand-protein complexes and investigate the effect of ligand-induced binding on RdRp structural dynamics under close to physiological settings. MD simulations were carried out with the AMBER14 force field implemented in YASARA Structure (version 21.12.19) with the simulation duration of 150 ns. At 300 K, the systems were solvated in a TIP3P water box with 0.9% of NaCl. To critically assess stability and strength of interactions, analyses were carried out on radius of gyration (Rg), dynamic cross-correlation matrix (DCCM), estimation of binding free energy through MM/PBSA, root mean square deviation (RMSD) and root mean square fluctuation (RMSF).

RESULTS AND DISCUSSION

Molecular Dynamics Simulations

The present work has generated an ensemble of RdRp conformations used in conducting structure-based virtual screening using molecular dynamics (MD) simulation. These simulations based upon the AMBER14 force field implemented in YASARA Structure proved the dynamic features of the RdRp over a time of 150 ns. To model physiological conditions, the simulation environment was made up of explicit water and 0.9 % NaCl at a temperature of 300 K. The essence of the ensemble docking strategy was extracting

appropriate representative conformations and proteins in the generated trajectories through clustering analysis. These different conformers allowed more accurate docking because they considered the receptor's conformational flexibility, which is usually ignored in the single-structure docking methods.

Conversion of Trajectory Data into Conformational Ensemble

Obtaining the ensemble of the conformations of proteins through molecular dynamics (MD) trajectory data is essential to depict how proteins change their structure over time. The conformations were analysed using YASARA command md_analyse.mcr, and then clustered by backbone RMSD values. Applying this clustering method, two protein variations were represented with 225 distinct conformations, including 87 conformations of the RdRp complex and 138 conformations of the apoprotein. An RMSD cutoff was set at 2.0 Å to ensure we only grouped structurally similar frames. This permitted a systematic study of the dynamics of the protein.

To maintain diversity within the clusters, a minimal separation ('rmsdmin') of 2.0 A was set between structural positions of each pair of the sample structures. The dataset was structurally unique because each cluster was represented by one single conformation stored as "Protein_cluster*yob". Such a criterion allowed a rich set of structural motifs to be expressed and played a part in establishing a large but non-redundant set of conformational states. Due to such a process, the apoprotein and ligand-bound RdRp conformational landscapes were resolved considerably, allowing more accurate subsequent analysis, such as ensemble docking.

Molecular Docking

Ensemble docking employed Cresset Flare (v5.0.0) designed to employ the XED polarisable force field. A total of 225 MD-derived RdRp structures were prepared and superimposed for docking. Docking was performed using the Lead Finder algorithm with a grid box centred on the native ligand’s binding site. All 328 curcuminoid compounds were parameterised using the GAFF2 AMBER force field to evaluate their inhibitory potential against DENV RdRp (Brooks et al., 1983).

Several compounds, i.e., 90, 175, 1CJS, 3DDX, 113, and 187 demonstrated stronger binding affinities than GTP and interacted with key catalytic residues, highlighting them as promising inhibitors. These compounds satisfied two critical criteria: superior binding strength and direct engagement with functional sites. The Lead Finder Virtual Screening (LF VSscore) was used for initial ranking (Stroganov et al., 2009), where more negative scores indicated stronger predicted interactions. This scoring helped prioritise top candidates for deeper analysis, including MD simulations and MM/PBSA calculations.

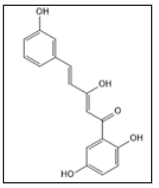
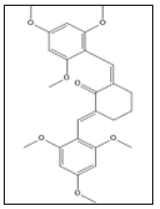
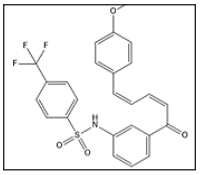
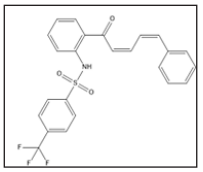
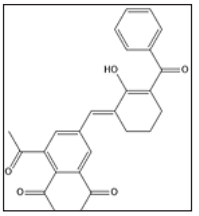
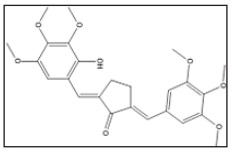
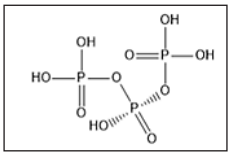
Table 1 summarises the binding affinities of compounds 90, 175, 1CJS, 3DDX, 113, and 187, with compound 90 exhibiting the strongest affinity compared to the others, including GTP. Table 2 details the molecular interactions of these high-affinity compounds with critical residues within the active site of RNA-dependent RNA polymerase (RdRp). Specifically, compounds 90, 175, 1CJS, 3DDX, 113, and 187 show consistent and specific binding to essential catalytic residues such as Arg729, Ser710, Thr794, Trp795, and Arg737, along with neighbouring amino acids that contribute to the structural and functional stability of the active site. The interaction patterns involve hydrogen bonding, electrostatic contacts, hydrophobic interactions, and π–π stacking, which act synergistically to stabilise the ligands and enhance their binding affinity.

Notably, Ser710 and Thr794 frequently form hydrogen bonds that help maintain the correct orientation of the ligand within the binding pocket. Arg729 and Arg737 engage in electrostatic interactions with polar functional groups on the ligands, strengthening their anchorage and contributing to overall binding energy. Trp795 often participates in π–π stacking interactions with aromatic moieties of the ligands, which is particularly important for stabilising planar and conjugated chemical scaffolds. Collectively, these interactions highlight the ability of the selected compounds to form stable and favourable contacts with RdRp, supporting their potential as effective inhibitors that may interfere with its enzymatic function.

In Figure 3, the best binding position of compound 1CJS in RdRp is represented with a docking score of -10.9 kcal/mol, depicting the interaction between RdRp and compound 1CJS at the major interacting residues in the binding pose. A pose view was utilised to gain insights into the residues responsible for the negative free energy. The key residues involved were Thr794, Arg729, Thr758, Ser710, Ser796, Ser791, Met761, Thr766, Thr793, Gln742, Leu734, Trp762, Met765, Ile740, and Cys709. Further analysis of the nonbonded interactions revealed that compound ICJS formed strong hydrogen bonds with the interacting residues Thr766, Arg729, Ser796, Cys709, and His711. Additionally, alkyl interactions with Arg737, Arg792, and Leu511 indicated the interaction of RdRp and compound 1CJS with the active site of the RdRp main protease.

In Figure 4, the compound ligand 175 demonstrated a scoring binding pose in RdRp with a score of -11.2 kcal/mol, highlighting the interaction between RdRp and the compound ligand 175 at the major interacting residues in the binding pose. A pose view was utilised to gain insights into the residues responsible for the negative free energy. The key residues involved were Thr794, Arg729, Thr758, Arg792, Ser710, Ser796, Ser741, Met761, Thr766, Thr793, Gln742, and Leu734. Further analysis of the nonbonded interactions revealed that compound 175 formed strong hydrogen bonds with the interacting residues Thr794 and Arg729. Moreover, alkyl and pi-alkyl interactions with Trp795, Ile740, Tyr758, His711, and Leu511 indicate the interaction of RdRp and compound 175 with the active site of the RdRp main protease.

Table 1
Docking results with the highest docking energies

Structure	ID	LF Vsscore (kcal/mol)
	90	-11.8
	175	-11.2
	ICJS	-10.9
	3DDX	–10.4
	113	-10.3
	187	-10.3
	GTP (control)	- 6.9

Note. The compound IDs represent different curcuminoid compounds. The LF Vsscore values indicate the predicted binding energies obtained from molecular docking simulations, where more negative values correspond to stronger predicted binding affinity. GTP was included as a control ligand for comparison

Table 2
Types of interactions between the top 3 molecular compounds with the highest binding affinities and the RdRp residues, with the original GTP ligand used as a positive control

ID	Hydrogen Bond Interaction	Pi-Pi T Shaped Interaction	Halogen Interaction	Pi-Alkyl Interaction	Pi-Sulphur /Anion	Carbon-Hydrogen Interaction	Pi-Pi Stacking
90	Thr793 Leu734 Cys709	Thr766		Met761 Cys709	Cys709	Arg737	Thr794
1CJS	Trp762 Thr794		Cys709 Ser710 His711	Arg737 Trp803 Leu511	Met761 Met765	Ser710 His711 Leu734 Ser796	
175	Tyr758 Arg792 Trp795 Arg729	His711 Trp795		His711 Trp705 Tyr758		Arg729 Cys709 Arg737	
GTP	Trp795 Arg729 Arg737					Thr794	

Note. The table summarises the types of non-covalent interactions between the top three curcuminoid compounds with the highest predicted binding affinities and dengue RdRp residues, with GTP included as a positive control. Interactions include conventional hydrogen bonds, Pi-Pi T-shaped, halogen, Pi-alkyl, Pi–sulfur/anion, carbon-hydrogen, and Pi–Pi stacking interactions. These interactions indicate how each compound stabilises within the active site and contributes to the overall binding affinity

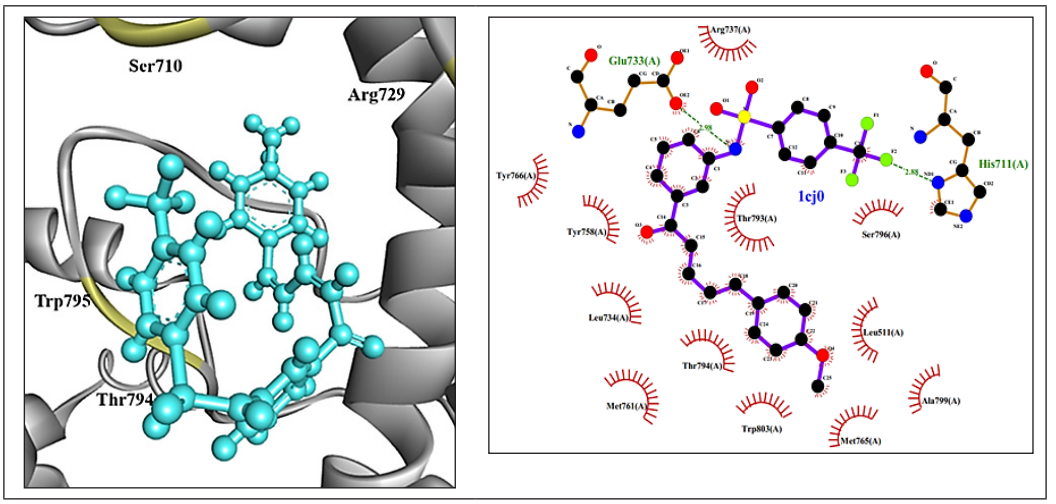


Figure 3. Residues at the active site of RdRp that interact with compound1CJS. Conventional H-bonds (green dashed line) are shown

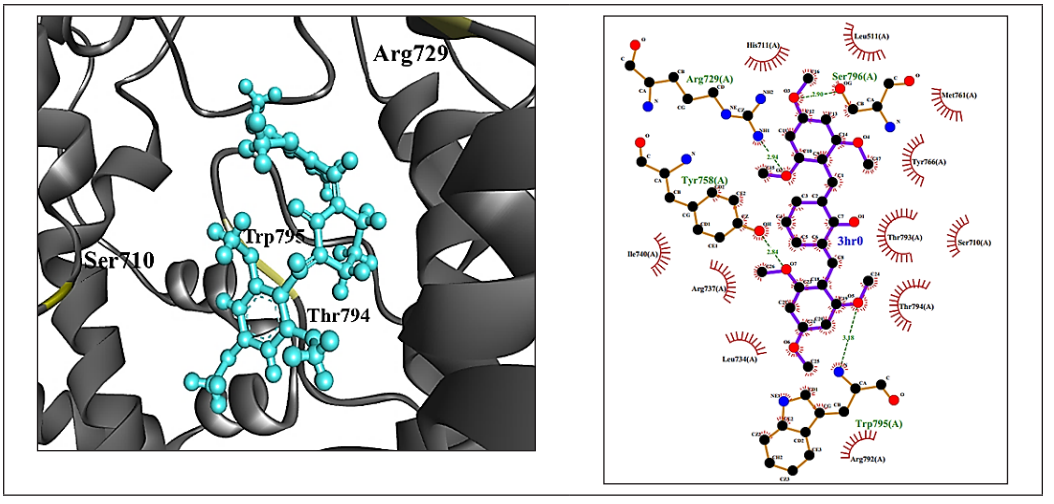


Figure 4. Residues at the active site of RdRp that interact with compound175. Conventional H-bonds (green dashed line) are shown

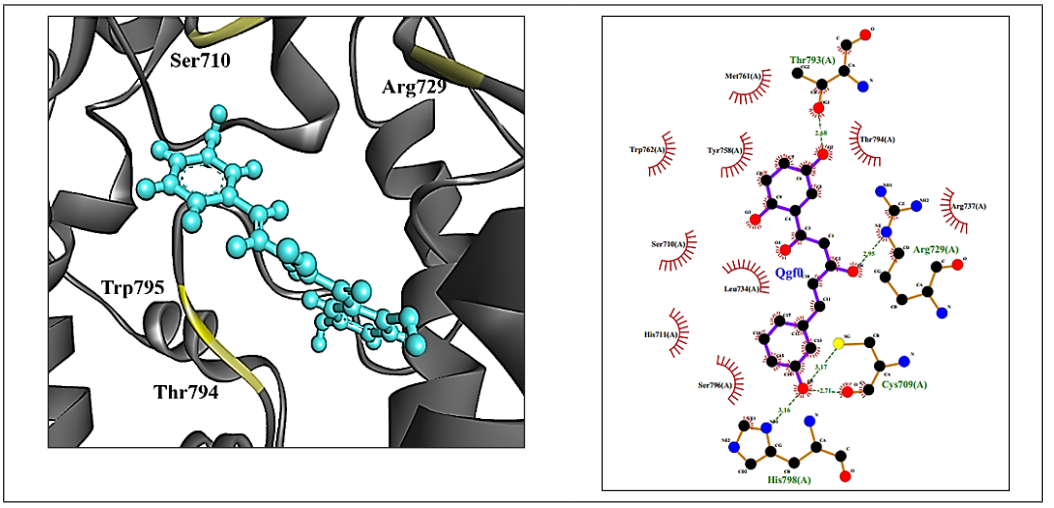


Figure 5. Residues at the active site of RdRp that interact with compound 90. Conventional H-bonds (green dashed line) are shown

As illustrated in Figure 5, compound 90 demonstrated the most favourable binding pose within the RdRp active site, achieving a docking score of -11.8 kcal/mol, indicative of strong binding affinity. The interaction was stabilised by a network of key amino acid residues, including Thr794, Arg729, Thr758, Ser710, Met761, Thr766, Thr793, Tyr606, Gly662, Ser661, Cys706, Leu734, Glu733, His798, and Leu511, which collectively formed a well-defined binding environment. A closer examination of non-bonded interactions revealed that compound 90 established multiple hydrogen bonds with Thr794, Thr793, Cys709,

Ser661, and Thr766, enhancing the specificity and strength of binding. Additionally, pi-alkyl interactions with Arg737 further contributed to the stability of the ligand-protein complex. These findings strongly suggest that compound 90 effectively occupies and interacts with the RdRp catalytic site, potentially interfering with its enzymatic function.

ADMET Analysis

ADMET, which stands for Absorption, Distribution, Metabolism, Excretion, and Toxicity, is an essential part of drug development. It helps us understand how a compound behaves in the human body; how well it is absorbed, where it goes, how it is broken down, how it is eliminated, and whether it might be toxic. Studying these properties early on makes it easier to identify promising drug candidates and avoid problems later in development.

Table 3 shows notable differences among the five compounds in terms of drug-likeness, solubility, and predicted oral absorption. Compounds 90 and 1CJS show the most favourable overall profiles, combining suitable molecular weight, good solubility, and high gastrointestinal absorption. Compound 1CJS emerges as the strongest candidate, as it has no PAINS or Brenk alerts and exhibits no major metabolic or permeability issues. Compound 90 also performs well, though the presence of several structural alerts suggests careful consideration. Compound 175 remains acceptable, with moderate drug-likeness, but its CYP2D6 inhibition slightly reduces its appeal. In contrast, compounds 3DDX and 133 demonstrate poor solubility, low absorption, and excessive lipophilicity, making them less suitable as lead compounds. Among them, Compound 133 is the weakest, showing additional P-gp substrate behaviour and a Lipinski rule violation. Overall, Compounds 90 and 1CJS appear to be the most promising ADME candidates for further optimisation and experimental evaluation.

MD Simulations

Molecular dynamics (MD) simulations were performed after the docking analysis to take solvent effects into consideration and assess the structural stability of the resultant ligand-protein complexes. The simulations reported that conformation relaxation of the ligand associated with an induced-fit process occurs after binding, and the RdRp receptor was also rearranged. The explicit inclusion of the water molecules in the modelled environment allowed a deeper comprehension of the solvent-mediated interactions and destabilisation and stabilisation of complexes. To a large extent, we were able to assess the stability and dynamic behaviour of the ligand-receptor complexes under near physiological conditions more realistically due to our solvent-aware methodology.

More so, the simulation provided valuable data regarding the system's evolution by providing a dynamic picture of the complex. In this dynamic view, a complete vision of the behaviour and interaction of the system is obtained since it becomes easier to differentiate docked and unstable conformations.

Table 3
ADME analysis of the selected compound with the highest binding affinity

Properties	175	3DDX	133	90	1CJS
Physicochemical					
Molecular Weight (g/mol)	451.61	487.49	457.46	298.29	416.47
Log P (Consensus)	4.35	5.32	5.42	2.34	3.99
TPSA (Å²)	72.45	80.85	71.62	97.99	88.51
Water Solubility (Log S ESOL)	-5.48 (Mod. (Soluble)	-6.26 (Poorly Soluble)	-6.53 (Poorly Soluble)	-3.88 (Soluble)	-4.73 (Mod. Soluble)
Pharmacokinetics					
GI Absorption	High	Low	Low	High	High
BBB Permeant	No	No	No	No	No
P-gp Substrate	No	No	Yes	No	No
CYP2D6 Inhibitor	Yes	No	No	No	No
CYP3A4 Inhibitor	No	Yes	No	No	Yes
Drug Likeness					
Lipinski Violations	0	0	1 (Log P)	0	0
Bioavailability Score	0.55	0.55	0.55	0.56	0.56
Alerts					
PAINS/Brenk Alerts	2 Alerts	2 Alerts	2 Alerts	4 Alerts	0 Alerts

Note. Physicochemical properties include molecular weight, Log P, TPSA, and water solubility. Pharmacokinetic properties cover gastrointestinal absorption, blood-brain barrier (BBB) permeability, P-gp substrate status, and CYP450 enzyme inhibition. Druglikeness is assessed based on Lipinski's rules and bioavailability score, while PAINS/Brenk alerts indicate potential structural liabilities. Compounds with optimal ADME profiles and minimal alerts are considered more favourable for drug development

Root Mean Square Deviation

To determine the stability of the structure and dynamics of RdRp complexes with the three best curcuminoid compounds, 150-ns molecular dynamics (MD) simulations were carried out (Figure 6). The root mean square deviation (RMSD) analysis was done to normalise the deviations of the changed structure with the initial structure allowed an impression of how each ligand impacts the consequences of the conformational moves in the RdRp. The simulations showed that the phenomenon of ligand binding led to individual dynamic reactions, showing the differences in the interaction of various ligands with the transformation of the structure of the RdRp.

There was a close match between the RMSD trajectories of compound 90 and that of the native ligand GTP during 70 to 100 ns, suggesting similar orientation in binding to the active site. Both systems showed similar structural dynamics, indicates a possible role of compound 90 as an analogue of GTP. In contrast, compounds 175 and 1CJS maintained consistently low RMSD values (below 3 Å) throughout the simulation, reflecting minimal

conformational fluctuations and high complex stability. Such rigidity may indicate that these compounds lock RdRp into a less flexible, possibly inactive conformation.

The unbound (apoprotein) RdRp also displayed relatively low RMSD values, confirming its inherent structural stability. However, both GTP and compound 90 induced slightly greater rigidity compared to the apoprotein form, which may correspond to the stabilisation of catalytically unfavourable states. Collectively, these findings suggest that curcuminoid-based inhibitors, particularly compound 90, can effectively modulate RdRp dynamics either by mimicking natural substrates or by imposing structural constraints. This reinforces their potential as promising antiviral candidates targeting DENV RdRp.

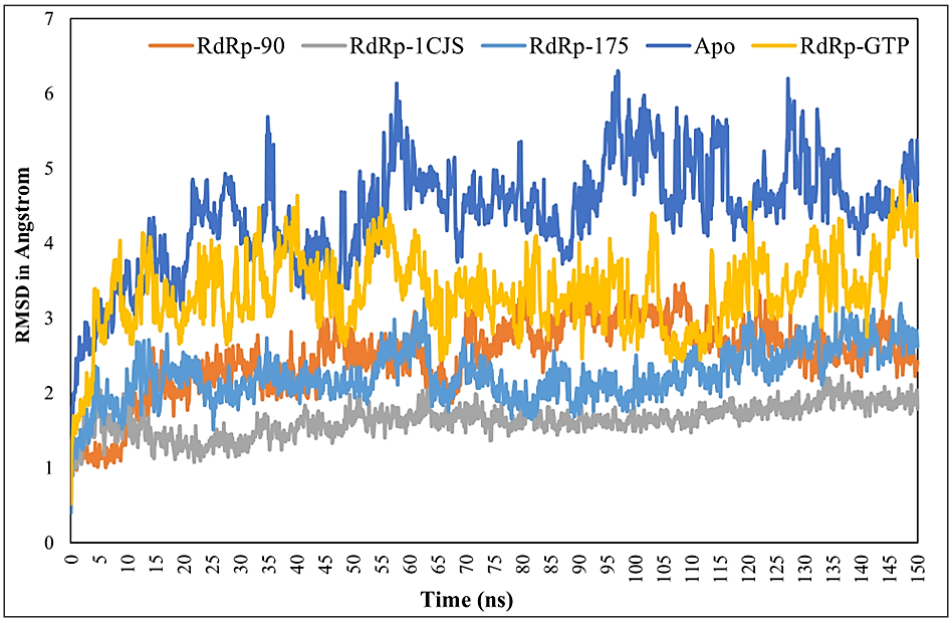


Figure 6. RMSD of complexes of RdRp over the 150 ns time scale

Root Mean Square Fluctuation

The root mean square fluctuation (RMSF) analysis offers critical insights into how ligand binding modulates the flexibility of DENV RdRp (Figure 7). In the unbound (apoprotein) form, the loop spanning residues 341-347, particularly residue 345, exhibited the highest flexibility (RMSF > 5 Å), indicating structural instability near the template tunnel. GTP binding markedly reduced fluctuations in this region, stabilising the loop. Notably, compound 90 elicited a comparable dampening effect, suggesting it mimics GTP's stabilising influence. In contrast, 1CJS and compound 175 produced only modest reductions in flexibility, reflecting weaker or less specific interactions with this region.

At the catalytic site, GTP and compound 90 both significantly reduced the flexibility of residues Ser710, Thr794, and Trp795, indicating strong and stabilising interactions essential for polymerase function. Conversely, compound 175 increased fluctuations at Trp795, suggesting possible destabilisation at a key catalytic position. Additionally, the segment spanning residues 472-486 displayed increased flexibility across all ligand-bound systems, possibly due to ligand-induced allosteric effects. The region Phe348-Arg352 exhibited over 50% reduced fluctuation in all complexes, indicating enhanced conformational stability and potentially improved catalytic precision. Meanwhile, the Thr585-Ser593 segment remained consistently stable, underscoring its likely role as part of the structural core. Collectively, these findings highlight compound 90's ability to replicate GTP's stabilising effect on RdRp dynamics, supporting its potential as a viable inhibitory candidate.

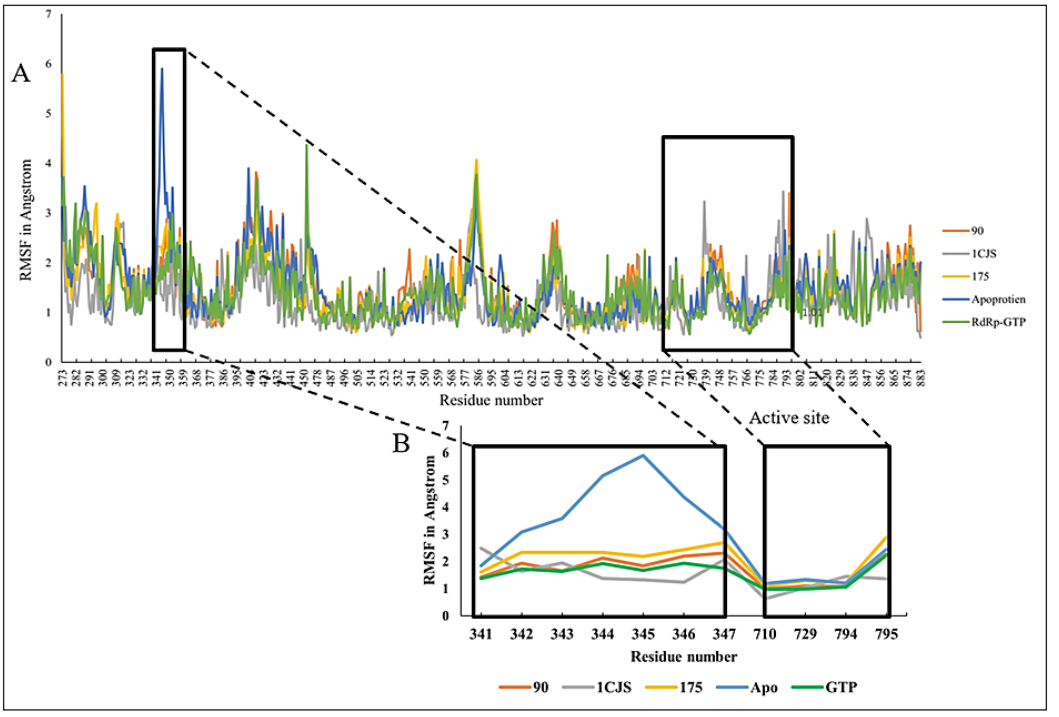


Figure 7. (A) Root mean square fluctuation (RMSF) profile of the RdRp-ligand complexes; (B) Highlighted regions of interest in the RMSF plot

Radius of Gyration

The radius of gyration (Rg)is a significant parameter in the analysis of protein compactness and structural stability in molecular dynamics analyses. Low Rg values characterise more compact and folded protein structures, and high values represent high flexibility or

non-compactness with a partial unfolding. Rg analysis was done in this study to study the relationship between the curcuminoid binding and the structural stability of DENV RdRp.

The ligand-bound RdRp complexes were consistently found to have a lower Rg than the unbound (apoprotein) form, consistent with the idea that binding ligands creates a more stable and compact protein conformation (Figure 8). Two of the tested compounds, GTP and compound 90, possessed almost the same Rg profiles (25.3-26.5 Å), meaning that Compound 90 almost perfectly mimics the RdRp folded architecture of GTP. Compounds 1CJS and 175 also contributed to structural stabilisation, albeit to a lesser extent. In contrast, the apoprotein RdRp showed greater Rg fluctuations, reflecting reduced conformational rigidity. These findings reinforce compound 90's potential to stabilise the polymerase structure similarly to the native ligand, supporting its candidacy as a potent RdRp inhibitor.

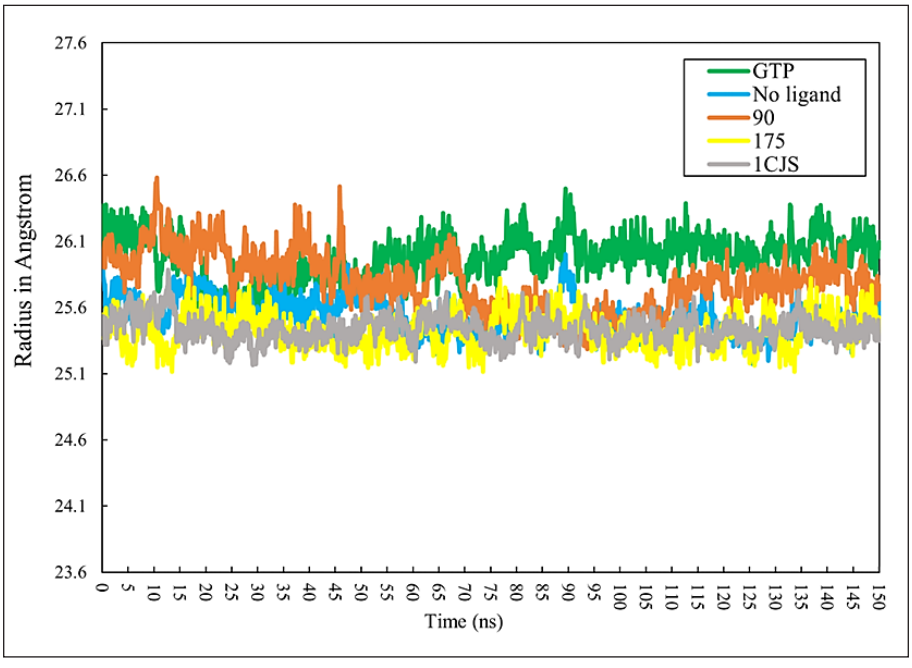


Figure 8. Analysis of the radius of gyration of the RdRp complexes over a 150 ns duration

Dynamic Cross-correlation Matrix

The dynamic cross-correlation matrix (DCCM) analysis of DENV RdRp illustrates how ligand binding modulates internal residue-residue motions during molecular dynamics simulations (Figure 9). In these plots, yellow regions indicate positively correlated, coordinated motions, while blue regions reflect anti-correlated, opposing movements. The red circle highlights the catalytic active-site region, where ligand binding exerts the

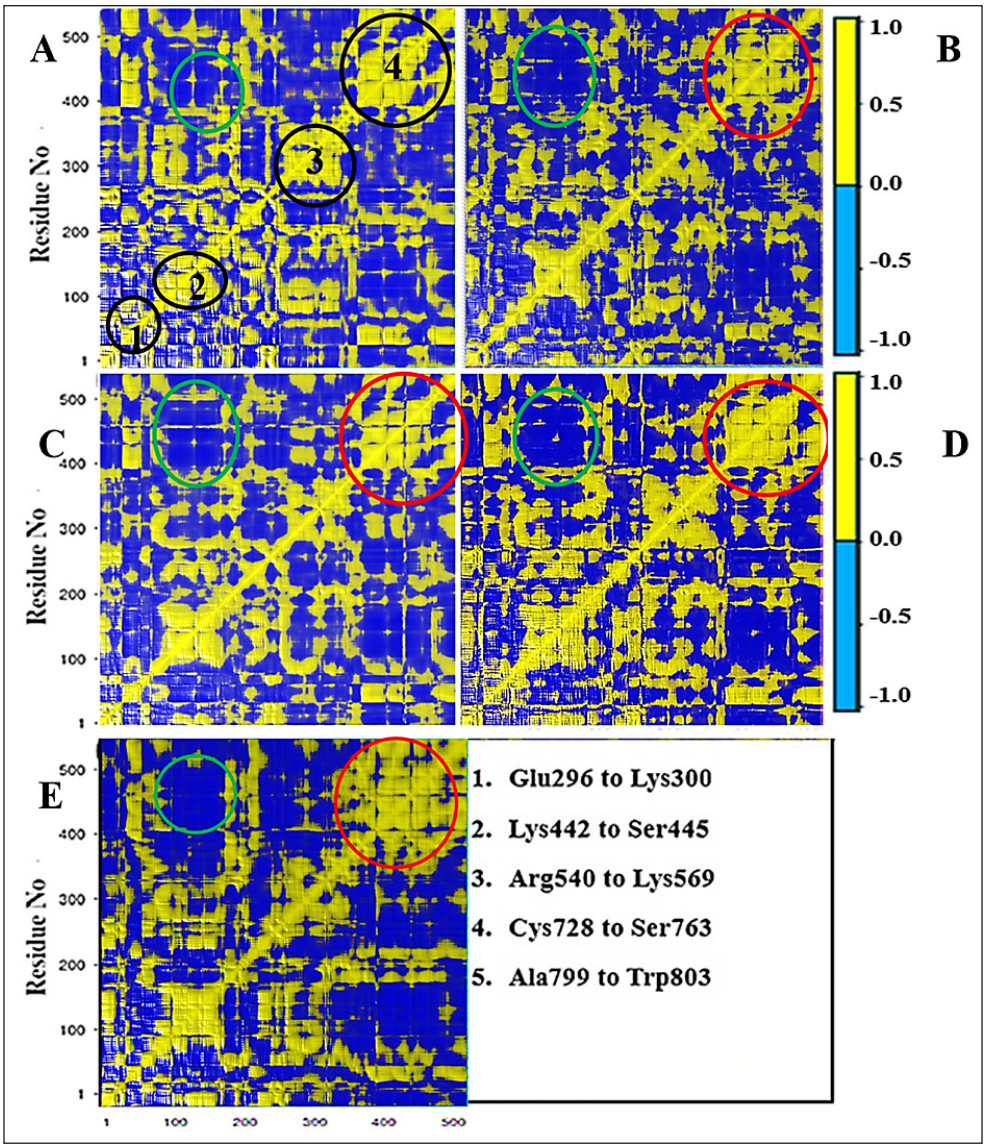


Figure 9. Dynamic cross-correlation map (DCCM) comparing the unbound RdRp structure with its various ligand-bound complexes

greatest influence, and the green circle marks an adjacent residue cluster that supports catalytic stability and intra-protein communication.

In the GTP-bound RdRp (Panel A), four key regions Lys300-Glu296, Lys442-Ser445, Arg540-Lys569, and Cys728-Ser763 exhibited strong positive correlations around both circles, reflecting well-organised dynamics that support efficient RNA synthesis. Compound 90 (Panel C) closely mimicked this pattern, restoring coordinated motions in the

same regions, suggesting that it not only occupies a similar binding pose as GTP but also replicates its effect on RdRp dynamics, supporting its potential as a competitive inhibitor.

Conversely, compound 175 (Panel B) disrupted these correlations, particularly within the catalytic region, weakening residue coordination, whereas 1CJS (Panel D) provided intermediate stabilisation. Notably, both 175 and 1CJS introduced shifts in anti-correlated interactions, especially in regions Cys728-Ser763 and Ala799-Trp803, toward positive correlations, indicating possible allosteric modulation that may stabilise RdRp in a more rigid, less catalytically active conformation. The apoprotein form (Panel E) displayed fragmented and weak correlations, emphasising the critical role of ligand binding in maintaining the structural and functional integrity of the polymerase.

Principal Component Analysis

The principal component analysis (PCA) results for compounds 175, 90, 1CJS, and the native ligand (GTP) are shown in Figure 10. In each subplot (A, B, C, and D), red dots represent positive conformations, i.e. trajectory frames that project in the positive direction along the principal component axis. These conformations correspond to one extreme of the dominant motion captured by PCA and are typically associated with an activated or open functional state of the protein. Conversely, blue dots represent negative conformations, projecting in the negative direction along the same axis, often reflecting a more compact, closed, or inactive structural state.

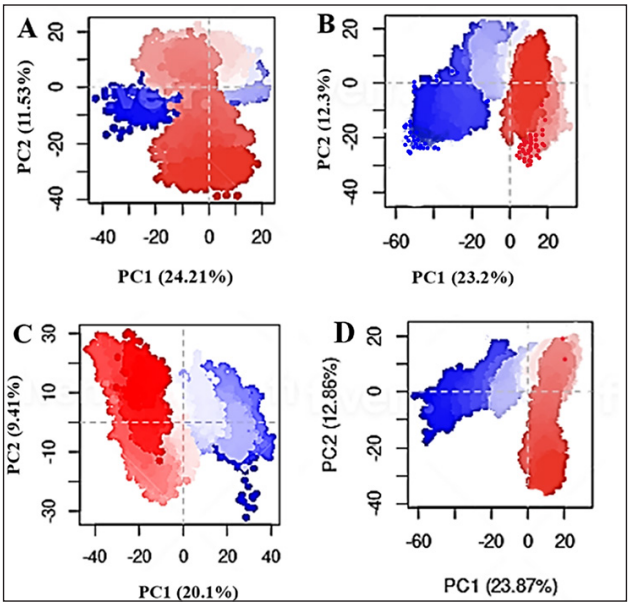


Figure 10. PCA analysis of the curcuminoids compound, where A represents compound 175; B represents compound 90; C represents compound 1CJS; and D represents the native ligand

Principal component analysis (PCA) revealed varying degrees of structural flexibility among the compounds. Compound 175 (Plot A) showed moderate conformational diversity with PC1 and PC2 explaining 24.21% and 11.53% of the variance, respectively. Compound 90 (Plot B) exhibited more pronounced clustering and separation, with PC1 and PC2 accounting for 23.2% and 12.3% respectively, indicating greater structural variation. Compound 1CJS (Plot C) explored a narrower conformational space, with lower variance explained (20.1% and 9.41%), suggesting restricted dynamics.

The native ligand GTP (Plot D) displayed distinct structural states, with PC1 and PC2 contributing 23.87% and 12.86%, respectively, resembling the variation seen with compound 90. Among all, compound 90 and GTP showed the highest structural diversity, while 1CJS remained the most conformationally constrained. Compound 175 showed intermediate behaviour. These findings highlight compound 90's dynamic flexibility, closely mimicking the native ligand's conformational behaviour, which may enhance its inhibitory potential.

Binding Free Energy Analysis

We employed the molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) method for binding free energy calculations using YASARA Structure and AMBER14 force field. MM/PBSA accurately estimates the binding free energies and obtains valuable information concerning interactions between the ligands and receptors. The binding free energy (DGbind) in kcal/mol was calculated by using the following formula:

$$\Delta Gbind = \Delta Gcomplex\ (minimised) - (\Delta Gligand\ (minimised) + \Delta Greceptor\ (minimised)$$

In the case that ΔGreceptor complex (minimised) is the minimised energy of the complex, ΔGreceptor ligand (minimised) is the minimised energy of the isolated ligand, and ΔGreceptor receptor (minimised) is the minimised energy of the isolated receptor. The three main contributions that we considered are: ΔG_{MM} (molecular mechanics interaction), ΔG_{PB} (polar solvation energy), and ΔG_{SA} (nonpolar solvation energy), to obtain DGbind. The term ΔG_{MM} considers the amount of electrostatic + van der Waals. The ΔG_{PB} and the ΔG_{SA} parameters correspond to the solvation energies' nonpolar and polar expressions.

With an average binding energy of -127.6 kcal/mol, Compound 90 exhibited the strongest binding affinity among all tested ligands. In contrast, compound 1CJS and Compound 175 showed substantially weaker affinities, at -33.2 and -31.3 kcal/mol, respectively, while the natural ligand GTP displayed a binding affinity of -89.9 kcal/mol (Table 4). These findings indicate that Compound 90 interacts with the active site far more strongly than both the native ligand and the other evaluated compounds.

Compound 90's much greater binding indicates that it may be a powerful RdRp inhibitor and competitive GTP mimic. MD modelling results confirm that it can efficiently occupy and stabilise the catalytic site. All these results point to Compound 90 as a promising antiviral candidate for more dengue virus research.

Table 4
MM/PBSA free energy of binding (kcal/mol)

Ligand	MM/PBSA kcal/mol
90	-127.6
1CJS	-33.2
175	-31.3
GTP (control)	-89.9

Binding Mechanism of Compound 90

Hydrogen bonds and hydrophobic interactions are crucial for ligand-protein complex stabilisation and binding affinity determination in computational drug development. In order to fix ligands at the binding site, hydrogen bonds, which are created between a hydrogen atom and electronegative atoms like oxygen or nitrogen, are essential. The stability of the complex is greatly increased by these interactions, which also improve binding specificity.

Tight binding within hydrophobic pockets is promoted by hydrophobic interactions, which happen when non-polar parts of the ligand and protein group together to avoid water. Supported by van der Waals forces, these interactions enhance total affinity and help keep the ligand oriented correctly. When combined, these pressures are essential for assessing and improving ligand binding during the drug discovery process.

The binding behaviour of compound 90 within the active site of DENV RdRp over a 150-ns molecular dynamics simulation is shown in Figure 11. The ligand demonstrated stable and specific binding, characterised by a network of hydrophobic contacts and hydrogen bonds that contributed to its strong affinity and conformational stability. Among the interacting residues, Arg729 emerged as the most critical, maintaining both hydrogen bonding and hydrophobic interactions for over 70% of the simulation time. This persistent engagement anchored the ligand effectively within the binding pocket, enhancing specificity and minimising positional drift.

Trp795 also played a significant role, forming consistent hydrophobic interactions (~71%) with the ligand's nonpolar region, likely through favourable van der Waals forces that further stabilised the complex. Additionally, residues Asp663 and Asp664 provided moderate support via transient hydrogen bonds and hydrophobic contacts, likely aiding in optimal ligand orientation. Interactions with Ser710 and Thr794 were less frequent, reflecting the inherent flexibility of nearby loop regions and the dynamic nature of the binding site.

Despite the variability in individual residue contributions, Compound 90 established a residue-specific interaction pattern, primarily anchored by Arg729 and Trp795, that maintained a stable and favourable pose throughout the trajectory.

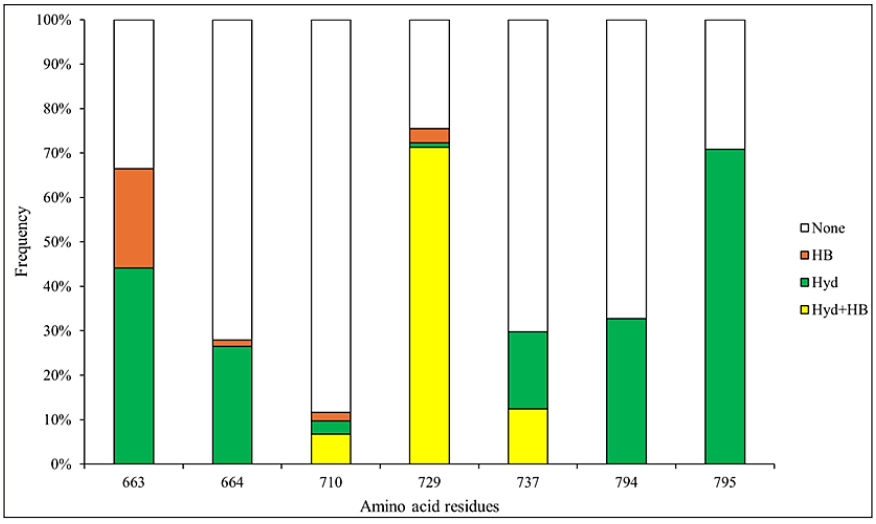


Figure 11. Frequency distribution of interactions between compound 90 and selected amino acid residues of RdRp during the 150 ns MD simulation

The MM/PBSA-calculated binding free energy of -127.6 kcal/mol further reinforces its strong binding affinity, exceeding that of the native ligand GTP and other tested compounds. These results underscore Compound 90's potential as a potent RdRp inhibitor and a promising candidate for antiviral drug development.

The 150-ns molecular dynamics simulation of the native ligand, revealing a stable binding profile predominantly maintained through hydrophobic interactions (Figure 12). Key residues such as Arg737, Thr794, and Trp795 consistently formed strong nonpolar contacts, indicating that the ligand was securely anchored within a hydrophobic pocket. These interactions, primarily mediated by van der Waals forces, contributed to a fixed ligand conformation and limited positional fluctuation features essential for maintaining binding stability.

The native ligand showed less hydrogen bonding than compound 90, which led to a less varied network of interactions. Interestingly, Asp663 and Asp664 showed minimal to no involvement, most likely as a result of peripheral placement in relation to the ligand or suboptimal spatial orientation. More prominent hydrophobic contacts maintained the overall stability of the binding even in their absence. Occasionally, Ser710 created hydrogen bonds, which provided slight polar stability without having a major impact on the orientation of the ligand.

Although less interactive than compound 90, the native ligand maintained steady contact with critical residues, affirming its role as a stable, albeit less optimised, binder. These findings highlight the native ligand's sufficient binding stability while emphasising the superior interaction profile of compound 90.

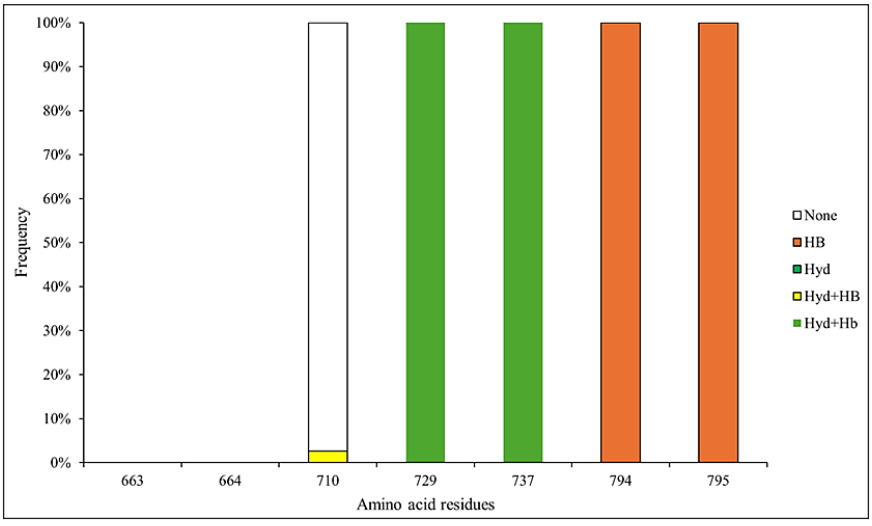


Figure 12. Frequency distribution of interactions between native ligand and selected amino acid residues of RdRp during the 150 ns MD simulation.

Curcumin has long been recognised for its broad antiviral properties, including activity against dengue virus. Previous studies showed that, at low concentrations (10-30 μM), curcumin reduces viral replication not by directly targeting viral enzymes but by affecting host-cell processes such as the actin cytoskeleton and the ubiquitin-proteasome system, leading to accumulation of viral proteins and reduced plaque formation (Padilla-S et al., 2014). Structural studies also revealed that curcumin can allosterically inhibit the NS2B–NS3 protease by binding to a non-active site cavity and destabilising its active conformation (Lim et al., 2020). However, these effects are relatively moderate and indirect. In our study, compound 90 showed a markedly different profile. It directly binds to the dengue RNA-dependent RNA polymerase RdRp, forming stable interactions with key catalytic residues, and molecular dynamics analyses demonstrated that it closely mimics the native ligand GTP in both structure and dynamic behaviour. Compound 90 combines strong specificity with potent inhibitory potential, directly interfering with viral replication. These findings suggest that rationally designed curcuminoids, such as compound 90, can overcome the limitations of native curcumin and represent promising leads for the development of more effective dengue antivirals.

CONCLUSION

In this study, we used computational approaches to identify potential inhibitors of dengue virus RdRp. Molecular dynamics simulations of 150 ns were performed to evaluate the stability and interactions of both apoprotein and ligand-bound RdRp. Ensemble docking of 328 curcuminoid compounds yielded three promising hits, with compound 90 emerging

as the strongest candidate, showing the highest binding affinity even surpassing the native ligand GTP. Post-simulation analyses, including RMSD, RMSF, radius of gyration, PCA, DCCM, and MM/PBSA, confirmed its stable binding and revealed that it closely mimics the dynamic behaviour of GTP within the active site, suggesting it could act as a competitive or pseudo-substrate inhibitor. Compound 90 formed persistent hydrogen bonds with key residues, including Ser710, Arg737, Thr794, Trp795, and the catalytic Asp663 and Asp664, which remained stable throughout the simulation. These results highlight Compound 90 as a highly promising RdRp inhibitor and support its further experimental validation as a potential antiviral agent against dengue virus.

ACKNOWLEDGEMENT

The authors would like to thank Priya Murugan and Yuan Chongjun for technical assistance. This research received no external funding.

REFERENCES

Adkins, J. C., & Faulds, D. (1998). Amprenavir. *Drugs, 55*(6), 837-842. https://doi.org/10.2165/00003495-199855060-00015

Balasubramanian, A., Pilankatta, R., Teramoto, T., Sajith, A. M., Nwulia, E., Kulkarni, A., & Padmanabhan, R. (2019). Inhibition of dengue virus by curcuminoids. *Antiviral Research, 162*, 71-78. https://doi.org/10.1016/j.antiviral.2018.12.002

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry, 4*(2), 187-217. https://doi.org/10.1002/jcc.540040211

Douglas, W. W., & Rubin, R. P. (1961). The role of calcium in the secretory response of the adrenal medulla to acetylcholine. *The Journal of Physiology, 159*(1), 40-57. https://doi.org/10.1113/jphysiol.1961.sp006791

Hassan, H. A., Hassan, A. R., Mohamed, E. A. R., Al-Khdhairawi, A., Karkashan, A., Attar, R., Allemailem, K. S., Abdulmonem, W. Al, Shimizu, K., Abdel-Rahman, I. A. M., & Allam, A. E. (2022). Conducting the RBD of SARS-CoV-2 Omicron variant with phytoconstituents from *Euphorbia dendroides* to repudiate the binding of spike glycoprotein using computational molecular search and simulation approach. *Molecules, 27*(9), Article 2929. https://doi.org/10.3390/molecules27092929

Khetarpal, N., & Khanna, I. (2016). Dengue fever: Causes, complications, and vaccine strategies. *Journal of Immunology Research, 2016*, Article 6803098. https://doi.org/10.1155/2016/6803098

Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry, 36*(13), 996-1007. https://doi.org/10.1002/jcc.23899

Lim, L., Dang, M., Roy, A., Kang, J., & Song, J. (2020). Curcumin allosterically inhibits the dengue NS2B-NS3 protease by disrupting its active conformation. *ACS Omega, 5*(40), 25677-25686. https://doi.org/10.1021/acsomega.0c00039

Lim, S. P., Noble, C. G., Nilar, S., Shi, P.-Y., & Yokokawa, F. (2018). Discovery of potent non-nucleoside inhibitors of dengue viral RNA-dependent RNA polymerase from fragment screening and structure-guided design. In *Methods in molecular biology* (pp. 187-198). https://doi.org/10.1007/978-981-10-8727-1_14

Mahmood, K., Zia, K. M., Zuber, M., Salman, M., & Anjum, M. N. (2015). Recent developments in curcumin and curcumin-based polymeric materials for biomedical applications: A review. *International Journal of Biological Macromolecules, 81*, 877-890. https://doi.org/10.1016/j.ijbiomac.2015.09.026

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation, 11*(8), 3696-3713. https://doi.org/10.1021/acs.jctc.5b00255

Markham, A. (2017). Baricitinib: First global approval. *Drugs, 77*(6), 697-704. https://doi.org/10.1007/s40265-017-0723-3

Martínez-Rosell, G., Giorgino, T., & De Fabritiis, G. (2017). PlayMolecule ProteinPrepare: A web application for protein preparation for molecular dynamics simulations. *Journal of Chemical Information and Modeling, 57*(7), 1511-1516. https://doi.org/10.1021/acs.jcim.7b00190

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics, 3*(1), Article 33. https://doi.org/10.1186/1758-2946-3-33

Padilla-S, L., Rodríguez, A., Gonzales, M. M., Gallego-G, J. C., & Castaño-O, J. C. (2014). Inhibitory effects of curcumin on dengue virus type 2-infected cells in vitro. *Archives of Virology, 159*(3), 573-579. https://doi.org/10.1007/s00705-013-1849-6

Prlina, I. M., Dahlen, C. P., & Bischoff, A. J. (1958). Sulfamethoxypyridazine, a new sulfonamide for urinary infections: A preliminary report. *Antibiotic Medicine & Clinical Therapy (New York, NY), 5*(4), 242-250.

Qureshi, A. I. (2020). Dengue virus infection. In A. I. Qureshi (Ed.), *Dengue virus disease* (1st ed., pp. 1-7). Elsevier. https://doi.org/10.1016/B978-0-12-818270-3.00001-1

Singh, P. K., Pathania, S., & Rawal, R. K. (2020). Exploring RdRp-remdesivir interactions to screen RdRp inhibitors for the management of novel coronavirus 2019-nCoV. *SAR and QSAR in Environmental Research, 31*(11), 857-867. https://doi.org/10.1080/1062936X.2020.1825014

Siviero, A., Gallo, E., Maggini, V., Gori, L., Mugelli, A., Firenzuoli, F., & Vannacci, A. (2015). Curcumin, a golden spice with a low bioavailability. *Journal of Herbal Medicine, 5*(2), 57-70. https://doi.org/10.1016/j.hermed.2015.03.001

Stroganov, O. V., Novikov, F. N., Stroylov, V. S., Kulkov, V., & Chilov, G. G. (2009). ChemInform abstract: Lead Finder: An approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. *ChemInform, 40*(16), Article 119992. https://doi.org/10.1002/chin.200916212

Weaver, S. C., & Vasilakis, N. (2009). Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infection, Genetics and Evolution, 9*(4), 523-540. https://doi.org/10.1016/j.meegid.2009.02.003

Yap, T. L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., Vasudevan, S. G., & Lescar, J. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *Journal of Virology, 81*(9), 4753-4765. https://doi.org/10.1128/JVI.02283-06

Pertanika J. Sci. & Technol. 34 (2): 787 - 822 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Optimisation of Silver Nanoparticle Synthesis Using Aqueous Extract of *Passiflora foetida* L. via Response Surface Methodology

Silvia Handayani[1], Alyza A. Azmi[1], R. Rudiyanto[2], Wan Iryani Wan Ismail[1], Kustiariyah Tarman[3,6], Faridah Abas[4], Wan M. Khairul[1,5], and M. Maulidiani[1,5]*

[1]*Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

[2]*Faculty of Food Science and Agrotechnology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

[3]*Department of Aquatic Products Technology, Faculty of Fisheries and Marine Science, IPB University, 16680 Bogor, West Java, Indonesia*

[4]*Department of Food Science, Faculty of Food Science and Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[5]*Advanced Nano Materials (ANOMA) Research Group, Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

[6]*Centre for Coastal and Marine Resources Studies, International Research Institute for Maritime,Ocean and Fisheries (i-MAR), IPB University, 16680 Bogor, West Java, Indonesia*

ABSTRACT

This study investigates the green synthesis of silver nanoparticles (PF-AgNPs) using the aqueous extract of the medicinal plant *Passiflora foetida* to enhance their biological activity. PF-AgNPs were characterised using UV-Vis spectroscopy, FTIR, XRD, SEM, and EDX analyses. The UV-Vis absorption peaks of PF-AgNPs were observed between 380 and 480 nm. Structural characterisation using FTIR and XRD confirmed that the functional groups of *P. foetida* metabolites acted as surface-capping agents, yielding crystallised AgNPs with a face-centred cubic (fcc) structure. SEM-EDX analysis further revealed that the synthesised PF-AgNPs were spherical and displayed a uniform distribution of silver at 66.39% under the optimised condition. The synthesis of PF-AgNPs was optimised using a Box–Behnken design via Response Surface Methodology (RSM), evaluating the effects of extract concentration (0.1%, 0.2%, and 0.3%), pH (4, 7, and 10), and temperature (30 °C, 50 °C, and 70 °C), with observations

ARTICLE INFO

Article history:
Received: 28 July 2025
Accepted: 26 February 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.09

E-mail addresses:
silvia.handayanni@gmail.com (Silvia Handayani)
alyza.azzura@umt.edu.my (Alyza A. Azmi)
rudiyanto@umt.edu.my (R. Rudiyanto)
waniryani@umt.edu.my (Wan Iryani Wan Ismail)
kustiaz@apps.ipb.ac.id (Kustiariyah Tarman)
faridah_abas@upm.edu.my (Faridah Abas)
wmkhairul@umt.edu.my (Wan M. Khairul)
maulidani@umt.edu.my (M. Maulidiani)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

taken at 1 and 24 hours. UV-Vis intensity, NO scavenging activity, and DPPH scavenging activity were used as the response variables in RSM modelling. All models yielded p-values of less than 0.05, indicating that the factors significantly influenced PF-AgNP synthesis, with pH identified as the most effective parameter. PF-AgNPs exhibited DPPH and NO inhibitory activities of 35.64% and 38.7%, respectively, both of which were higher than those of the extract (17.16%). These findings highlight a sustainable and statistically guided approach for synthesising bio-functional AgNPs with promising potential for future biomedical applications.

Keywords: Antioxidant, green synthesis, *Passiflora foetida*, response surface methodology, silver nanoparticles

INTRODUCTION

Nanotechnology has been one of the fastest-growing fields and has rapidly become a powerful tool for tackling challenges in various areas, from medicine and pharmaceuticals to the food industry (De Silva et al., 2021; Naganthran et al., 2022). Among metallic nanoparticles, silver nanoparticles (AgNPs) have attracted particular attention due to their unique properties, including high conductivity and chemical stability. These features have made them useful in a wide range of applications, such as anti-inflammatory, antibacterial, and antioxidant activities (Alex et al., 2025; Barkat et al., 2017; Gherasim et al., 2020; Kaushal et al., 2023; Sytu & Camacho, 2018; Temur et al., 2024).

Green synthesis of AgNPs has gained popularity over traditional physical and chemical methods, largely because they are both eco-friendly and cost-effective (Alex et al., 2025; Pourmortazavi et al., 2015). Various biological resources have been used for AgNP production, including seaweed *Cladophora fascicularis* (Krishnamoorthi et al., 2022) and edible mushroom *Agaricus bisporus* (Rajasekar et al., 2019). The characteristics of the resulting nanoparticles can be influenced by several factors, including the type and concentration of biological reducers, precursor salt levels, reaction time, pH, and temperature (Crisan et al., 2021). Plant extracts are highly useful among various green methods due to their simple, fast, and economical synthesis with high reaction rates. Their phytochemicals, such as tannins, flavonoids, glycosides, alkaloids, polyphenols, enzymes and proteins act as reducing, stabilising and capping agents. They can also play the role of natural surfactants to promote NPs formation and stability (Abalkhil et al., 2017; Wang et al., 2021; Wasilewska et al., 2023).

Passiflora foetida, commonly known as passionflower, is a climbing plant of the genus *Passiflora* in the *Passifloraceae* family. The species has been widely distributed across tropical and subtropical regions due to its high adaptability and fast growth as a perennial tree. *P. foetida* has been reported to be a rich source of bioactive compounds. Thus, it might serve as a promising candidate for mediating the green synthesis of AgNPs. Its leaves have been used in folk medicine for the treatment of tuberculosis, stress, insomnia, helminthiasis, coughs, and colds (Díaz-Velasco et al., 2022; Patil et al., 2015).

Flavonoids, phenolics, glycosides, cyanogenic compounds, passifloricins, passifetilactone, α-pyrones, polypeptides, and alkaloids are the constituents of *P. foetida* associated with its biological activities (Nguyen et al., 2015; Song et al., 2018; Van Linh et al., 2022). However, this plant is frequently seen as a crop-harming weed. Previously, studies reported that the utilisation of plant extracts in the synthesis of AgNPs exhibited higher antioxidant activities compared to crude extracts (Gecer, 2025; Moges, 2022; Mazumder, 2025; Thapa et al., 2026). Although *P. foetida*-mediated AgNP synthesis has been reported, optimisation procedures for synthesis parameters such as pH, extract concentration, and temperature are lacking (Elangovan et al., 2022; Palanisamy et al., 2024). Traditional approaches for optimising reaction parameters vary only one factor at a time while keeping others constant. These methods are time-consuming and do not account for the simultaneous effects and interactions among multiple variables. In contrast, this study applied Response Surface Methodology (RSM) to systematically optimise *P. foetida*-mediated AgNPs synthesis. This approach provides a predictive model for reaction parameters, enabling maximised nanoparticle yield and improved physicochemical properties, representing the first optimisation model for this plant-based synthesis (Abu-Elghait et al., 2025; Ebrahimzadeh et al., 2020; Sati et al., 2025).

Therefore, this study focused on optimising the synthesis of AgNPs using *P. foetida* aqueous extract (PF-AgNPs) through RSM modelling, taking into account parameters such as pH, temperature, and extract concentration. Before PF-AgNPs synthesis, the chemical profile of *P. foetida* aqueous extract was analysed using ^{1}H NMR to identify the bioactive compounds that contribute to the reduction and stabilisation of silver ions. The UV-Vis, FTIR, XRD, and SEM analyses were carried out to confirm the formation and physicochemical characteristics of the obtained PF-AgNPs.

MATERIALS AND METHODS

Materials and Chemicals

The leaves of *P. foetida* were collected around the coastal areas of Universiti Malaysia Terengganu (UMT), the east coast of Peninsular Malaysia. The plant was identified by a Botanist, Dr Shamsul Khamis and the voucher specimen (ID068/ 2023) was deposited at the Herbarium Universiti Kebangsaan Malaysia (UKM), Malaysia. Leaf samples were cleaned with tap water and stored at -80°C before freeze-drying. A 20 g freeze-dried powder sample was extracted using distilled water at a ratio of 1:20 with ultrasonication at a temperature of 40°C. The water extract filtrate was then stored at -80°C for two days and subsequently freeze-dried to obtain a *P. foetida* aqueous extract.

All the chemicals and reagents used in this study were purchased from Sigma-Aldrich, Merck, Systerm, and Bendosen. The deuterium oxide (D_2O) containing 0.1% trimethylsilylpropanoic acid (TSP) sodium salt from Merck (Darmstadt, Germany)

was used for ^{1}H NMR analysis at 500 MHz (Varian Inc., California, USA). The chemicals used in AgNPs synthesis comprised silver nitrate ($AgNO_3$, Systerm), sodium hydroxide (1N NaOH, Bendosen), hydrochloric acid (1N HCl, Bendosen), and distilled water. In addition, chemicals including sodium nitroprusside (Solarbio), phosphate-buffered saline (PBS, Gibco), Griess reagent (Sigma-Aldrich), gallic acid (Sigma-Aldrich), quercetin (Aldrich), 2,2-diphenyl-1-picrylhydrazyl (DPPH, Sigma-Aldrich), methanol (Merck), and dimethyl sulfoxide (DMSO, Solarbio) were used for antioxidant assays. All the chemicals were of analytical grade.

Chemical Profiling of *P. foetida* Aqueous Extract by ^{1}H NMR Analysis

About 30 mg of *P. foetida* aqueous extract was added with 0.7 mL of D_2O containing 0.1% TSP, and the solution was sonicated at room temperature for 15 min (Maulidiani et al., 2018). After that, the solution was centrifuged at 13000 rpm for 5 min, and the supernatant was taken and transferred to an NMR tube. The ^{1}H NMR analysis was performed at a frequency of 500 MHz at room temperature 25 °C, with the parameters applied including number of scans of 64 and relaxation delay time of 2.0 s. The residual water signal at δ_H 4.70-4.90 ppm was suppressed through a pre-saturation method. Metabolite identification was conducted by comparing the characteristic signals of metabolites present in the ^{1}H NMR extract spectrum with the reference standards available via Chenomx Profiler software (Chenomx Inc., Edmonton, Canada) and literature data.

Design of Experiment via Response Surface Methodology (RSM) Modelling

Response Surface Methodology (RSM) was utilised to develop predictive models for optimising the experimental conditions of AgNPs synthesis. A Box-Behnken design (BBD) with three factors at three levels was employed to assess the effects of temperature, *P. foetida* extract concentration, and pH on the UV intensity of the resulting AgNPs (Table 1). A total of 14 experimental combinations of the independent variables. To model and analyse the relationship between the independent variables (temperature, *P. foetida* extract concentration, and pH) and the response variable (UV-Vis spectroscopy intensity), a second-order multiple linear regression model was used. This relationship is expressed by the following Equation 1:

$$Y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum \beta_{ij} x_i x_j \quad [1]$$

where: In the regression model, Y represents the response variable (UV intensity), while x denotes the independent variables, with x_i and x_j representing the experimental

parameters, where i, j ranges from 1 to k. The model includes β as the regression coefficients, where β_0 is the intercept, β_i and β_{ii} represent the linear and quadratic coefficients, and β_{ij} indicates the interaction coefficients between variables. The term ϵ accounts for the random error. Model evaluation and statistical significance were assessed at a 95% confidence interval with $p < 0.05$.

Table 1
Selected parameters and their three-level

Factor	Units	Minimum	Maximum	Coded Low	Coded High	Mean
A: Temperature	ºC	30	70	-1	+1	50
B: Concentration of *P. foetida* extract	g/mL	0.1	0.3	-1	+1	0.2
C: pH		4	10	-1	+1	7

Synthesis of *P. foetida*-Silver Nanoparticles (PF-AgNPs)

$AgNO_3$ was used for the eco-friendly synthesis of AgNPs with *P. foetida* as the reducing and capping agents. The process involved mixing freeze-dried *P. foetida* extract solution with a 1 mM $AgNO_3$ solution in a 1:9 ratio within a 40 ml volume. The mixture was left at room temperature for 24 hours to observe at a range of 380-480 nm the reduction of Ag^+ ions to Ag^0 using a UV-Vis spectrophotometer (UV-1900i, Shimadzu Corp, UK) (Egodawaththa et al., 2022).

Characterisation of PF-AgNPs

The formation of PF-AgNPs was monitored using a UV–Vis spectrophotometer (UV-1800 Shimadzu 240V, Japan) within a wavelength range of 200 to 800 nm (Akash & Rehman, 2020). The functional groups were responsible for the reduction of Ag^+ to Ag^0, and the stabilisation of NPs was characterised using FT-IR spectroscopy (FT-IR INVENIO-S spectrometer, Bruker, Germany) with an attenuated total reflectance (ATR) accessory within the range of 4000–400 cm^{-1} at a resolution of 4 cm^{-1} with 16 scans. X-ray diffraction (XRD) was used to examine the crystallographic structure using a high-resolution Rigaku SmartLab, Japan, equipped with Cu-Kα radiation (λ = 1.5406 Å) at an operating voltage of 40 kV and current of 30 mA. Meanwhile, the diffraction was recorded at a range of 20–80°C (2θ) with a step size of 0.01° and a scan rate of 2°/min, and the crystallite size of the NPs was calculated using the Debye-Scherrer equation (Singh et al., 2018). The surface morphology of the synthesised AgNPs was characterised by scanning electron microscopy (SEM) performed using the VPSEM model LEO1455, and the elemental composition was determined by energy dispersive X-ray spectroscopy (EDX) analysis.

Before SEM imaging, AgNPs were coated with a thin layer of gold using a sputtering device and carried out at an accelerating voltage of 20 kV to obtain detailed surface morphology and elemental composition.

Antioxidant Assays

The antioxidant assays were performed using DPPH and nitric oxide (NO) scavenging methods Abdul-Hamid et al. (2019). Briefly, for the DPPH assay, 50 µL of the sample was mixed with 100 µL of DPPH reagent (5.9 mg/100 mL) in a 96-well plate and incubated for 30 minutes in the dark (Shata et al., 2025). Then, the absorbance was measured at λ_{max} = 515 nm. Quercetin was used as a positive control and methanol as a negative control in this experiment, and the percentage of inhibition was calculated using Equation 2. For the NO scavenging assay, 60 µL of the sample was mixed with 10 mM of PBS solution of sodium nitroprusside (SNP) in a 96-well plate and incubated for 150 minutes under light. Afterwards, 60 µL of Griess reagent (1% sulphanilamide, 0.1% NED, and 2.5% H_3PO_4) was added, and absorbance was measured at λ_{max} = 550 nm (Mohamad Ali et al., 2020; Sreejayan & Rao, 1997). The positive control was quercetin, while the negative control was DMSO.

$$\%\text{Inhibition} = \frac{(Ac - As)}{(As)} \text{ x } 100 \qquad [2]$$

Where A_c is the absorbance of the control or blank, A_s is the absorbance of the plant extract or AgNPs or standard.

Statistical Data Analysis

Analysis of variance (ANOVA) was used to determine the statistical significance of the regression model and its parameters. Model adequacy was assessed based on the *P*-value (< 0.05), F-value, R^2, adjusted R^2, and lack-of-fit statistics. All statistical analyses and 2D surface plot visualisations were conducted using the RSM package in R-Studio (https://www.r-project.org/), which enabled the identification of optimal synthesis conditions with a maximum of variable responses, indicating efficient AgNPs formation.

RESULTS AND DISCUSSION

Chemical Profile of *P. foetida* Aqueous Extract

In a green synthesis of NPs, it is hypothesised that the presence of the -OH group in primary and secondary plant metabolites is attributed to their capability to reduce silver ions to AgNPs.

In this study, the metabolite profile of *P. foetida* extract was analysed using ^{1}H NMR spectroscopy. Mainly, the aqueous extract of *P. foetida* consisted of aromatic (δ_H 6.0–8.5 ppm), glycoside (δ_H 3.1–3.9 ppm), and aliphatic (δ_H 0.7–2.5 ppm) compounds (Figure 1). Table 2 presents the identification of ^{1}H NMR characteristics of the metabolites in the *P. foetida* aqueous extract. Several aromatic compounds from the flavonoid class were identified, including vitexin derivatives at δ_H 8.01 (d, J= 8.5 Hz, H2' and H6'), 7.04 (d, J=8.5 Hz, H3' and H5'), and 6.69 (s, H3) as well as apigenin at δ_H 7.89 (d, J= 8.5 Hz, H2' and H6'), 6.84 (d, J= 8.5 Hz, H3' and H5') and 6.67 (s, H3). Amino acids such as phenylalanine, 4-hydroxyphenyllactic acid, glutamic acid, glutamine, alanine, valine, isoleucine, and leucine were also identified in the ^{1}H NMR spectrum. Further observation demonstrated that glycosidic compounds such as flavonoid glycosides were the major components of *P. foetida* aqueous extract. The chemical constituents of *P. foetida,* such as flavonoids, have been linked to its pharmacological effects, including antioxidant (Foudah et al., 2019; Ning et al., 2023), anti-inflammatory (Nguyen et al., 2015), and anticancer activities (Ponsuwan et al., 2024).

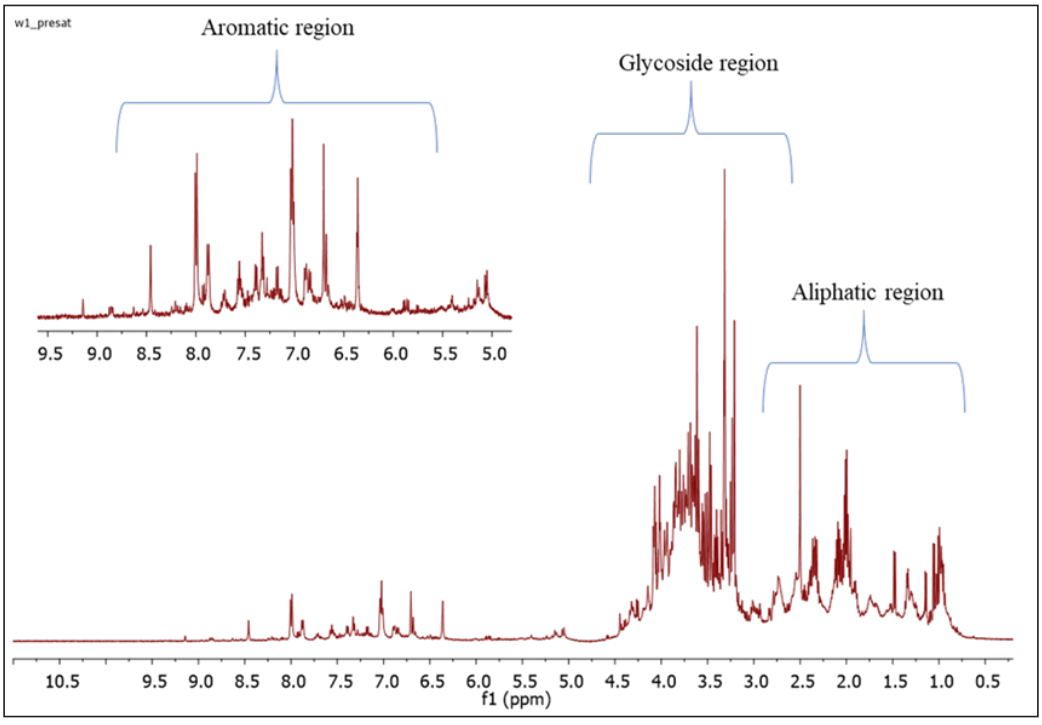


Figure 1. ^{1}H NMR spectrum of *P. foetida* aqueous extract in D_2O containing 0.1% TSP

Table 2
^{1}H NMR characteristics of the metabolites in P. foetida *aqueous extract*

HMDB ID	Compounds	Class	Chemical shift
HMDB0000687	Leucine	Amino acids	3.73 (m, ov.), 1.75–1.67 (m, ov.), 0.94 (t, J =4.75 Hz)
HMDB0250806	Valine		3.62 (d, J =4.45 Hz), 2.26 (m, ov.), 1.04 (d, J=7.05 Hz), 0.99 (d, J=7.00 Hz)
HMDB0000167	Threonine		4.25 (m, ov.), 3.60 (d, ov.), 1.33 (d, J=6.50 Hz)
HMDB0001310	Alanine		3.79 (q, ov.), 1.47 (d, J= 7.25 Hz)
HMDB0001080	4-aminobutyrate		3.01 (m, ov.), 2.29 (t, ov.), 1.90 (m, ov.)
HMDB0000162	Proline		4.12 (m, ov.), 3.41 (m, ov.), 3.33 (m, ov.), 2.34 (m, ov.), 2.06–1.98 (m, ov.)
HMDB0000020	4-hydroxy phenyllactic acid		7.16 (d, J= 8.50 Hz), 6.83 (d, J=8.55 Hz), 4.20 (m, ov.), 2.99 (m, ov.), 2.80 (m, ov.)
HMDB0000904	Citrulline		6.36 (br s), 3.75 (m, ov.), 3.15 (m, ov.), 1.84 (m, ov.), 1.59 (m, ov.)
HMDB0000172	Isoleucine		3.67 (d, ov.), 1.99 (m, ov.), 1.46 (m, ov.), 1.25 (m, ov.), 0.99 (d, J=7.00 Hz), 0.94 (t, J= 7.45 Hz)
HMDB0000182	Lysine		3.75 (t, ov.), 3.01 (t, J=8.15 Hz), 1.91 (m, ov.), 1.72 (m, ov.), 1.48 (m, ov.)
HMDB0000696	Methionine		3.85 (m, ov.), 2.63 (t, ov.), 2.19 (m, ov.), 2.13 (s)
HMDB0000159	Phenylalanine		7.42 (m, ov.), 7.37 (m, ov.), 7.31 (d, J=6.80 Hz), 3.99 (m, ov.), 3.28 (m, ov.), 3.12 (m, ov.)
HMDB0000191	Aspartate		3.89 (m, ov.), 2.79 (m, ov.), 2.67 (m, ov.)
HMDB0000929	Tryptophan		7.72 (d, J=7.70 Hz), 7.54 (d, J=7.65 Hz), 7.31 (s), 7.26 (m, ov.), 7.20 (m, ov.), 4.06 (m, ov.), 3.46 (dd, J=14.80 Hz, 4.9 Hz)
HMDB0000158	Tyrosine		7.19 (d, ov.), 6.87 (d, J=8.45 Hz), 3.93 (m, ov.), 3.20 (dd, ov.), 3.05 (m, ov.)
HMDB0000296	Uridine	Nucleosides	7.86 (d, J=8.55 Hz), 5.90 (m, ov.), 4.34 (m, ov.), 4.24 (t, ov.), 4.12 (m, ov.), 3.90 (m, ov.), 3.80 (m, ov.)

Table 2 (*continued*)

HMDB ID	Compounds	Class	Chemical shift
HMDB0304632	Glucose	Sugars	5.22 (d, J=3.65 Hz), 4.64 (d, J=8.10 Hz), 3.89(dd, ov.), 3.82 (m, ov.), 3.71 (m, ov.), 3.48 (m, ov.), 3.40 (m, ov.), 3.22 (m, ov.)
HMDB0000258	Sucrose		5.39 (d, J=3.45 Hz), 4.21 (d, ov.), 4.04 (t, J=8.80 Hz), 3.88 (m, ov.), 3.76 (t, 9.70 Hz) 3.66 (d, J=11.5 Hz), 3.56 (m, ov.)
HMDB0000660	Fructose		4.10 (m, ov.), 4.03 (dd, J=12.50 Hz, 1.20 Hz), 3.99 (m, ov.), 3.89–3.78 (m, ov.), 3.71–3.67 (m, ov.) 3.59–3.54 (m, ov.)
HMDB0000765	Mannitol		3.86 (m, ov.), 3.79 (m, ov.), 3.74 (m, ov.), 3.67 (m, ov.)
HMDB0000163	Maltose		5.39 (d, J=3.45 Hz), 5.22 (d, J=3.85 Hz), 4.65 (d, ov.), 3.97 (m, ov.), 3.97–3.58 (m, ov.), 3.42 (t, ov.), 3.28 (m, ov.)
HMDB0000646	Arabinose		5.29 (d, ov.), 5.23 (m, ov.), 4.51 (d, J=7.60 Hz), 4.11–3.64 (m, ov.), 3.51 (m, ov.)
HMDB0000098	Xylose		5.19 (d, ov.), 4.56 (d, J=7.85 Hz), 3.92 (m, ov.), 3.67– 3.62 (m, ov.), 3.51 (m, ov.), 3.42 (t, ov.), 3.30 (t, ov.), 3.22 (m, ov.)
HMDB0000975	Trehalose		5.18 (d, J=6.55 Hz), 3.86–3.76 (m, ov.), 3.46 (t, J=9.90 Hz)
HMDB0000956	Tartaric acid	Organic acids	4.34 (s)
HMDB0000254	Succinic acid		2.38 (s), 2.43 (s)
HMDB0005807	Gallic acid		7.02 (s)
HMDB0000691	Malonic acid		3.09 (s)
HMDB0000148	Glutamic acid		3.75 (m, ov.), 2.36–2.04 (m, ov.)
HMDB0000641	Glutamine		6.87 (s), 3.77 (t, ov.), 2.47–2.11 (m, ov.)
HMDB0000639	Mucic acid		4.25 (s), 3.95 (s)
HMDB0000625	Gluconic acid		4.13 (d, ov.), 4.03 (t, ov.), 3.84 (m, ov.), 3.77 (m, ov.), 3.66 (m, ov.)
HMDB0000097	Choline	Vitamins	4.06 (m, ov.), 3.51 (m, ov.), 3.19 (s)
HMDB0000244	Riboflavin		7.99 (s), 7.97 (s), 5.13 (m, ov.), 4.96–3.73 (m, ov.), 2.48 (s)

Table 2 (*continued*)

HMDB ID	Compounds	Class	Chemical shift
HMDB0128581	Hispidulin	Flavonoids	8.00 (d, J=8.75 Hz), 6.47 (s)
HMDB0301981	Isoorientin		6.71 (s)
HMDB0030614	Orientin		6.65 (s), 6.33 (s)
HMDB0005800	Luteolin		7.39 (d, J= 7.15 Hz), 6.85 (d, J= 8.35 Hz), 6.57 (s), 6.52 (d, J=3.45 Hz), 6.32 (d, J=3.85 Hz)
HMDB0030614	Vitexin		8.01 (d, J = 8.50 Hz), 7.04 (d, J = 8.50 Hz), 6.69 (s)
HMDB0002124	Apigenin		7.89 (d, J=8.50 Hz), 6.84 (d, J=8.50 Hz), 6.67 (s)
HMDB0253708	Isovitexin		6.67 (s)
HMDB0005801	Kaempferol		8.01 (d, J=9.00 Hz), 7.04 (d, J=9.00 Hz)

Note. Ov. is overlapped with other signals. Letters in parentheses refers to the signal multiplicities: s, singlet; d, doublet; t, triplet; dd, doublet of doublets; q, quartet; and m, multiplet

Optimisation of PF-AgNPs Synthesis using RSM Modelling

The optimisation experiments were conducted to synthesise the PF-AgNPs according to the BBD matrix, and evaluated the effects of temperature (X_1), extract concentration (X_2), and pH (X_3) on UV-Vis intensity, NO scavenging activity, and DPPH free radical activity (Table 3). The correlation matrix, as summarised in Figure 2a, indicated that pH exhibited the strongest positive correlation with UV-Vis intensity (r = 0.83), while temperature and concentration showed weaker correlations (0.28 and 0.16, respectively). This finding reinforces the hypothesis that pH plays a key role in the formation of AgNPs. Importantly, pH emerged as the most statistically significant individual factor ($P < 0.001$) (Seifipour et al., 2020), which highlighted the dominant influence of pH on UV absorbance intensity compared to other synthesis parameters. The synergistic effect between pH and temperature further suggests a cooperative influence on nanoparticle synthesis efficiency, underscoring the importance of optimising these conditions to enhance both structural and functional outcomes.

The regression model for UV-Vis intensity was statistically significant (F = 5.52; $P = 0.0052$), with a high coefficient of determination ($R^2 = 0.985$) on Figure 2b, a low root mean square error (RMSE = 0.097), and no significant lack of fit (P = 0.15), indicating good model adequacy. Among the three variables that were shown as a linear model,

Table 3
A Box-Behnken design under RSM for the synthesis of PF-AgNPs

Sample	Run	Factor 1	Factor 2	Factor 3	Respond 1	Respond 2	Respond 3
		Temp. ºC	Conc. (g/mL)	pH	Intensity (a.u)	NO %	DPPH %
A1B2C1	1	30	0.1	7	0.32	14.58	25.19
A2B1C1	2	30	0.2	4	0.26	10.83	27.00
A2B3C1	3	30	0.2	10	1.04	28.97	19.93
A3B2C1	4	30	0.3	7	0.62	22.21	14.29
A1B1C2	5	50	0.1	4	ND	10.94	20.35
A1B3C2	6	50	0.1	10	1.25	23.99	7.81
A2B2C2	7	50	0.2	7	0.74	22.21	15.02
A2B2C2	8	50	0.2	7	0.68	22.72	18.47
A3B1C2	9	50	0.3	4	1.77	35.15	19.60
A3B3C2	10	50	0.3	10	2.07	20.48	10.38
A1B2C3	11	70	0.1	7	0.72	27.18	19.04
A2B1C3	12	70	0.2	4	ND	21.18	28.44
A2B3C3	13	70	0.2	10	2.92	38.70	35.64
A3B2C3	14	70	0.3	7	0.94	18.23	23.28

Note. A= Concentration; B = pH; C = Temperature; ND = Not Detected; % = Inhibition (with the inhibition of *P. foetida* aqueous extract being 17.16%).

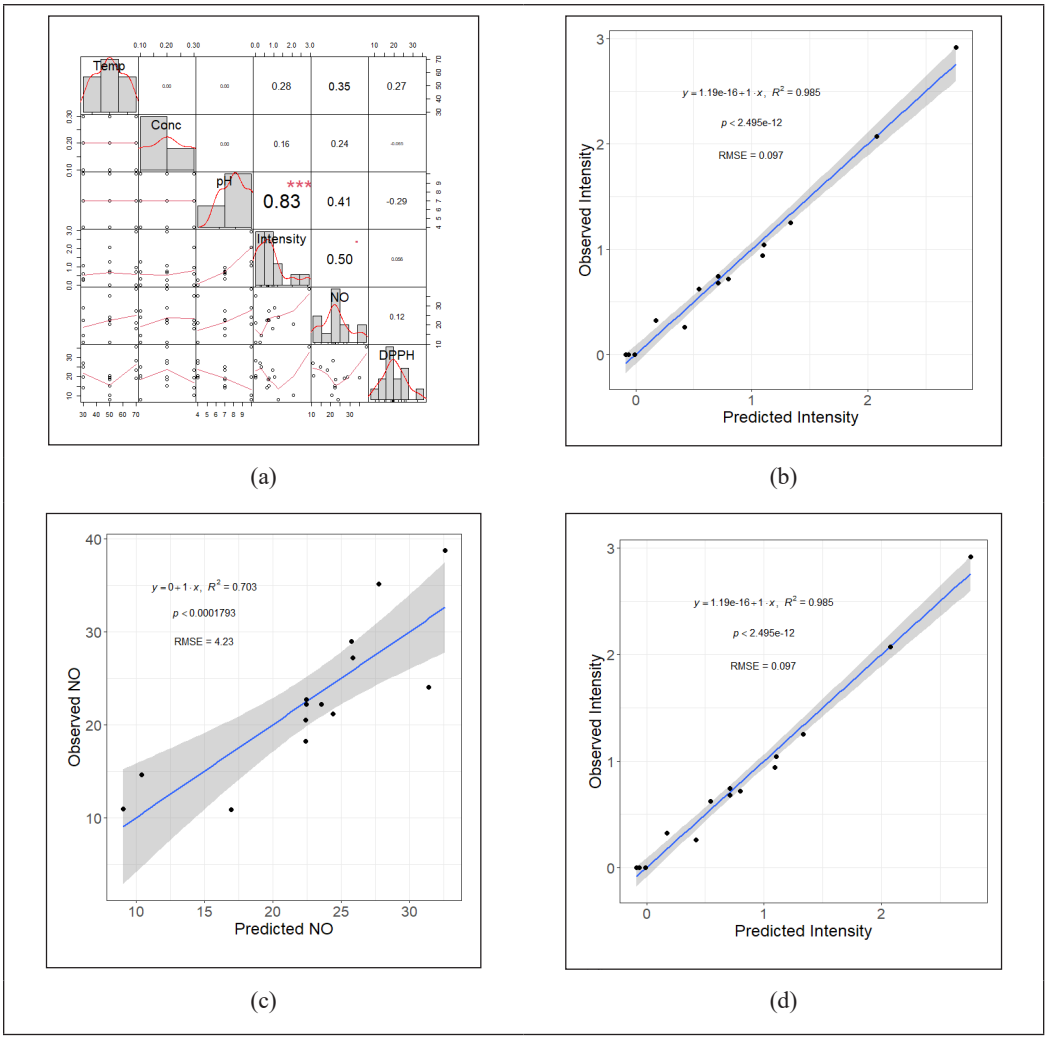


Figure 2. (a) Correlation between all factors with variable responses. Regression values; (b) intensity response; (c) NO scavenging response; and (d) DPPH scavenging response

pH (X_3) had the most significant effect ($P < 0.001$), followed by temperature (X_1; $P < 0.05$), while extract concentration was not statistically significant ($P = 0.059$). It is important to emphasise that the P values from this data set fall short of the strict < 0.0001 threshold (Patil et al., 2024). which suggests that the model lacks statistical significance. However, terms within the model that exhibit P values < 0.05 are still regarded as significant, indicating a greater probability (up to 5%) that the observed results occurred by chance (Desai et al., 2008; Ndikau et al., 2017). Further, the significant interaction was also observed between temperature and pH (X_1X_3; $P < 0.01$), showing the two-factor interaction. The maximum intensity (~2.95) at high temperature (~70°C), pH (~10), and higher concentration (~0.28%)

was suggested by the response surface plots (Figure 3a). However, at pH 4, NPs agglomeration was more evident, particularly at high temperatures, which may compromise product stability (Al-Gebory & Mengüç, 2018; Cativa et al., 2020).

The regression equation is graphically represented by two-dimensional response surface plots for NO scavenging activity that are statistically significant (F = 4.01; $P = 0.01598$) in Table 4, according to an R^2 of approximately 0.703 and an RMSE of 4.23 (Figure 2c). The model did not, however, appear to adequately capture the variability of the data, as indicated by a significant lack of fit ($P = 0.02899$). No individual variable or interaction term was statistically significant ($P > 0.05$). For the interaction factor BC (X_2X_3; $P < 0.155$), there is a trend toward interaction, although it cannot yet be statistically categorised as significant. The surface response plots under various pH conditions show slight curvature, indicating a tendency for a two-factor interaction between concentration and pH, although not significant plots (Figure 3b).

The model for DPPH scavenging activity was statistically significant (F = 4.73; $P = 0.0091$) in Table 4, with an R^2 of approximately 0.857 and RMSE of 2.68 (Figure 2d), as well as no significant lack of fit ($P = 0.30685$). The adequacy of the quadratic model was checked by ANOVA, and the temperature (X_1^2) was significant ($P = 0.0391$). The response surface plots in Figure 3c show a downward curve due to the DPPH activity decreasing in a range of moderate temperatures (~50–60°C), which marked the green area as dark, and the red to white area indicates higher values of DPPH activity.

Canonical path (Appendix) and response surface analyses (Figure 3a, 3b and 3c) showed that the stationary points for intensity, NO, and DPPH corresponded to saddle points, indicating the absence of true interior optima within the experimental design space. For all responses, predicted values decreased toward the centre of the canonical distance and increased toward the boundaries, confirming that optimisation is governed by boundary conditions rather than interior stationary points. Accordingly, particle swarm optimisation was applied to identify the maximum predicted responses within the experimental region. The optimal conditions for intensity were a temperature of 70 °C, a concentration of 0.30 g/mL, and pH 10, yielding a predicted value of 2.97 (a.u.). For NO, the maximum response (39.16 %) was predicted at 70 °C, 0.10 g/mL, and pH 10, while DPPH reached its optimum (31.96 %) at 70 °C, 0.242 g/mL, and pH 10. These results indicate that all three responses are strongly influenced by high temperature and alkaline pH, whereas the effect of concentration is variable, suggesting that careful adjustment of these factors can maximise the desired outcomes. The predicted optima provide a clear guide for experimental validation, which will further confirm the conditions for maximum response. These results were similar to earlier reports, thus confirming the effectiveness of RSM in determining the optimal conditions for the synthesis of plant-based AgNPs with desired physicochemical and biological properties (Aliero et al., 2024; Fernando & Zhou, 2019; Krupa et al., 2016).

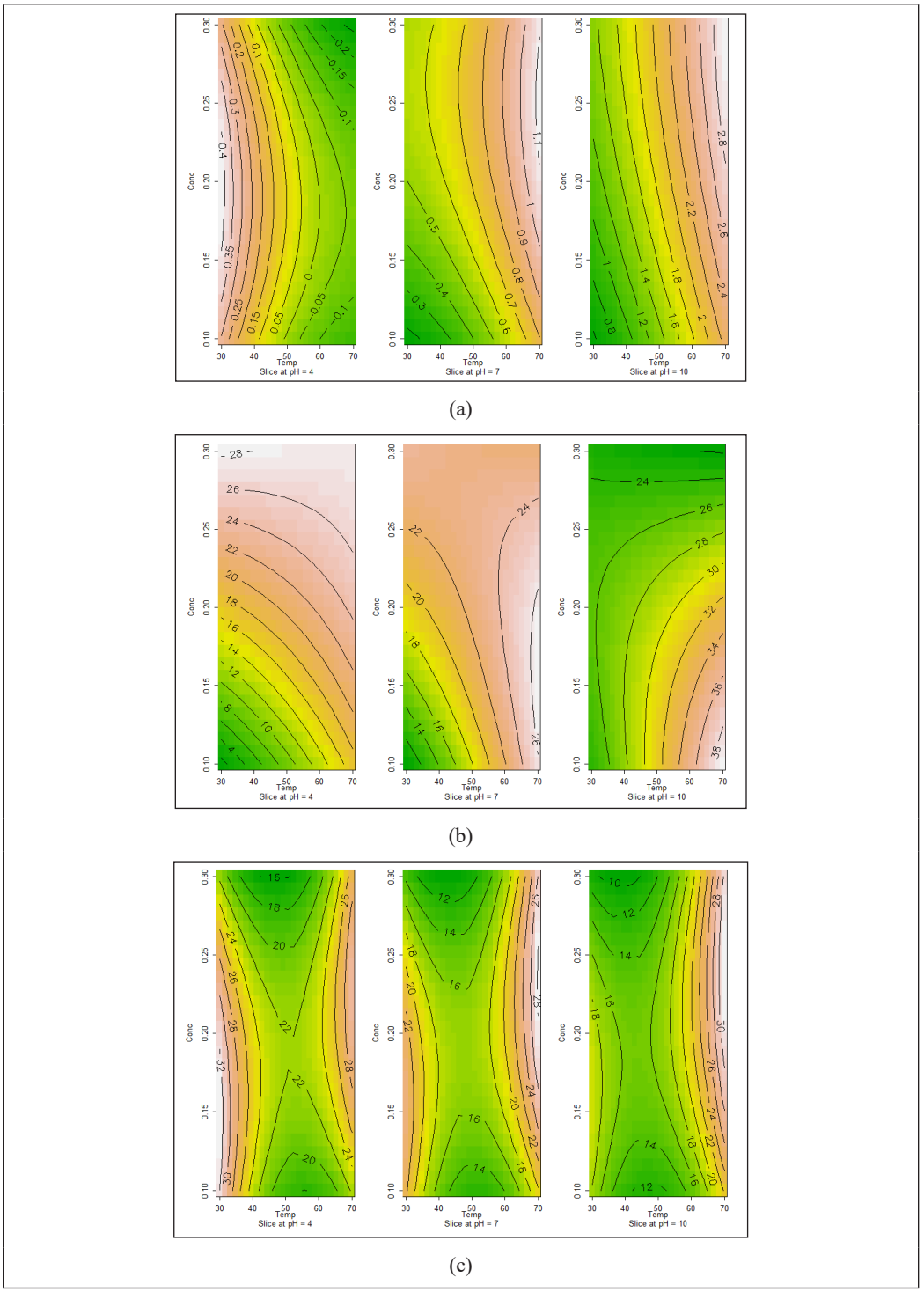


Figure 3. 2D response surface curves showing correlation: (a) intensity response; (b) NO scavenging response; and (c) DPPH scavenging response

Table 4
Optimisation studies of variable response using ANOVA

Source	Sum of Squares	Mean Squares	F-value	P-value	Remarks
Intensity responses					
Model	0.71	0.128623	5.52	0.005259	Significant
A-Temperature	0.2925	0.064311	4.5482	0.010432*	Significant
B-Concentration	0.1675	0.064311	2.6045	0.059763	
C-pH	0.8775	0.064311	13.6446	0.000167 **	Significant
AB	-0.02	0.09095	-0.2199	0.836715	
AC	0.535	0.09095	5.8824	0.004174	Significant
BC	0.205	0.09095	2.254	0.087259	
A^2	0.0825	0.101685	0.8113	0.462693	
B^2	-0.1425	0.101685	-1.4014	0.233717	
C^2	0.2625	0.101685	2.5815	0.06123	
Residuals	0.1324	0.03309			
Lack of fit	0.1306	0.04352	24.1759	0.148148	
Pure error	0.0018	0.0018			
NO Scavenging					
Model	22.465	5.6	4.0116	0.01598*	Significant
A-Temperature	3.5875	2.8	1.2812	0.26934	
B-Concentration	2.4225	2.8	0.8652	0.43575	
C-pH	4.255	2.8	1.5196	0.20323	
AB	-4.145	3.9598	-1.0468	0.35429	
AC	-0.155	3.9598	-0.0391	0.97065	
BC	-6.93	3.9598	-1.7501	0.155	
A^2	0.1825	4.4272	0.0412	0.96909	
B^2	-2.0975	4.4272	-0.4738	0.66036	
C^2	2.2725	4.4272	0.5133	0.63479	
Residuals	250.881	62.720			
Lack of fit	250.751	83.584	642.7049	0.02899	
Pure error	0.130	0.130			
DPPH Scavenging					
Model	16.7450	3.5432	4.726	0.009131	Significant
A-Temperature	2.4987	1.7716	1.4105	0.231226	
B-Concentration	-0.6050	1.7716	-0.3415	0.749913	
C-pH	-2.7037	1.7716	-1.5262	0.201669	
AB	3.7850	2.5054	1.5107	0.20538	
AC	3.5675	2.5054	1.4239	0.227574	
BC	0.8300	2.5054	0.3313	0.75706	
A^2	8.4612	2.8011	3.0206	0.039138*	Significant
B^2	-4.7562	2.8011	-1.6980	0.164746	
C^2	2.5463	2.8011	0.9090	0.414764	
Residuals	100.43	25.108			
Lack of fit	94.48	31.494	5.2920	0.30685	
Pure error	5.95	5.951			

UV-visible Spectroscopy Analysis

The formation of AgNPs can be observed by the colour change of the reaction mixture from yellowish to brownish-red, indicating the development and stabilisation of the NPs, as seen in Figure 4a. This also indicates that the phytochemicals of *P. foetida* were involved in the biosynthesis of AgNPs as a reducing agent. The distinctive surface plasmon resonance (SPR) band of the formed AgNPs have absorption bands of 380-480 nm (Egodawaththa et al., 2022). According to our observations, synthesis with a 24-h incubation time exhibited an absorption band for SPR compared to a 1-h incubation. Thus, for the RSM design experiments, synthesis with a 24-h incubation was chosen. For instance, the optimal condition A2B3C3 has SPR at 420 nm. Similar findings were reported for AgNPs synthesised from *Cicer arietinum* (Mouriya et al., 2023), *Sambucus ebulus* (Karan et al., 2024), *Diospyros malabarica* (Bharadwaj et al., 2021), *Lythrum salicaria* (Corciovă et al., 2022), and *Pyrus pashia* (Khanal et al., 2023).

The temperature effect has been evident in the increased synthesis rate of AgNPs formation with rising temperature. Previous studies indicate that higher temperature causes more thermal activity and a faster rate of crystal formation around the nucleus, which increases absorption (Liu et al., 2020; Xu et al., 2020). This improvement was marked by the growth in the size of NPs (Ebrahimzadeh et al., 2020). Figure 4b. illustrates the variation in the absorption spectrum of PF-AgNPs synthesised at temperatures of 30 and 70ºC, with a constant composition of extract concentration of 0.2% and pH 10. The temperature of 50ºC was not included in this variation as it adheres to the standard design of the BBD. At 0 hours, both samples (Figure 4b) exhibit the same colour with the starting SPR at 384 nm. After 24 hours at temperatures of 30ºC and 70ºC, the respective SPRs are 385 nm and 420 nm, exhibiting reddish-brown and brownish-red colours. The peak plasmon resonance shifts to longer wavelengths (redshift), and the intensity is greater, especially at the temperature condition of 70ºC. The resulting brownish-red colour indicated a significantly higher formation of AgNPs. These results are consistent with previous reports suggesting that an increase in temperature accelerates the reaction rate and enlarges the particle diameter (Liu et al., 2020). Optimisation studies confirm a temperature of around 70°C as ideal. An increase in extract concentration enhances the bioactive compounds involved in the reduction process.

Figure 4c illustrates the variation in the absorption spectrum of PF-AgNPs synthesised at extract concentrations of 0.1% and 0.3% (w/v), with constant pH and temperature compositions set at pH 10 and 50ºC. The concentration of 0.2% was not included in this variation as it adheres to the standard design of the BBD. At 0 hours, both exhibit the same SPR at 381 nm with higher intensity at the 0.3% concentration. After 24 hours, a shift in SPR occurs at extract concentrations of 0.1% and 0.3%, with respective SPRs of 415 nm and 406 nm, both exhibiting a brownish-red colour. The peak plasmon resonance shifts to longer

wavelengths (redshift), with a larger plasmon shift at the 0.1% concentration, indicating the formation of larger NPs compared to the 0.3% concentration. This study showed that the higher the extract concentration, the smaller the resulting NPs size compared to the lower concentration. It matched with earlier studies of AgNPs of *Spinacia oleracea* extract; 2% extract had a size of 173 nm, while 10% extract had a size of nm (Miranda et al., 2022). Although the sizes of AgNPs are also affected by other factors, the RSM result demonstrated that the extract concentration is not a significant factor.

pH is another crucial factor influencing the size, shape, and morphology of the synthesised AgNPs. The main influence of the reaction pH is its ability to alter the electrical charges of biomolecules, which can modify their reducing and capping abilities and the subsequent growth of NPs (Jain & Mehata, 2017). Figure 4d demonstrated the variation in the absorption spectrum of AgNPs synthesised at pH 4 and 10, with constant compositions of extract concentration 0.2% and temperature 50ºC, where pH 7 was not included in this variation as it adheres to the standard design of the BBD. At 0 hours, the detected SPR peak is only at pH 10, at 385 nm, indicating the formation of AgNPs. After 24 hours, a shift in SPR occurs at pH 10 to 406 nm, while at pH 4 experiences a decrease in intensity and no detected SPR peak. This significant difference is also supported by the colours, with pH 4 and pH 10 exhibiting light yellow to light purple (experiencing agglomeration) and yellowish to brownish-red, respectively. pH plays a role in regulating the redox reaction environment during the NPs formation. A lower pH results in a higher aggregation rate, leading to agglomeration (Fernando et al., 2020; Lau et al., 2016). A low pH creates an acidic condition where interparticle energy becomes neutral due to the presence of protons entering the neutrally charged AgNPs (Fernando & Zhou, 2019). On the other hand, pH 10 (higher pH) creates a base condition, leading to the stabilisation of hydroxyl ions and strengthening the negative charge (Wang et al., 2020). Similarly, in the earlier studies, pH 10 was identified as the optimal condition for the biosynthesis of AgNPs (Mardiyanto et al., 2023; Ndikau et al., 2017; Ndulini et al., 2024; Seifipour et al., 2020).

FT-IR Analysis

FT-IR spectroscopy analysis was conducted to identify the characteristic functional groups of *P. foetida* aqueous extract and PF-AgNPs that are related to the reduction and stabilisation of the NPs (Figure 5). The FT-IR spectrum of the *P. foetida* aqueous extract showed O-H stretching at 3256.48 cm^{-1}, indicating the presence of carbohydrates, proteins and polyphenols (Alkhulaifi et al., 2020; Bharadwaj et al., 2021; Kumari et al., 2020). The stretching vibrations of the C≡N (nitrite or nitrile) group were identified at 2118.59 cm^{-1}, and the C≡C (alkyne) around 2100 cm^{-1} (Jain & Mehata, 2017). The peak at 1593.13 cm^{-1} suggested the C=C stretching vibrations in double bonds for aromatic compounds or alkenes, as well as the vibrations of the carbonyl group C=O or amide (Arokiyaraj et

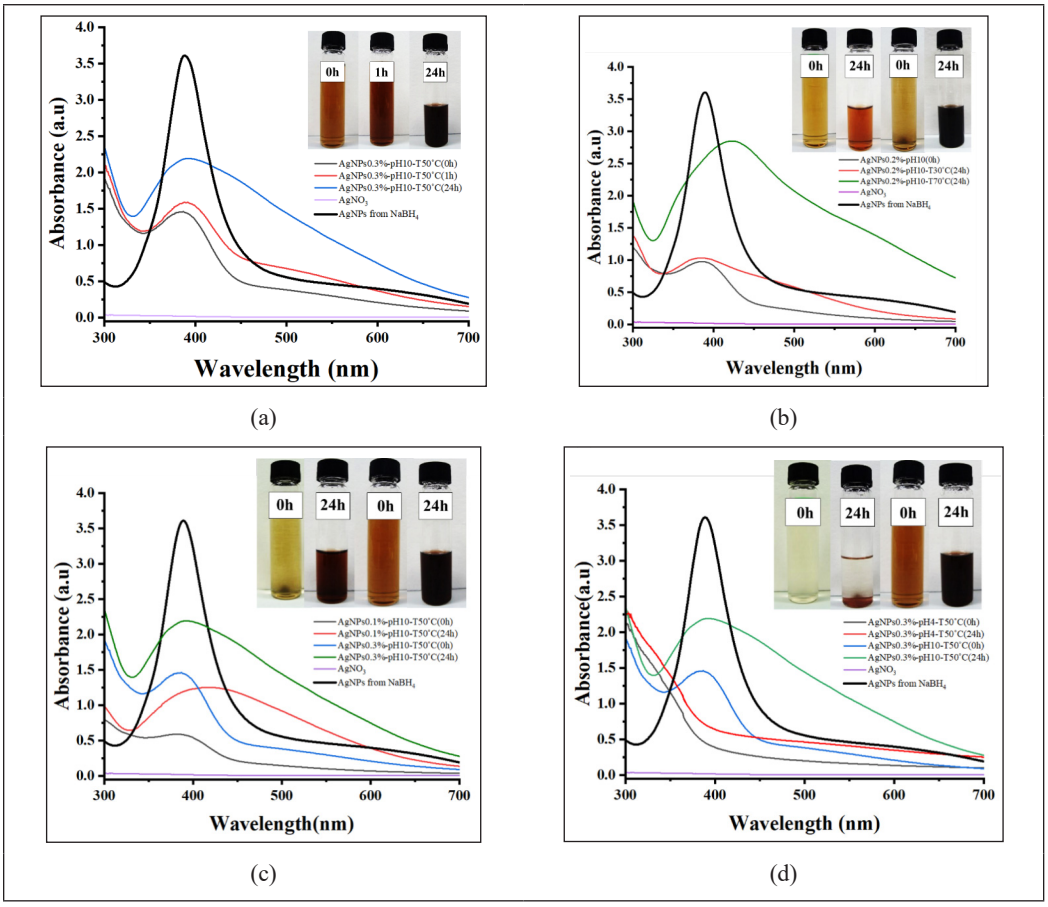


Figure 4. UV-visible spectroscopy results of the green synthesised PF-AgNPs: (a) effect of incubation time; (b) effect of temperature; (c) effect of extract concentration; and (d) effect of pH

al., 2014; Jaison et al., 2023). The bending (deformation) of C–H in aliphatic compounds was investigated by the presence of a band at 1399.44 cm^{-1} (Dobrucka, 2017). This band may also be associated with the symmetric vibration of the NO_2 group or the bending (deformation) of the C–H group in methyl ($-CH_3$) or methylene ($-CH_2^-$) groups (Mousavi et al., 2018). The stretching vibrations of the C-O bond of the alcohols and phenols at 1046.59 cm^{-1} (Riaz et al., 2022) or C-N, and 590.89 cm^{-1} indicates stretching vibrations of C-Cl (Nguyen et al., 2015).

Based on Figure 5a and Figure 5b, changes in the functional groups, as shown by the shift in absorption bands in comparison to the spectrum of *P. foetida* aqueous extracts as a standard, confirmed the successful formation of PF-AgNPs. Although the shifts were not highly pronounced, they were sufficient to suggest interactions between the active compounds in the extract and the synthesised PF-AgNPs. The vibrational data for the identified functional groups are summarised in Table 5. Remarkably, C-O-C stretching

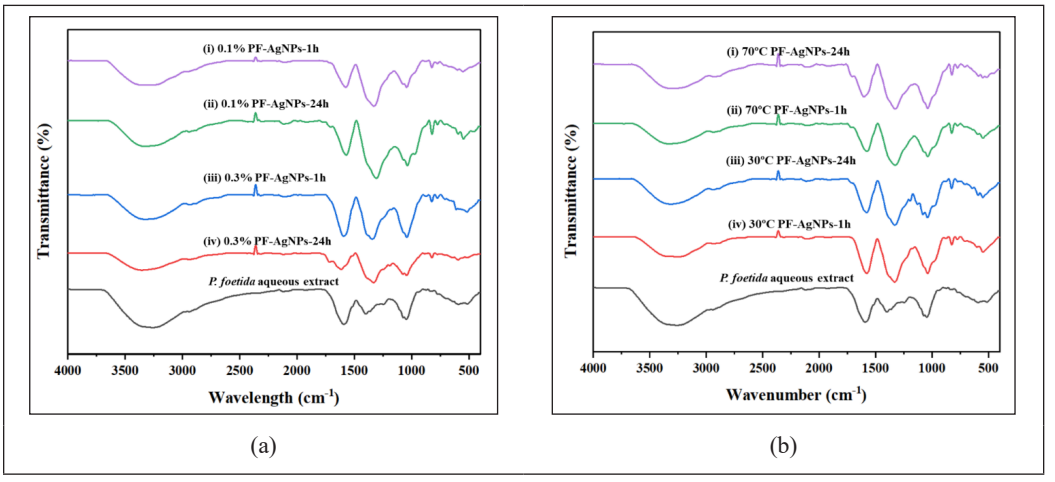


Figure 5. FTIR analysis of PF-AgNPs: (a) effect of concentration and (b) effect of temperature

Table 5
FT-IR peak assignments of P. foetida aqueous extract and synthesised PF-AgNPs

Possible assignments	*P. foetida* extract	PF-AgNPs (cm^{-1})			
		i	ii	iii	iv
Effect of different concentrations and reaction times (pH and temperature constant)					
v(O-H, H-bonding)	3256.48	3353.52	3326.07	3325.64	3352.54
v(C≡C or C≡N)	2118.59	ND	2115.72	ND	ND
v(C=C), Amide II	1593.13	1613.42	1592.9	1570.12	1574.11
v(C–H aliphatic)	1399.44	1331.61	1343.94	1307.72	1329.74
v(C–O)	1046.59	1043.08	1041.89	1036.77	1042.71
v(Aromatic C–H)	ND	823.57	824.1	822.85	823.48
v(Aromatic C–H), alkenes	ND	ND	775.32	ND	ND
v(C–Br or C–I)	590.89	594.36	517.12	547.62	552.47
Effect of different temperature and reaction times (pH and concentration constant)					
v(O-H, H-bonding)	3256.48	3280.98	3324.07	3238.57	3230.11
v(C≡C or C≡N)	2118.59	2114.89	2111.97	2116.14	2117.01
v(C=C), Amide II	1593.13	1603.29	1575.14	1577.99	1576.57
v(C–H aliphatic)	1399.44	1325.45	1326.8	1330.39	1330.27
v(C–O)	1046.59	1038.14	1038.44	1038.73	1039.13
v(Aromatic C–H)	ND	824.19	824.66	824.52	824.42
v(Aromatic C–H), alkenes	ND	773.51	774.18	775.1	775.13
v(C–Br or C–I)	590.89	550.01	548.55	547.27	547.36

Note. ND is not detected

identifies the fingerprint region of AgNPs spectra as identified at 1043, 1041, 1036, 1042, 1038, 1038, 1038, and 1039 cm^{-1} (Table 5). A larger shift in vibrational frequency frequently indicates a stronger interaction between the NPs and the capping agents. It is because the interaction can significantly alter the electronic environment around the functional groups involved, leading to more pronounced shifts in their vibrational frequencies (Jumabaev et al., 2025; Upadhyaya & Devi, 2018). Also, new peaks at 822.85-824.66 cm^{-1} and 773.51–753.2 cm^{-1} indicate C-H bending vibrations typical of aromatic compounds and alkenes, as well as suggest the presence of aromatic rings or unsaturated hydrocarbons (Khan et al., 2025). Additionally, the bands within the range of 517.12-594.36 cm^{-1} were identified as C-X stretching vibrations, as from halogenated compounds such as C-Cl or C-Br. These findings suggest that bioactive compounds such as flavonoids and alkaloids present in *P. foetida* extract play key roles in the reduction, formation, and stabilisation of the PF-AgNPs.

X-Ray Diffraction Analysis

The PF-AgNPs diffraction was compared with that of diffraction AgNPs prepared using $NaBH_4$ as a reducing agent as a standard reference for AgNPs formation. The XRD diffraction peaks of the $NaBH_4$-AgNPs appeared at 2θ angles of 38.08°, 44.25°, 64.37°, and 77.31°, whereas the PF-AgNPs incubated for 1 hour and 24 hours exhibited closely matching peaks at 2θ values of 38.08°, 44.26°, 64.39°, and 77.33°, and 38.07°, 44.24°, 64.36°, and 77.29°, respectively. These four peaks correspond to the (111), (200), (220), and (311) crystallographic planes of a face-centred cubic (fcc) structure, in accordance with the Joint Committee on Powder Diffraction Standards (JCPDS) file No. 04-0783.

These results confirm that the synthesised PF-AgNPs were crystalline and exhibited an fcc structure, which matched the values established earlier by (Anguraj et al., 2024; Elangovan et al., 2022; Khatun et al., 2024). The other peaks observed in Figure 6a were also identified as phytochemicals from the plant extract residue (Behravan et al., 2019; Elangovan et al., 2022; Jain & Mehata, 2017). Additionally, the diffractogram revealed a slight broadening in the (111) plane after about 24 hours of incubation as opposed to 1 hour, which suggested the beginning of NPs aggregation. Peak broadening may result from strain and defects in the crystal lattice caused by this aggregation process (Gulyaev et al., 2022).

SEM and EDX Analyses

The SEM analysis was conducted to investigate the synthesised PF-AgNPs' surface morphology, along with the EDX analysis to determine the elemental composition. The SEM image in Figure 6a showed that smaller NPs agglomerated, resulting in the formation of bigger-sized AgNPs agglomerates with a roughly spherical shape. This phenomenon was caused by an imbalance between attractive Van der Waals forces and repulsive electrostatic forces (Zhao et al., 2016). The short incubation time for PF-AgNPs did not provide

adequate time for phytochemicals to stabilise the NPs' surface effectively. Therefore, these agglomerations occurred (Giorgi et al., 2019). Such characteristics suggested that the formation of polydisperse AgNPs is due to variations in the phytochemical content of the *P. foetida* aqueous extract, which acts as both a reducing and stabilising agent.

According to Figure 6b, it was observed at 24-hour incubation and exhibited a more compact and uniform morphology, with relatively well-dispersed particles, although minor agglomeration was still observed. These results were consistent with the earlier established studies that AgNPs synthesised from *Crataegus pentagyna* at 1 µm magnification revealed improved stability and definition with longer incubation times (Ebrahimzadeh et al., 2020). The study concluded that extended incubation significantly enhances the formation of well-defined and stable AgNPs, owing to the increasingly effective stabilising action of phytochemicals over time. Further validated the presence of elemental silver, which confirmed the successful synthesis of PF-AgNPs using EDX analysis (Table 6),

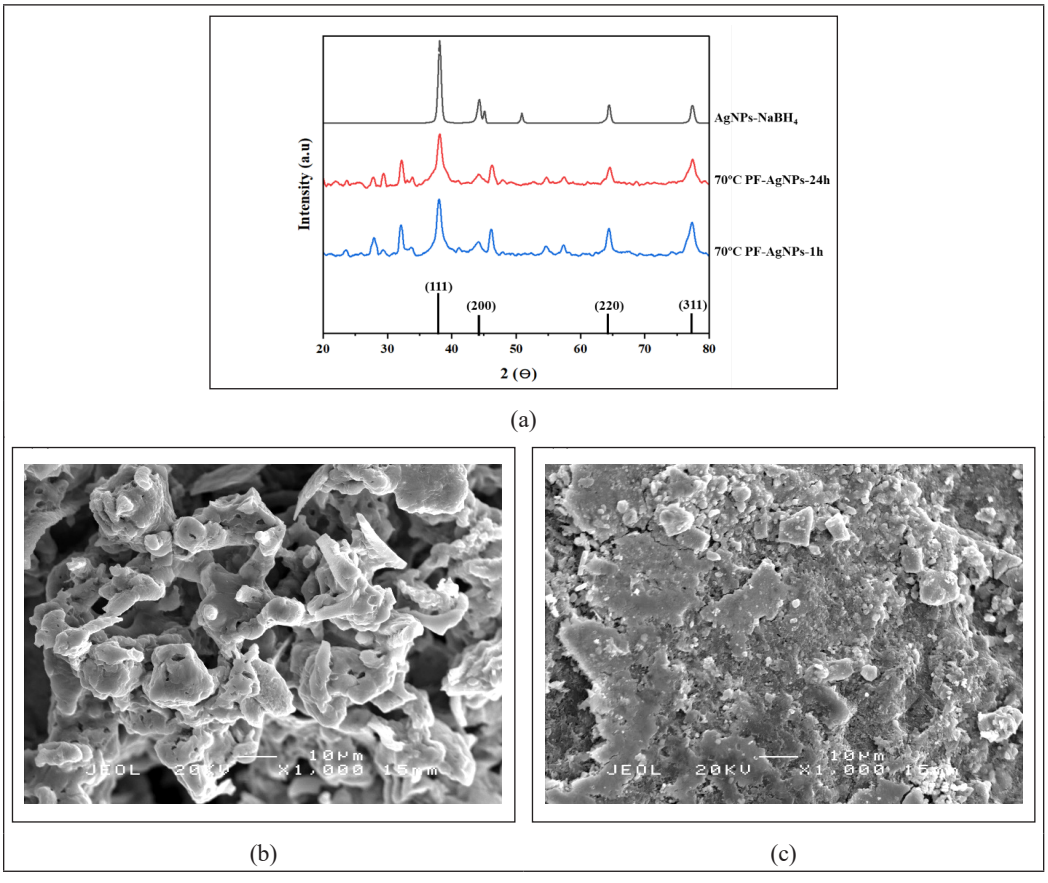


Figure 6. (a) XRD characterisation of PF-AgNPs; (b) SEM image after 1 h of incubation (10 µm scale, ×1000 magnification); (c) SEM image after 24 h of incubation (10 µm scale, ×1000 magnification) under optimal conditions

Table 6
Elemental composition using SEM-EDX at 1 and 24 h incubation time

Element	Mass (%) at 1 h	Mass (%) at 24 h	Atom (%)
Carbon (C)	18.33	8.75	28.40
Oxygen (O)	26.31	14.80	36.05
Sodium (Na)	4.91	1.61	2.72
Silicon (Si)	0.39	0.47	0.65
Chlorine (Cl)	2.89	4.41	4.84
Potassium (K)	1.60	0.85	0.85
Calcium (Ca)	0.53	2.13	2.07
Silver (Ag)	44.91	66.39	23.99

which revealed that, especially in this area, PF-AgNPs incubated for 24-hours contained a higher silver content (66.39%) compared to the 1-hour sample (44.91%). This result matched the result of the UV-Vis intensity, where 24-hour incubation has a higher intensity than 1-hour incubation time. Apart from this, the EDX analysis also detected elements such as Ca, K, Fe, and Si, potentially originating from residual plant extract components of *P. foetida*, which serve as capping and stabilising agents for the NPs' surfaces.

Proposed Mechanism of PF-AgNPs Synthesis

According to Ajitha et al. (2018), in the mechanism of the biosynthesised AgNPs formation, the Ag^{+} ions are reduced by antioxidant phytochemicals to Ag^{0} atoms, which then proceed through nucleation to create small clusters. The clusters then develop into petal-like structures and finally form flower-like superstructures with the assistance of bio-organic capping agents. The oxygen-containing functional groups (-OH and -COOH) participate in electrostatic attraction and complexation, enabling the particle to be stabilised. Phytochemicals not only function as reducing agents, but they also hinder further aggregation by capping the NPs through electron-rich hydroxyl groups. The study of Density of States (DOS) has also revealed that the interaction between Ag atoms and phenolic compounds is weakly adsorbed by the Van der Waals force or hydrogen bonds (Syafiuddin et al., 2017; Wang et al., 2021). This adsorption is reversible, does not change the basic structure of AgNPs and retains their stability and inherent properties such as electrical conductivity and catalytic behaviour (Brahmkhatri et al., 2015). The proposed reaction mechanism is summarised in Figure 7. This mechanism is supported by the UV-Vis spectra that displayed a characteristic SPR band at 420 nm and altered in colour from yellowish to brownish-red using optimum conditions. This reduction is pH-dependent to a certain extent, with basic conditions (pH 10) favouring the deprotonation of hydroxyl groups and enhancing reducing power capabilities. Further support is provided by the XRD patterns, which indicate that the NPs have an fcc crystalline structure, and FTIR spectra indicating the interaction of functional groups between the NPs and their capping agent.

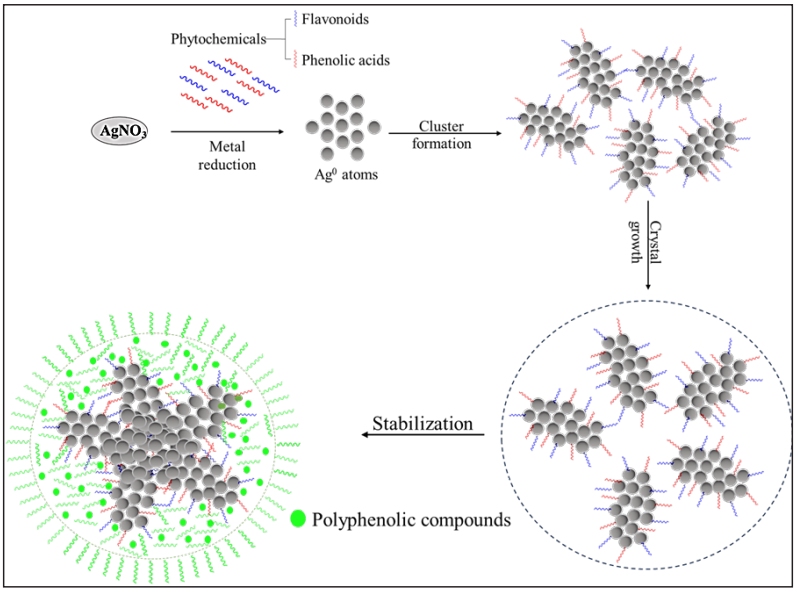


Figure 7. Mechanism of PF-AgNPs synthesis capped with *P. foetida* leaf extract, inspired by Ajitha et al. (2018)

Antioxidant Mechanism of PF-AgNPs

The antioxidant activity of PF-AgNPs is contributed by the combined actions of both the capping agents, such as phytochemicals, and the AgNPs (Figure 8a) (Javed et al., 2020). The antioxidative reactions of the stable purple DPPH radicals in an ideal model system result in their reduction to a pale, yellow-coloured DPPH–H through hydrogen atom transfer (HAT) and single electron transfer (SET) mechanisms, indicating DPPH radical scavenging capability. (Fouegue et al., 2018; Rohman et al., 2023; Wang et al., 2023). Polyphenolic groups in PF-AgNPs may donate hydrogen atoms or electrons to the DPPH radicals. These bioactive compounds are conjugated on the surface of AgNPs, especially those containing many hydroxyl groups, such as flavonoids. The AgNPs core also contributes to the reduction activity by: (1) increasing the amount of surface area available for interaction with free radicals; (2) facilitating electron transfer due to its metallic properties, and 3) stabilising phenoxy radicals generated by hydrogen donation, which prevents chain extension (Naseri et al., 2025; Park et al., 2016; Sytu & Camacho, 2018). In this assay, the control solution was dark purple, opposed to yellowish purple for PF-AgNPs with an inhibition of 35.64% (higher than that obtained in crude extract, i.e., 17.16%) under optimised conditions.

In NO radical scavenging assay, sodium nitroprusside (SNP) in aqueous solution at pH 7.4 undergoes spontaneous decomposition to generate NO• radicals that either interact with oxygen to produce nitrite ions (NO_2^-). Neutralisation of these reactive nitrogen species by PF-AgNPs occurs due to radical scavenging or metal chelation, as depicted in Figure 8b, adopted from Can et al. (2022) with a little modification. Antioxidant compounds

(ArOH) from phytochemical sources work as donors of hydrogen atoms or electrons for NO• radicals (Ermiş et al., 2025; Fraisse et al., 2018). In the hydrogen atom transfer mechanism, ArOH donates a hydrogen atom (H•) to neutralise the radical species, while ArOH itself is oxidised to a resonance-stabilised phenoxy radical. Furthermore, the antioxidants may operate through a SET mechanism to NO•, where an electron is donated to HNO, giving NO^- subsequently protonated to HNO. The formed ArO• radicals are stable, lack the initiation mechanism for a chain reaction and result in a decrease of free NO• radicals. This reduction in NO• level is witnessed with the decreased generation of nitrite in Griess reagent assay, and this indicates that PF-AgNPs have antioxidant scavenging ability (Can et al., 2022). In the experiment, PF-AgNPs exhibited the NO inhibitory activity at 38.70%.

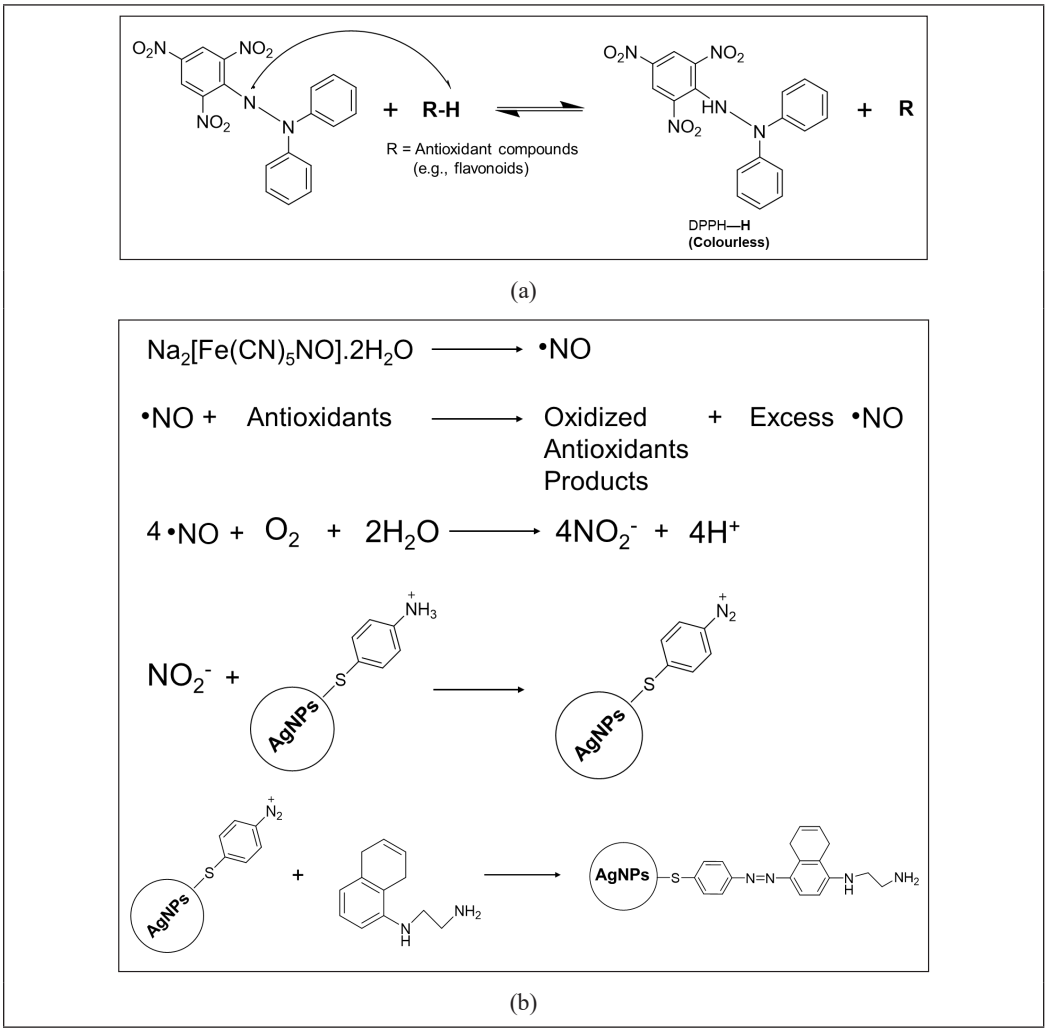


Figure 8. Proposed mechanisms of (a) DPPH; and (b) NO radical scavenging by antioxidants in PF-AgNPs

CONCLUSION

In this study, the green synthesis of PF-AgNPs was successfully optimised through Response Surface Methodology (RSM), identifying pH as the most influential factor in AgNPs formation. The optimal biosynthesis condition, 0.2% (w/v) extract at pH 10 and 70°C incubated for 24 hours, produced PF-AgNPs with significantly enhanced antioxidant activity compared to other formulations. The chemical profiling of the extract using ^{1}H NMR confirmed the presence of bioactive compounds, including flavonoid glycosides and other phytoconstituents, which are considered to play a role as natural reducing and stabilising agents during nanoparticle formation. Several other techniques, UV-Vis spectroscopy, FTIR, XRD, and SEM, have supported the synthesised nanoparticles to confirm that functional groups of phytochemical extracts helped the formation and stabilisation of nanoparticles and supported the synthesised nanoparticles' face-centred cubic (fcc) crystalline structure of silver, while SEM images revealed relatively uniform particles. Although polydispersity and lack of fit were observed in the NO-scavenger model, PF-AgNPs still exhibited enhanced antioxidant activity compared to the crude extract. Moreover, the therapeutic properties of PF-AgNPs could be evaluated for other biological activities, such as antimicrobial and anticancer effects. The integration of RSM with comprehensive physicochemical characterisation establishes a solid foundation for the development of eco-friendly AgNPs from natural resources. However, future studies, including in vivo evaluation and long-term stability assessment, should ideally incorporate particle size and zeta potential measurements using DLS to monitor nanoparticle homogeneity and colloidal stability, thereby supporting the sustainable production of low-cost PF-AgNPs with potential applications in biomedical, cosmetic, and environmental fields.

ACKNOWLEDGEMENT

We would like to thank the Ministry of Higher Education (MOHE) Malaysia for the financial support through the Fundamental Research Grant Scheme (FRGS/1/2022/STG02/UMT/02/1).

LIST OF ABBREVIATIONS

PF-AgNPs	:	*Passiflora foetida*-mediated silver nanoparticles
UV-Vis	:	Ultraviolet-visible spectroscopy
FTIR	:	Fourier transform infrared spectroscopy
XRD	:	X-ray diffraction
SEM	:	Scanning electron microscopy
EDX	:	Energy dispersive X-ray
SPR	:	Surface plasmon resonance

RSM	:	Response surface methodology
NO	:	Nitric oxide
DPPH	:	2,2-Diphenyl-1-picrylhydrazyl
^{1}H NMR	:	Proton nuclear magnetic resonance
ATR	:	Attenuated total reflectance
NP	:	Nanoparticle
RMSE	:	Root means square error
ANOVA	:	Analysis of variance
BBD	:	Box-Behnken design

REFERENCES

Abalkhil, T. A., Alharbi, S. A., Salmen, S. H., & Wainwright, M. (2017). Bactericidal activity of biosynthesised silver nanoparticles against human pathogenic bacteria. *Biotechnology & Biotechnological Equipment, 31*(2), 411-417. https://doi.org/10.1080/13102818.2016.1267594

Abdul-Hamid, N. A., Abas, F., Ismail, I. S., Tham, C. L., Maulidiani, M., Mediani, A., Swarup, S., & Umashankar, S. (2019). ^{1}H-NMR-based metabolomics to investigate the effects of *Phoenix dactylifera* seed extracts in LPS-IFN-γ-induced RAW 264.7 cells. *Food Research International, 125*, 108565.https://doi.org/10.1016/j.foodres.2019.108565

Abu-Elghait, M., Soliman, M. K. Y., Azab, M. S., & Salem, S. S. (2025). Response surface methodology: Optimisation of myco-synthesised gold and silver nanoparticles by *Trichoderma saturnisporum*. *Biomass Conversion and Biorefinery, 15*(3), 4211-4224. https://doi.org/10.1007/s13399-023-05188-4

Ajitha, B., Reddy, Y. A. K., Jeon, H. J., & Ahn, C. W. (2018). Synthesis of silver nanoparticles in an eco-friendly way using *Phyllanthus amarus* leaf extract: Antimicrobial and catalytic activity. *Advanced Powder Technology, 29*(1), 86-93. https://doi.org/10.1016/j.apt.2017.10.015

Akash, M. S. H., & Rehman, K. (2020). Ultraviolet-visible (UV-Vis) spectroscopy. In M. S. H. Akash & K. Rehman (Eds.), *Essentials of pharmaceutical analysis* (pp. 29-56). Springer Nature Singapore.https://doi.org/10.1007/978-981-15-1547-7_3

Alex, S. B., S., M. S., & L., M. S. (2025). Evaluation of antimicrobial, anti-inflammatory and cytotoxic effects of silver nanoparticles synthesised from *Cynodon dactylon*. *Natural Product Research, 39*(4), 779-786. https://doi.org/10.1080/14786419.2023.2290154

Al-Gebory, L., & Mengüç, M. P. (2018). The effect of pH on particle agglomeration and optical properties of nanoparticle suspensions. *Journal of Quantitative Spectroscopy and Radiative Transfer, 219*, 46-60. https://doi.org/10.1016/j.jqsrt.2018.07.020

Aliero, A. S., Zawawi, N. A., Malek, N. A. N. N., Usman, B. J., Asraf, M. H., Matmin, J., & Isah, M. (2024). Harnessing palm oil mill effluent for the green synthesis of silver nanoparticles: Optimisation via response surface methodology and assessment of antibacterial activity. *Journal of Inorganic and Organometallic Polymers and Materials*, 1-17. https://doi.org/10.1007/s10904-024-03478-6

Alkhulaifi, M. M., Alshehri, J. H., Alwehaibi, M. A., Awad, M. A., Al-Enazi, N. M., Aldosari, N. S., Hatamleh, A. A., & Abdel-Raouf, N. (2020). Green synthesis of silver nanoparticles using *Citrus limon* peels and

evaluation of their antibacterial and cytotoxic properties. *Saudi Journal of Biological Sciences, 27*(12), 3434-3441. https://doi.org/10.1016/j.sjbs.2020.09.031

Anguraj, A., Michael, H. S. R., Sugumaran, S., Madhusudhanan, G. R., & Sivaraman, R. K. (2024). A comparative study on biosynthesised silver nanoparticles from *Hylocereus undatus* fruit peel and their therapeutic applications. *Discover Nano, 19*(1), Article 49. https://doi.org/10.1186/s11671-024-03995-w

Arokiyaraj, S., Valan Arasu, M., Vincent, S., Oh, Y.-K., Kim, K. H., Choi, K.-C., Choi, S. H., & Prakash, N. U. (2014). Rapid green synthesis of silver nanoparticles from *Chrysanthemum indicum* L. and its antibacterial and cytotoxic effects: An in vitro study. *International Journal of Nanomedicine, 9*, 379-388. https://doi.org/10.2147/IJN.S53546

Barkat, M. A., Harshita, Beg, S., Naim, M. J., Pottoo, F. H., Singh, S. P., & Ahmad, F. J. (2017). Current progress in synthesis, characterisation and applications of silver nanoparticles: Precepts and prospects. *Recent Patents on Anti-Infective Drug Discovery, 13*(1), 53-69. https://doi.org/10.2174/1574891X12666171006102833

Behravan, M., Hossein Panahi, A., Naghizadeh, A., Ziaee, M., Mahdavi, R., & Mirzapour, A. (2019). Facile green synthesis of silver nanoparticles using *Berberis vulgaris* leaf and root aqueous extract and its antibacterial activity. *International Journal of Biological Macromolecules, 124*, 148-154.https://doi.org/10.1016/j.ijbiomac.2018.11.101

Bharadwaj, K. K., Rabha, B., Pati, S., Choudhury, B. K., Sarkar, T., Gogoi, S. K., Kakati, N., Baishya, D., Kari, Z. A., & Edinur, H. A. (2021). Green synthesis of silver nanoparticles using *Diospyros malabarica* fruit extract and assessment of their antimicrobial, anticancer and catalytic reduction of 4-nitrophenol (4-NP). *Nanomaterials, 11*(8), Article 1999.https://doi.org/10.3390/nano11081999

Brahmkhatri, V. P., Chandra, K., Dubey, A., & Atreya, H. S. (2015). An ultrastable conjugate of silver nanoparticles and protein formed through weak interactions. *Nanoscale, 7*(30), 12921-12931.https://doi.org/10.1039/C5NR03047A

Can, Z., Keskin, B., Üzer, A., & Apak, R. (2022). Detection of nitric oxide radical and determination of its scavenging activity by antioxidants using spectrophotometric and spectrofluorometric methods. *Talanta, 238*, 122993. https://doi.org/10.1016/j.talanta.2021.122993

Cativa, N. M., Dell'Erba, I. E., Waiman, C. V., Arenas, G. F., Ceolín, M., Giovanetti, L. J., Ramallo-López, J. M., Eliçabe, G., & Hoppe, C. E. (2020). Tuning the photothermal effect of carboxylated-coated silver nanoparticles through pH-induced reversible aggregation. *Langmuir, 36*(46), 13998-14008. https://doi.org/10.1021/acs.langmuir.0c02528

Corciovă, A., Mircea, C., Burlec, A. F., Fifere, A., Moleavin, I. T., Sarghi, A., Tuchiluş, C., Ivănescu, B., & Macovei, I. (2022). Green synthesis and characterisation of silver nanoparticles using a *Lythrum salicaria* extract and in vitro exploration of their biological activities. *Life, 12*(10), Article 1643.https://doi.org/10.3390/life12101643

Crisan, M. C., Teodora, M., & Lucian, M. (2021). Copper nanoparticles: Synthesis and characterisation, physiology, toxicity and antimicrobial applications. *Applied Sciences, 12*(1), Article 141. https://doi.org/10.3390/app12010141

De Silva, C., Nawawi, N. M., Abd Karim, M. M., Abd Gani, S., Masarudin, M. J., Gunasekaran, B., &

Ahmad, S. A. (2021). The mechanistic action of biosynthesised silver nanoparticles and its application in aquaculture and livestock industries. *Animals, 11*(7), Article 2097.https://doi.org/10.3390/ani11072097

Desai, K. M., Survase, S. A., Saudagar, P. S., Lele, S. S., & Singhal, R. S. (2008). Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimisation: Case study of fermentative production of scleroglucan. *Biochemical Engineering Journal, 41*(3), 266-273. https://doi.org/10.1016/j.bej.2008.05.009

Díaz-Velasco, S., Delgado, J., Peña, F. J., & Estévez, M. (2022). Protein oxidation marker, α-amino adipic acid, impairs proteome of differentiated human enterocytes: Underlying toxicological mechanisms. *Biochimica et Biophysica Acta-Proteins and Proteomics, 1870*(7), Article 140797.https://doi.org/10.1016/j.bbapap.2022.140797

Dobrucka, R. (2017). Synthesis of titanium dioxide nanoparticles using *Echinacea purpurea* herba. *Iranian Journal of Pharmaceutical Research, 16*(2), 756-762. https://pmc.ncbi.nlm.nih.gov/articles/PMC5603885/

Ebrahimzadeh, Z., Salehzadeh, A., Naeemi, A. S., & Jalali, A. (2020). Silver nanoparticles biosynthesised by *Anabaena flos-aquae* enhance apoptosis in a breast cancer cell line. *Bulletin of Materials Science, 43*(1), Article 92. https://doi.org/10.1007/s12034-020-2064-1

Egodawaththa, N. M., Knight, A. L., Ma, J., Knight, D. A., Guisbert, E., & Nesnas, N. (2022). Synthesis and characterisation of ligand-stabilised silver nanoparticles and comparative antibacterial activity against *Escherichia coli*. *International Journal of Molecular Sciences, 23*(23), Article 15251.https://doi.org/10.3390/ijms232315251

Elangovan, D., Rahman, H. B. H., Dhandapani, R., Palanivel, V., Thangavelu, S., Paramasivam, R., & Muthupandian, S. (2022). Coating of wallpaper with green synthesised silver nanoparticles from *Passiflora foetida* fruit and its illustrated antifungal mechanism. *Process Biochemistry, 112*, 177-182. https://doi.org/10.1016/j.procbio.2021.11.027

Ermiş, S., Türk, Ş., Koç, Ö. K., Üzer, A., & Apak, R. (2025). Turn-on fluorometric assay based on functionalised gold nanoparticles for measurement of nitric oxide radical scavenging activity of phenol-, amine-, and thiol-type antioxidants. *Food Chemistry*, Article 146108. https://doi.org/10.1016/j.foodchem.2025.146108

Fernando, I., Tay, Y. Y., Karunasekera, H., & Zhou, Y. (2020). Observation of the interactions of silver nanoparticles (AgNPs) mediated by acid in aquatic matrices using in situ liquid cell transmission electron microscopy. *Analytica Chimica Acta, 1104*, 47-52. https://doi.org/10.1016/j.aca.2019.12.072

Fernando, I., & Zhou, Y. (2019). Impact of pH on the stability, dissolution and aggregation kinetics of silver nanoparticles. *Chemosphere, 216*, 297-305. https://doi.org/10.1016/j.chemosphere.2018.10.122

Fouegue, A. D. T., Mama, D. B., Ghogomu, J. N., Elie, Y., & Etoh, M.-A. (2018). The substitution effect on reaction enthalpies of antioxidant mechanisms of juglone and its derivatives in gas and solution phase: DFT study. *Journal of Chemistry, 2018*(1), Article 1958047. https://doi.org/10.1155/2018/1958047

Fraisse, D., Degerine-Roussel, A., Bred, A., Ndoye, S. F., Vivier, M., Felgines, C., & Senejoux, F. (2018). A novel HPLC method for direct detection of nitric oxide scavengers from complex plant matrices and its application to *Aloysia triphylla* leaves. *Molecules, 23*(7), Article 1574.https://doi.org/10.3390/molecules23071574

Gherasim, O., Puiu, R. A., Bîrca, A. C., Burduşel, A. C., & Grumezescu, A. M. (2020). An updated review

on silver nanoparticles in biomedicine. *Nanomaterials, 10*(11), Article 2318. https://doi.org/10.3390/nano10112318

Giorgi, F., Coglitore, D., Curran, J. M., Gilliland, D., Macko, P., Whelan, M., Worth, A., & Patterson, E. A. (2019). The influence of inter-particle forces on diffusion at the nanoscale. *Scientific Reports, 9*(1), Article 12325. https://doi.org/10.1038/s41598-019-48754-5

Gulyaev, V. P., Petrov, P. P., & Stepanova, K. V. (2022). Experimental validation of safety coefficient values of 09G2S steel by X-ray diffraction method. *Procedia Structural Integrity, 40*, 180-184.https://doi.org/10.1016/j.prostr.2022.04.024

Jain, S., & Mehata, M. S. (2017). Medicinal plant leaf extract and pure flavonoid mediated green synthesis of silver nanoparticles and their enhanced antibacterial property. *Scientific Reports, 7*(1), Article 15867. https://doi.org/10.1038/s41598-017-15724-8

Jaison, J. P., Balasubramanian, B., Gangwar, J., James, N., Pappuswamy, M., Anand, A. V., Al-Dhabi, N. A., Valan Arasu, M., Liu, W.-C., & Sebastian, J. K. (2023). Green synthesis of bioinspired nanoparticles mediated from plant extracts of Asteraceae family for potential biological applications. *Antibiotics, 12*(3), Article 543. https://doi.org/10.3390/antibiotics12030543

Javed, R., Zia, M., Naz, S., Aisida, S. O., Ain, N. ul, & Ao, Q. (2020). Role of capping agents in the application of nanoparticles in biomedicine and environmental remediation: Recent trends and future prospects. *Journal of Nanobiotechnology, 18*(1), Article 172. https://doi.org/10.1186/s12951-020-00704-4

Jumabaev, A., Khudaykulov, B., Holikulov, U., Norkulov, A., Subbiah, J., Al-Dossary, O. M., Hushvaktov, H., Absanov, A., & Issaoui, N. (2025). Molecular structure, vibrational spectral assignments, MEP, HOMO–LUMO, AIM, NCI, RDG, ELF and LOL properties of acetophenone and its solutions based on DFT calculations. *Optical Materials, 159*, Article 116683. https://doi.org/10.1016/j.optmat.2025.116683

Karan, T., Gonulalan, Z., Erenler, R., Kolemen, U., & Eminagaoglu, O. (2024). Green synthesis of silver nanoparticles using *Sambucus ebulus* leaves extract: Characterisation, quantitative analysis of bioactive molecules, antioxidant and antibacterial activities. *Journal of Molecular Structure, 1296*, Article 136836. https://doi.org/10.1016/j.molstruc.2023.136836

Kaushal, A., Khurana, I., Yadav, P., Allawadhi, P., Banothu, A. K., Neeradi, D., Thalugula, S., Barani, P. J., Naik, R. R., Navik, U., Bharani, K. K., & Khurana, A. (2023). Advances in therapeutic applications of silver nanoparticles. *Chemico-Biological Interactions, 382*, Article 110590.https://doi.org/10.1016/j.cbi.2023.110590

Khan, A., Younis, T., Anas, M., Ali, M., Shinwari, Z. K., Khalil, A. T., Munawar, K. S., Mohamed, H. E. A., Hkiri, K., & Maaza, M. (2025). *Withania coagulans*-mediated green synthesis of silver nanoparticles: Characterisation and assessment of their phytochemical, antioxidant, toxicity and antimicrobial activities. *BMC Plant Biology, 25*(1), Article 574. https://doi.org/10.1186/s12870-025-06533-7

Khanal, L. N., Dhakal, P. P., Kandel, M. R., Acharya, D., Baral, E. R., Chhetri, K., & Kalauni, S. K. (2023). Stem bark-mediated green synthesis of silver nanoparticles from *Pyrus pashia*: Characterisation, antioxidant and antibacterial properties. *Inorganics, 11*(6), Article 263. https://doi.org/10.3390/inorganics11060263

Khatun, H., Alam, S., Aziz, M. A., Karim, M. R., Rahman, M. H., Rabbi, M. A., & Habib, M. R. (2024).

Plant-assisted green preparation of silver nanoparticles using leaf extract of *Dalbergia sissoo* and their antioxidant, antibacterial and catalytic applications. *Bioprocess and Biosystems Engineering, 47*(8), 1347-1362. https://doi.org/10.1007/s00449-024-03029-w

Krishnamoorthi, R., Mahalingam, P. U., & Malaikozhundan, B. (2022). Edible mushroom extract engineered AgNPs as safe antimicrobial and antioxidant agents with no significant cytotoxicity on human dermal fibroblast cells. *Inorganic Chemistry Communications, 139*, Article 109362. https://doi.org/10.1016/j.inoche.2022.109362

Krupa, A. N. D., Abigail, M. E. A., Santhosh, C., Grace, A. N., & Vimala, R. (2016). Optimisation of process parameters for the microbial synthesis of silver nanoparticles using 3-level Box–Behnken design. *Ecological Engineering, 87*, 168-174. https://doi.org/10.1016/j.ecoleng.2015.11.030

Kumari, R., Saini, A. K., Kumar, A., & Saini, R. V. (2020). Apoptosis induction in lung and prostate cancer cells through silver nanoparticles synthesised from *Pinus roxburghii* bioactive fraction. *JBIC Journal of Biological Inorganic Chemistry, 25*(1), 23-37. https://doi.org/10.1007/s00775-019-01729-3

Lau, C. P., Abdul-Wahab, M. F., Jaafar, J., Chan, G. F., & Rashid, N. A. A. (2016). Effect of pH and biological media on polyvinylpyrrolidone-capped silver nanoparticles. *AIP Conference Proceedings, 1756*(1), Article 080002. https://doi.org/10.1063/1.4958781

Liu, H., Zhang, H., Wang, J., & Wei, J. (2020). Effect of temperature on the size of biosynthesised silver nanoparticles: Deep insight into microscopic kinetics analysis. *Arabian Journal of Chemistry, 13*(1), 1011-1019. https://doi.org/10.1016/j.arabjc.2017.09.004

Mardiyanto, M., Apriani, E. F., & Heylken, F. P. (2023). The role of temperature and pH in the synthesis of silver nanoparticles using *Areca catechu* L. seed extract as bioreductor. *Farmacia, 71*(2), 244-253.https://doi.org/10.31925/farmacia.2023.2.3

Maulidiani, M., Mediani, A., Abas, F., Park, Y. S., Park, Y.-K., Kim, Y. M., & Gorinstein, S. (2018). ^{1}H NMR and antioxidant profiles of polar and non-polar extracts of persimmon (*Diospyros kaki* L.): Metabolomics study based on cultivars and origins. *Talanta, 184*, 277-286. https://doi.org/10.1016/j.talanta.2018.02.084

Miranda, A., Akpobolokemi, T., Chung, E., Ren, G., & Raimi-Abraham, B. T. (2022). pH alteration in plant-mediated green synthesis and its resultant impact on antimicrobial properties of silver nanoparticles (AgNPs). *Antibiotics, 11*(11), Article 1592. https://doi.org/10.3390/antibiotics11111592

Mohamad Ali, B., Velavan, B., Sudhandiran, G., Sridevi, J., & Sultan Nasar, A. (2020). Radical dendrimers: Synthesis, anti-tumour activity and enhanced cytoprotective performance of TEMPO free radical functionalised polyurethane dendrimers. *European Polymer Journal, 122*, Article 109354.https://doi.org/10.1016/j.eurpolymj.2019.109354

Mouriya, G. K., Mohammed, M., Azmi, A. A., Khairul, W. M., Karunakaran, T., Amirul, A. A. A., Ramakrishna, S., Santhanam, R., & Vigneswari, S. (2023). Green synthesis of *Cicer arietinum* waste-derived silver nanoparticles for antimicrobial and cytotoxicity properties. *Biocatalysis and Agricultural Biotechnology, 47*, Article 102573. https://doi.org/10.1016/j.bcab.2022.102573

Mousavi, B., Tafvizi, F., & Zaker Bostanabad, S. (2018). Green synthesis of silver nanoparticles using *Artemisia turcomanica* leaf extract and the study of anti-cancer effect and apoptosis induction on gastric cancer cell line (AGS). *Artificial Cells, Nanomedicine, and Biotechnology, 46*(sup1), 499-510.https://doi.org/1

0.1080/21691401.2018.1430697

Naganthran, A., Verasoundarapandian, G., Khalid, F. E., Masarudin, M. J., Zulkharnain, A., Nawawi, N. M., Karim, M., Che Abdullah, C. A., & Ahmad, S. A. (2022). Synthesis, characterisation and biomedical application of silver nanoparticles. *Materials, 15*(2), Article 427.https://doi.org/10.3390/ma15020427

Naseri, M., Irannejad, M., Mehdilo, A., & Neisiani, R. K. (2025). Green synthesis of silver nanoparticles: A review of the methods and antioxidant properties of nanoparticles. *Amirkabir Journal of Civil Engineering, 57*(6), 939-972. https://doi.org/10.22060/ceej.2025.23714.8202

Ndikau, M., Noah, N. M., Andala, D. M., & Masika, E. (2017). Green synthesis and characterisation of silver nanoparticles using *Citrullus lanatus* fruit rind extract. *International Journal of Analytical Chemistry, 2017*, Article 8108504. https://doi.org/10.1155/2017/8108504

Ndulini, S. F., Mthembu, M. S., & Pullabhotla, V. S. R. (2024). *Ochrobactrum* bacterial-mediated pH-dependent synthesis of silver oxide nanoparticles. *BioNanoScience, 14*(3), 2092-2102. https://doi.org/10.1007/s12668-024-01470-5

Nguyen, T. Y., To, D. C., Tran, M. H., Lee, J. S., Lee, J. H., Kim, J. A., Woo, M. H., & Min, B. S. (2015). Anti-inflammatory flavonoids isolated from *Passiflora foetida*. *Natural Product Communications, 10*(6), Article 634. https://doi.org/10.1177/1934578X1501000634

Palanisamy, D. S., Gounder, B. S., Selvaraj, K., Kandhasamy, S., Alqahtani, T., Alqahtani, A., Chidambaram, K., Arunachalam, K., Alkahtani, A. M., Chandramoorthy, H. C., Sharma, N., Rajeshkumar, S., & Marwaha, L. (2024). Synergistic antibacterial and mosquitocidal effect of *Passiflora foetida*-synthesised silver nanoparticles. *Brazilian Journal of Biology, 84*, Article e263391. https://doi.org/10.1590/1519-6984.263391

Park, J., Cha, S.-H., Cho, S., & Park, Y. (2016). Green synthesis of gold and silver nanoparticles using gallic acid: Catalytic activity and conversion yield toward the 4-nitrophenol reduction reaction. *Journal of Nanoparticle Research, 18*(6), Article 166. https://doi.org/10.1007/s11051-016-3466-2

Patil, A., Lade, B., & Paikrao, H. (2015). A scientific update on *Passiflora foetida*. *European Journal of Medicinal Plants, 5*(2), 145-155.https://doi.org/10.9734/EJMP/2015/12015

Patil, N., Yadav, P., & Gogate, P. R. (2024). Ultrasound-assisted intensified enzymatic extraction of total phenolic compounds from pomegranate peels. *Separation and Purification Technology, 350*, Article 127967. https://doi.org/10.1016/j.seppur.2024.127967

Ponsuwan, K., Nathabumroong, S., Lekphrom, R., Sorin, S., Saengboonmee, C., Senawong, T., Tontapha, S., & Schevenels, F. T. (2024). Passifetilactones A–E, fatty acid lactones from the fruit and flowers of *Passiflora foetida* with cytotoxic activity. *Journal of Natural Products, 87*(6), 1652-1659.https://doi.org/10.1021/acs.jnatprod.4c00463

Pourmortazavi, S. M., Taghdiri, M., Makari, V., & Rahimi-Nasrabadi, M. (2015). Procedure optimisation for green synthesis of silver nanoparticles by aqueous extract of *Eucalyptus oleosa*. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 136*, 1249-1254. https://doi.org/10.1016/j.saa.2014.10.010

Rajasekar, P., Palanisamy, S., Anjali, R., Vinosha, M., Thillaieswari, M., Malaikozhundan, B., Boomi, P., Saravanan, M., You, S. G., & Prabhu, N. M. (2019). *Cladophora fascicularis*-mediated silver nanoparticles: Assessment of their antibacterial activity against *Aeromonas hydrophila*. *Journal of Cluster Science,*

31(4), 673-683. https://doi.org/10.1007/s10876-019-01674-w

Riaz, M., Altaf, M., Ahmad, P., Khandaker, M. U., Osman, H., Eed, E. M., & Shakir, Y. (2022). Biogenic synthesis of Ag nanoparticles of 18.27 nm by *Zanthoxylum armatum* and determination of biological potentials. *Molecules, 27*(4), Article 1166. https://doi.org/10.3390/molecules27041166

Rohman, R., Nath, R., & Kar, R. (2023). Revisiting hydrogen atom transfer reactions through a simple and accurate theoretical model: Role of hydrogen bond energy in polyphenolic antioxidants. *Computational and Theoretical Chemistry, 1223*, Article 114097. https://doi.org/10.1016/j.comptc.2023.114097

Sati, A., Ranade, T. N., Mali, S. N., Ahmad Yasin, H. K., & Pratap, A. (2025). Silver nanoparticles (AgNPs): Comprehensive insights into bio/synthesis, key influencing factors, multifaceted applications and toxicity: A 2024 update. *ACS Omega, 10*(8), 7549-7582. https://doi.org/10.1021/acsomega.4c11045

Seifipour, R., Nozari, M., & Pishkar, L. (2020). Green synthesis of silver nanoparticles using *Tragopogon collinus* leaf extract and study of their antibacterial effects. *Journal of Inorganic and Organometallic Polymers and Materials, 30*(8), 2926-2936. https://doi.org/10.1007/s10904-020-01441-9

Shata, A. M., Rejili, M. S., El-Naggar, M. M., El-Asser, S. A., Saleh, A. A., Hafez, E. E., Haouala, F., & Youssef, A. S. (2025). Exploring the mechanisms and mode of action of bioactive compounds from marine *Streptomyces albidoflavus* against breast carcinoma cells. *BMC Biotechnology, 25*(1), Article 48.https://doi.org/10.1186/s12896-025-00991-5

Singh, J., Dutta, T., Kim, K.-H., Rawat, M., Samddar, P., & Kumar, P. (2018). Green synthesis of metals and their oxide nanoparticles: Applications for environmental remediation. *Journal of Nanobiotechnology, 16*(1), Article 84. https://doi.org/10.1186/s12951-018-0408-4

Song, Y., Wei, X.-Q., Li, M.-Y., Duan, X.-W., Sun, Y.-M., Yang, R.-L., Su, X.-D., Huang, R.-M., & Wang, H. (2018). Nutritional composition and antioxidant properties of the fruits of a Chinese wild *Passiflora foetida*. *Molecules, 23*(2), Article 459. https://doi.org/10.3390/molecules23020459

Sreejayan, & Rao, M. N. A. (1997). Nitric oxide scavenging by curcuminoids. *Journal of Pharmacy and Pharmacology, 49*(1), 105-107. https://doi.org/10.1111/j.2042-7158.1997.tb06761.x

Syafiuddin, A., Salmiati, Salim, M. R., Beng Hong Kueh, A., Hadibarata, T., & Nur, H. (2017). A review of silver nanoparticles: Research trends, global consumption, synthesis, properties and future challenges. *Journal of the Chinese Chemical Society, 64*(7), 732-756. https://doi.org/10.1002/jccs.201700067

Sytu, M. R. C., & Camacho, D. H. (2018). Green synthesis of silver nanoparticles (AgNPs) from *Lenzites betulina* and the potential synergistic effect of AgNPs and capping biomolecules in enhancing antioxidant activity. *BioNanoScience, 8*(3), 835-844. https://doi.org/10.1007/s12668-018-0548-x

Temur, N., Gundes, N. S., Korkmaz, B., Ozkaya, Z. A., & Ocsoy, I. (2024). Silver nanoparticles in dentistry. In *Silver nanoparticles for drug delivery* (pp. 265-288). Elsevier.https://doi.org/10.1016/B978-0-443-15343-3.00008-5

Upadhyaya, P., & Devi, T. G. (2018). Raman band shape analysis and DFT study of CO and CH stretching modes of N,N-dibutyl formamide in DMSO solvent. *Vibrational Spectroscopy, 96*, 110-117. https://doi.org/10.1016/j.vibspec.2018.03.006

Van Linh, N., Trung Tuong, N., Xuan Phong, P., Trang, D. T., Nhiem, N. X., Hoai An, D., & Huu Tai,

B. (2022). New phenylethanoid and other compounds from *Passiflora foetida* L., with their nitric oxide inhibitory activities. *Natural Product Communications, 17*(11), Article 141163. https://doi.org/10.1177/1934578X221141163

Wang, L., Periyasami, G., Aldalbahi, A., & Fogliano, V. (2021). The antimicrobial activity of silver nanoparticles biocomposite films depends on silver ion release behaviour. *Food Chemistry, 359*, Article 129859.https://doi.org/10.1016/j.foodchem.2021.129859

Wang, M., Mo, F., Li, H., Li, Y., Zhang, S., Zhu, L., Li, Z., Xu, J., Deng, N., & Wang, K. (2021). Adsorption based on weak interaction between phenolic hydroxyl and carboxyl groups and silver nanoparticles in aqueous environment: Experimental and DFT-D3 exploration. *Journal of Environmental Chemical Engineering, 9*(6), Article 106816. https://doi.org/10.1016/j.jece.2021.106816

Wang, X., Sun, T., Zhu, H., Han, T., Wang, J., & Dai, H. (2020). Roles of pH, cation valence and ionic strength in the stability and aggregation behaviour of zinc oxide nanoparticles. *Journal of Environmental Management, 267*, Article 110656. https://doi.org/10.1016/j.jenvman.2020.110656

Wang, Y., Liu, C., Tang, S., Tian, J., Wang, Y., & Yang, Y. (2023). Thermodynamics, kinetics and structure-activity relationship of hydroxyanthraquinones scavenging free radicals. *Food Bioscience, 53*, Article 102705. https://doi.org/10.1016/j.fbio.2023.102705

Wasilewska, A., Klekotka, U., Zambrzycka, M., Zambrowski, G., Święcicka, I., & Kalska-Szostko, B. (2023). Physico-chemical properties and antimicrobial activity of silver nanoparticles fabricated by green synthesis. *Food Chemistry, 400*, Article 133960. https://doi.org/10.1016/j.foodchem.2022.133960

Xu, M., Liu, G., Li, M., Huo, M., Zong, W., & Liu, R. (2020). Probing the cell apoptosis pathway induced by perfluorooctanoic acid and perfluorooctane sulfonate at the subcellular and molecular levels. *Journal of Agricultural and Food Chemistry, 68*(2), 633-641. https://doi.org/10.1021/acs.jafc.9b07072

Zhao, R., Luo, Y., & Pendry, J. B. (2016). Transformation optics applied to van der Waals interactions. *Science Bulletin, 61*(1), 59-67.https://doi.org/10.1007/s11434-015-0958-x

APPENDIX

1. Results of canonical path analysis for the intensity response

canonical.path(model2, dist = seq(-2, 2, by = 0.1))

	dist	x1	x2	x3	Temp	Conc	pH	yhat
1	-2.0	-3.211	0.567	-1.503	-14.22	0.2567	2.491	2.388
2	-1.9	-3.155	0.580	-1.422	-13.10	0.2580	2.734	2.208
3	-1.8	-3.098	0.593	-1.340	-11.96	0.2593	2.980	2.035
4	-1.7	-3.041	0.606	-1.259	-10.82	0.2606	3.223	1.873
5	-1.6	-2.985	0.619	-1.177	-9.70	0.2619	3.469	1.719
6	-1.5	-2.928	0.631	-1.096	-8.56	0.2631	3.712	1.575
7	-1.4	-2.871	0.644	-1.015	-7.42	0.2644	3.955	1.441
8	-1.3	-2.815	0.657	-0.933	-6.30	0.2657	4.201	1.315
9	-1.2	-2.758	0.670	-0.852	-5.16	0.2670	4.444	1.199
10	-1.1	-2.702	0.683	-0.770	-4.04	0.2683	4.690	1.092
11	-1.0	-2.645	0.695	-0.689	-2.90	0.2695	4.933	0.995
12	-0.9	-2.588	0.708	-0.608	-1.76	0.2708	5.176	0.906
13	-0.8	-2.532	0.721	-0.526	-0.64	0.2721	5.422	0.827
14	-0.7	-2.475	0.734	-0.445	0.50	0.2734	5.665	0.758
15	-0.6	-2.418	0.747	-0.363	1.64	0.2747	5.911	0.697
16	-0.5	-2.362	0.760	-0.282	2.76	0.2760	6.154	0.646
17	-0.4	-2.305	0.772	-0.201	3.90	0.2772	6.397	0.604
18	-0.3	-2.248	0.785	-0.119	5.04	0.2785	6.643	0.572
19	-0.2	-2.192	0.798	-0.038	6.16	0.2798	6.886	0.549
20	-0.1	-2.135	0.811	0.044	7.30	0.2811	7.132	0.534
21	0.0	-2.079	0.824	0.125	8.42	0.2824	7.375	0.530
22	0.1	-2.022	0.836	0.207	9.56	0.2836	7.621	0.535
23	0.2	-1.965	0.849	0.288	10.70	0.2849	7.864	0.548
24	0.3	-1.909	0.862	0.369	11.82	0.2862	8.107	0.572
25	0.4	-1.852	0.875	0.451	12.96	0.2875	8.353	0.604
26	0.5	-1.795	0.888	0.532	14.10	0.2888	8.596	0.646
27	0.6	-1.739	0.900	0.614	15.22	0.2900	8.842	0.697
28	0.7	-1.682	0.913	0.695	16.36	0.2913	9.085	0.758
29	0.8	-1.625	0.926	0.776	17.50	0.2926	9.328	0.827
30	0.9	-1.569	0.939	0.858	18.62	0.2939	9.574	0.906
31	1.0	-1.512	0.952	0.939	19.76	0.2952	9.817	0.995
32	1.1	-1.455	0.964	1.021	20.90	0.2964	10.063	1.093
33	1.2	-1.399	0.977	1.102	22.02	0.2977	10.306	1.199
34	1.3	-1.342	0.990	1.183	23.16	0.2990	10.549	1.315
35	1.4	-1.286	1.003	1.265	24.28	0.3003	10.795	1.441
36	1.5	-1.229	1.016	1.346	25.42	0.3016	11.038	1.575
37	1.6	-1.172	1.029	1.428	26.56	0.3029	11.284	1.720
38	1.7	-1.116	1.041	1.509	27.68	0.3041	11.527	1.872
39	1.8	-1.059	1.054	1.591	28.82	0.3054	11.773	2.036
40	1.9	-1.002	1.067	1.672	29.96	0.3067	12.016	2.208
41	2.0	-0.946	1.080	1.753	31.08	0.3080	12.259	2.388

2. Results of canonical path analysis for the NO response

canonical.path(model2, dist = seq(-2, 2, by = 0.1))

	dist	x1	x2	x3	Temp	Conc	pH	yhat
1	-2.0	-1.129	1.827	-1.390	27.42	0.3827	2.830	40.454
2	-1.9	-1.105	1.775	-1.308	27.90	0.3775	3.076	38.733
3	-1.8	-1.082	1.724	-1.225	28.36	0.3724	3.325	37.099
4	-1.7	-1.058	1.672	-1.143	28.84	0.3672	3.571	35.555
5	-1.6	-1.034	1.621	-1.060	29.32	0.3621	3.820	34.094
6	-1.5	-1.011	1.569	-0.978	29.78	0.3569	4.066	32.730
7	-1.4	-0.987	1.518	-0.896	30.26	0.3518	4.312	31.457
8	-1.3	-0.963	1.467	-0.813	30.74	0.3467	4.561	30.262
9	-1.2	-0.940	1.415	-0.731	31.20	0.3415	4.807	29.161
10	-1.1	-0.916	1.364	-0.648	31.68	0.3364	5.056	28.143
11	-1.0	-0.892	1.312	-0.566	32.16	0.3312	5.302	27.216
12	-0.9	-0.869	1.261	-0.483	32.62	0.3261	5.551	26.378
13	-0.8	-0.845	1.209	-0.401	33.10	0.3209	5.797	25.628
14	-0.7	-0.821	1.158	-0.319	33.58	0.3158	6.043	24.969
15	-0.6	-0.797	1.107	-0.236	34.06	0.3107	6.292	24.394
16	-0.5	-0.774	1.055	-0.154	34.52	0.3055	6.538	23.910
17	-0.4	-0.750	1.004	-0.071	35.00	0.3004	6.787	23.511
18	-0.3	-0.726	0.952	0.011	35.48	0.2952	7.033	23.202
19	-0.2	-0.703	0.901	0.094	35.94	0.2901	7.282	22.982
20	-0.1	-0.679	0.850	0.176	36.42	0.2850	7.528	22.850
21	0.0	-0.655	0.798	0.258	36.90	0.2798	7.774	22.806
22	0.1	-0.632	0.747	0.341	37.36	0.2747	8.023	22.850
23	0.2	-0.608	0.695	0.423	37.84	0.2695	8.269	22.982
24	0.3	-0.584	0.644	0.506	38.32	0.2644	8.518	23.204
25	0.4	-0.561	0.592	0.588	38.78	0.2592	8.764	23.512
26	0.5	-0.537	0.541	0.671	39.26	0.2541	9.013	23.910
27	0.6	-0.513	0.490	0.753	39.74	0.2490	9.259	24.393
28	0.7	-0.490	0.438	0.835	40.20	0.2438	9.505	24.965
29	0.8	-0.466	0.387	0.918	40.68	0.2387	9.754	25.629
30	0.9	-0.442	0.335	1.000	41.16	0.2335	10.000	26.379
31	1.0	-0.418	0.284	1.083	41.64	0.2284	10.249	27.220
32	1.1	-0.395	0.232	1.165	42.10	0.2232	10.495	28.145
33	1.2	-0.371	0.181	1.248	42.58	0.2181	10.744	29.163
34	1.3	-0.347	0.130	1.330	43.06	0.2130	10.990	30.261
35	1.4	-0.324	0.078	1.412	43.52	0.2078	11.236	31.449
36	1.5	-0.300	0.027	1.495	44.00	0.2027	11.485	32.733
37	1.6	-0.276	-0.025	1.577	44.48	0.1975	11.731	34.101
38	1.7	-0.253	-0.076	1.660	44.94	0.1924	11.980	35.558
39	1.8	-0.229	-0.127	1.742	45.42	0.1873	12.226	37.094
40	1.9	-0.205	-0.179	1.825	45.90	0.1821	12.475	38.740
41	2.0	-0.182	-0.230	1.907	46.36	0.1770	12.721	40.448

3. Results of canonical path analysis for the DPPH response

> canonical.path(model2, dist = seq(-2, 2, by = 0.1))

	dist	x1	x2	x3	Temp	Conc	pH	yhat
1	-2.0	-2.190	-0.384	0.218	6.20	0.1616	7.654	52.327
2	-1.9	-2.094	-0.370	0.244	8.12	0.1630	7.732	48.713
3	-1.8	-1.999	-0.357	0.271	10.02	0.1643	7.813	45.313
4	-1.7	-1.903	-0.343	0.297	11.94	0.1657	7.891	42.071
5	-1.6	-1.808	-0.329	0.323	13.84	0.1671	7.969	39.042
6	-1.5	-1.712	-0.316	0.350	15.76	0.1684	8.050	36.167
7	-1.4	-1.617	-0.302	0.376	17.66	0.1698	8.128	33.505
8	-1.3	-1.521	-0.288	0.403	19.58	0.1712	8.209	30.998
9	-1.2	-1.426	-0.275	0.429	21.48	0.1725	8.287	28.707
10	-1.1	-1.330	-0.261	0.455	23.40	0.1739	8.365	26.576
11	-1.0	-1.235	-0.247	0.482	25.30	0.1753	8.446	24.644
12	-0.9	-1.139	-0.233	0.508	27.22	0.1767	8.524	22.884
13	-0.8	-1.044	-0.220	0.534	29.12	0.1780	8.602	21.327
14	-0.7	-0.949	-0.206	0.561	31.02	0.1794	8.683	19.946
15	-0.6	-0.853	-0.192	0.587	32.94	0.1808	8.761	18.741
16	-0.5	-0.758	-0.179	0.613	34.84	0.1821	8.839	17.733
17	-0.4	-0.662	-0.165	0.640	36.76	0.1835	8.920	16.896
18	-0.3	-0.567	-0.151	0.666	38.66	0.1849	8.998	16.253
19	-0.2	-0.471	-0.137	0.692	40.58	0.1863	9.076	15.790
20	-0.1	-0.376	-0.124	0.719	42.48	0.1876	9.157	15.514
21	0.0	-0.280	-0.110	0.745	44.40	0.1890	9.235	15.421
22	0.1	-0.185	-0.096	0.771	46.30	0.1904	9.313	15.513
23	0.2	-0.089	-0.083	0.798	48.22	0.1917	9.394	15.791
24	0.3	0.006	-0.069	0.824	50.12	0.1931	9.472	16.249
25	0.4	0.102	-0.055	0.851	52.04	0.1945	9.553	16.899
26	0.5	0.197	-0.042	0.877	53.94	0.1958	9.631	17.724
27	0.6	0.293	-0.028	0.903	55.86	0.1972	9.709	18.743
28	0.7	0.388	-0.014	0.930	57.76	0.1986	9.790	19.940
29	0.8	0.484	0.000	0.956	59.68	0.2000	9.868	21.330
30	0.9	0.579	0.013	0.982	61.58	0.2013	9.946	22.887
31	1.0	0.675	0.027	1.009	63.50	0.2027	10.027	24.653
32	1.1	0.770	0.041	1.035	65.40	0.2041	10.105	26.580
33	1.2	0.866	0.054	1.061	67.32	0.2054	10.183	28.708
34	1.3	0.961	0.068	1.088	69.22	0.2068	10.264	31.009
35	1.4	1.057	0.082	1.114	71.14	0.2082	10.342	33.510
36	1.5	1.152	0.096	1.140	73.04	0.2096	10.420	36.172
37	1.6	1.248	0.109	1.167	74.96	0.2109	10.501	39.048
38	1.7	1.343	0.123	1.193	76.86	0.2123	10.579	42.077
39	1.8	1.439	0.137	1.219	78.78	0.2137	10.657	45.320
40	1.9	1.534	0.150	1.246	80.68	0.2150	10.738	48.720
41	2.0	1.630	0.164	1.272	82.60	0.2164	10.816	52.334

Pertanika J. Sci. & Technol. 34 (2): 823 - 844 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Green Synthesis of Carbon Dots from Empty Fruit Bunch Using Hydrothermal for Tetracycline Removal

Wan Nuraishah Wan Ishak[1], Ying Pei Lim[1]*, Huey Ling Tan[1], Noor Fitrah Abu Bakar[1], Ng Law Yong[2,3], and Chun Yang Yin[4]

[1]School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[2]Department of Chemical Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia

[3]Centre for Advanced and Sustainable Materials Research (CASMR), Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia

[4]College of Engineering and Computer Science, VinUniversity, Vinhomes Ocean Park, Gia Lam, 10000 Hanoi, Vietnam

ABSTRACT

Tetracycline (TCs) residues in wastewater present serious environmental and health risks, necessitating the development of efficient and sustainable remediation strategies. This study reports the green synthesis of carbon dots (CDs) derived from oil palm empty fruit bunches (EFB) through a one-pot hydrothermal process, offering an eco-friendly alternative for antibiotic degradation. The synthesised CDs were characterised by SEM-EDX, TEM, UV-DRS, FTIR, and XRD analyses, confirming their structural and optical properties. Photocatalytic experiments were conducted under various conditions, including different catalyst dosages (0.5-2.5 g/L), initial TCs concentrations (5-25 ppm), and light sources (solar, visible, and UV). The CDs achieved a maximum degradation efficiency of 50.14% at 1 g/L dosage and 5 ppm TCs concentration under UV light. Moreover, the CDs demonstrated good stability and reusability across multiple cycles. Mechanistic analysis suggests that photocatalytic degradation primarily involves the generation of hydroxyl (•OH) and superoxide ($O_2•^-$) radicals. These findings highlight the potential of biomass-derived CDs as an efficient, sustainable, and reusable photocatalyst for antibiotic removal from wastewater.

Keywords: Antibiotic degradation, biomass valorisation, carbon nanomaterials, photocatalytic activity, sustainable wastewater treatment

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 06 June 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.10

E-mail addresses:
2022288228@student.uitm.edu.my (Wan Nuraishah Wan Ishak)
hueyling@uitm.edu.my (Huey Ling Tan)
fitrah@uitm.edu.my (Noor Fitrah Abu Bakar)
lyng@utar.edu.my (Law Yong Ng)
chunyang.y@vinuni.edu.vn (Chun Yang Yin)
yingpei@uitm.edu.my (Ying Pei Lim)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Antibiotic contamination in water bodies has become a global environmental concern due to its role in promoting antimicrobial resistance (AMR), which the World Health Organisation (2021) has identified as a major public health threat of the 21st century. Among the various antibiotics, tetracyclines (TCs) are the second most widely used group after sulfonamides, with extensive applications in human healthcare, veterinary medicine, and aquaculture (Niu et al., 2013). The frequent use and poor metabolisation rates of TCs result in their continuous discharge into the environment, contributing to the development of antibiotic-resistant bacteria and genes. In Malaysia, for example, Thiang et al. (2021) reported the presence of 23 antibiotics, including a high detection frequency of TCs in aquaculture farms. Similar pollution has been recorded globally, from Europe to North America and Asia, highlighting its widespread environmental relevance. Therefore, the removal of TCs from wastewater is a pressing global necessity.

Conventional wastewater treatment methods, such as ozonation, enzymatic degradation, activated carbon adsorption, and chlorination, have shown limited effectiveness in completely removing persistent pharmaceutical pollutants (Abejón et al., 2015; Gopal et al., 2020; Saadati et al., 2016). As a result, advanced oxidation processes (AOPs), particularly photocatalysis, have garnered significant attention as sustainable and promising alternatives. Photocatalysis, a green AOP that utilises light energy to activate semiconductor materials that generate reactive oxygen species (ROS) capable of mineralising organic contaminants, offering an eco-friendly and efficient method for antibiotic degradation (Hak et al., 2020). In this study, carbon dots (CDs) were synthesised using empty fruit bunch (EFB), a lignocellulosic agricultural waste generated in large quantities during crude palm oil (CPO) production. EFB is a fibrous byproduct obtained after the removal of oil-rich fruits from fresh fruit bunches during the sterilisation and threshing steps in palm oil milling. According to Hasanudin et al. (2023), approximately 20-22 million tonnes of EFB are produced annually in Malaysia, making it one of the most abundant biomass wastes in the country. Due to its high cellulose and lignin content, EFB is a promising, renewable carbon source for the synthesis of nanomaterials. Utilising EFB not only adds value to agricultural waste but also supports circular bioeconomy initiatives and reduces the environmental burden associated with biomass disposal. This aligns with the findings of Hasanudin et al. (2023), who emphasised the underutilised potential of EFB in Malaysia's palm oil industry.

Among various photocatalysts, carbon dots (CDs) have recently attracted significant attention due to their unique optical properties, tunable band gaps (2.47-4.5 eV), abundant surface functionalities, and excellent electron transfer capabilities (Jumardin et al., 2021; Khairol Anuar et al., 2021; Sutanto et al., 2020). CDs possess numerous surface functional groups and exhibit a hybrid structure of amorphous and nanocrystalline domains (Liu et al., 2020). With sizes below 10 nm, CDs are considered highly promising nanomaterials

for environmental remediation due to their strong photoluminescence, good absorption properties, and environmentally benign nature (Akbar et al., 2021; Liu et al., 2020). Among carbon nanomaterials such as nanotubes, fullerenes, and graphene, CDs have shown promise in photocatalysis because of their tunable band gap and rich surface chemistry (Jamaludin et al., 2018). Several studies have reported the application of CDs derived from agricultural biomass such as rice husk, banana peel, and sugarcane bagasse for pollutant degradation (Liu et al., 2020). CDs synthesised from oil palm biomass, including EFB, have been reported. However, their specific application in the photocatalytic degradation of antibiotics such as tetracycline remains significantly underexplored.

Recently, CDs have emerged as versatile photocatalysts for wastewater treatment, demonstrating high efficiency in degrading organic pollutants and antibiotics. Notably, CDs achieved 96% and 98% degradation of methylene blue and Congo red, respectively (Sawalha et al., 2021). Furthermore, CDs possess the ability to be tailored via functionalisation for specific applications. For instance, Liu et al. (2020) reported the synthesis of magnetic polyethyleneimine-functionalised carbon quantum dots, incorporating CDs into magnetic nanomaterials, which achieved an impressive uranium removal efficiency of 88%. CDs also show promising performance in removing polycyclic aromatic hydrocarbons (PAHs) such as benzopyrene, with an adsorption efficiency of 93.9% with the use of 9mg fatty acid-coated magnetic nanoparticles at a temperature of 25 °C (Cui et al., 2021). Apart from dyes and PAHs, CDs can be combined with other photocatalysts to degrade antibiotics. For example, CDs with bismuth vanadate and tantalates achieved 89.3% and 87.1% removal of amoxicillin and ciprofloxacin, respectively (Abinaya et al., 2021; Bai et al., 2022; Mukherjee et al., 2021).

Despite their promising photocatalytic activity, critical knowledge gaps persist regarding the reusability and underlying mechanisms of CDs. While numerous studies highlight their efficacy, comprehensive investigations into this area remain limited. However, existing studies have rarely explored the photocatalytic mechanism, reusability, and radical interaction pathways of biomass-derived CDs in antibiotic degradation, particularly in the context of tetracycline. Bridging these gaps is essential for realising the full potential of CDs in practical wastewater treatment applications. CDs can be synthesised using two main methods, which are the top-down approach and the bottom-up approach (Liu et al., 2020). The chosen method significantly impacts CDs' characteristics, including surface functional groups, size, and photoluminescence properties (Khairol Anuar et al., 2021). Top-down methods (arc-discharge, laser ablation) offer precise size control but require expensive equipment, harsh conditions, and often yield low quantities. Bottom-up approaches (hydrothermal, microwave, pyrolysis) are simple, low-cost, and scalable, yet can suffer from broader size distributions and longer reaction times. Among these, hydrothermal synthesis, classified as a bottom-up method, is widely favoured due

to its simplicity, environmental friendliness, cost-efficiency, and feasibility for large-scale production (Liu et al., 2020).

This study aims to contribute to the advancement of green nanomaterials by utilising EFB, a major agricultural waste product in Malaysia and Southeast Asia, for the sustainable synthesis of CDs. This valorisation not only promotes environmental sustainability by mitigating EFB disposal issues but also introduces a novel, renewable photocatalyst for antibiotic removal. Specifically, this study synthesises CDs via a hydrothermal method using EFB as the carbon precursor, characterises their structural, optical, and morphological properties using TEM, XRD, UV-DRS, and FTIR, evaluates their photocatalytic performance for tetracycline degradation under different operational parameters, and investigates the underlying photocatalytic mechanism through reactive species trapping experiments. To our knowledge, this is the first study utilising hydrothermally synthesised CDs derived from EFB for the photocatalytic degradation of tetracycline, addressing a critical gap in antibiotic wastewater remediation using palm biomass.

MATERIALS AND METHODS

Material

Empty fruit bunches were sourced from a palm oil plantation located in Selangor, Malaysia. Tetracycline (TCs > 95%) antibiotics with the molecular formula of $C_{22}H_{24}N_2O_8$ were supplied by Merck Millipore. All chemicals were analytical grade and were used without further purification. Deionised (DI) water served as the solvent for solution preparation. UV lamps, Kintons KT6-UVC (2 units x 15 W) were procured from a local Malaysian supplier.

Synthesis of Carbon Dots

The EFB was subjected to a rigorous washing procedure with tap water to eliminate any adhering contaminants or debris. Subsequently, the EFB was spread out in a monolayer and exposed to direct sunlight until complete desiccation was achieved. The dried EFB was then mechanically grinded into smaller particles using a grinding machine. Finally, the grinded EFB was sieved through a mesh of 0.25 mm mesh to obtain a uniform particle size distribution. The sieved EFB particles were then mixed with deionised (DI) water. The synthesis of CDs was modified according to previous literature hydrothermal methods (Egorova et al., 2018; Mahat & Shamsudin, 2020), as illustrated in Figure 1. Briefly, 6 g of EFB powder was mixed with 150 mL of deionised (DI) water in a Teflon-lined autoclave reactor. The reactor was then subjected to hydrothermal treatment at 200 °C for 3 hours. Thereafter, the reactor was allowed to cool down overnight. The resultant brown solution

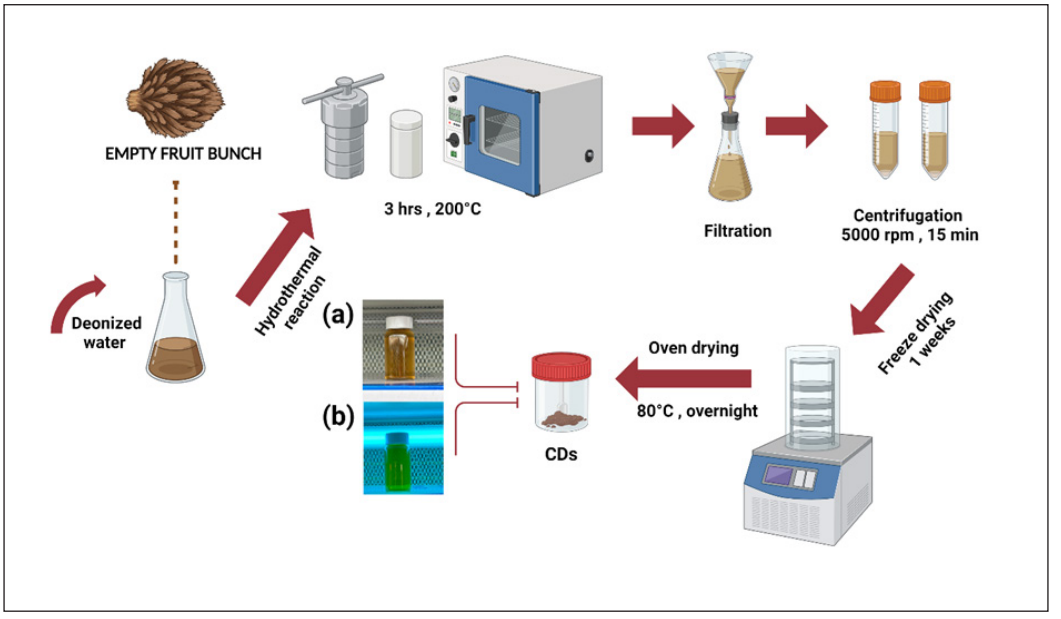


Figure 1. Synthesis route of carbon dots (CDs) derived from empty fruit bunch via the hydrothermal method. (a) CDs under daylight and (b) CDs under UV light

was centrifuged at 5000 rpm for 15 minutes to remove precipitates and larger particles. The final solution containing the synthesised CDs was then freeze-dried, followed by oven-drying at 80 °C, and stored for further characterisation.

Characterisation of Synthesised CDs

The functional group of the synthesised CDs was determined by Fourier Transform Infrared (FTIR) with a spectral range of 400 cm^{-1} to 4000 cm^{-1}. The crystalline structure of the photocatalyst was analysed using X-ray diffraction (Rigaku Ultima IV XRD) using monochromatised Cu Kα radiation of 40 kV and 30 mA with a scan rate of 8 min^{-1} between $2\theta = 10°$ to 90° with Cu Kα ($\lambda = 1.504$ Å). The morphology of the photocatalyst's structure was observed by Transmission Electron Microscope (TEM) using JEOL, 2100F at 200 kV. The CDs were dropped onto a copper grid and vacuum-dried for preparation before TEM characterisation. The elemental composition of CDs was determined by HITACHI scanning electron microscopy-energy-dispersive X-ray analysis (SEM-EDX) using 15 kV accelerating voltage. For SEM–EDX analysis of the carbon dots (CDs), each powder sample was mounted on a carbon-taped stub and sputter-coated with a thin layer of gold (~5 nm) using a Denton Desk V coater (Denton Vacuum, USA) to prevent charging effects. EDAX spectra were acquired on the freshly synthesised CDs, before any photocatalytic testing. The band gap was determined by Ultraviolet Diffuse Reflectance Spectroscopy (UV-DRS)

using a Shimadzu 3600. The diffuse reflectance spectroscopy was performed on the prepared sample using a PerkinElmer Lambda 750 UV-Vis/NIR spectrometer (Cary 300-800).

Photocatalytic Degradation of Tetracycline

The photocatalytic degradation test was conducted by mixing 1 g/L of synthesised CDs with 50 mL of 15 ppm TCs in a 100 mL Pyrex beaker over 180 minutes. Before light irradiation, the photocatalyst was stirred with tetracycline solution (15 ppm) in the dark for 30 minutes to establish adsorption-desorption equilibrium. This control step was conducted to assess the adsorption contribution and ensure any subsequent removal of tetracycline under light was due to photocatalytic degradation. Photocatalysis was then initiated by illuminating the solution with 2 units of 15 W UV lamps positioned 15 cm above the beaker. 5 mL of the solution was withdrawn from the beaker every 30 minutes by using a 0.22μm syringe filter. The concentration of TCs was determined from absorbance readings at 357nm using a Perkin Elmer UV-Vis/NIR Lamda 750. The photocatalytic degradation of TCs was repeated using different CDs dosages (0.5 g/L, 1.0 g/L, 1.5 g/L, 2.0 g/L, 2.5 g/L), TCs concentration (5 ppm, 10 ppm, 15 ppm, 20 ppm, 25 ppm) and source of light (solar light, visible light, UV light). For comparison with UV-induced degradation, additional photocatalytic tests were conducted under natural sunlight (10,000-30,000 lux intensity, clear sky, 12:00-2:00 PM) and visible light using an 80 W lamp with an emission peak at 400 nm. Control experiments were also conducted to evaluate photolysis and adsorption. For photolysis, the TCs solution was exposed to UV light without the addition of CDs, while for adsorption, the TCs solution was mixed with CDs and kept in the dark for 30 minutes. The percentage of TCs removal, E, was calculated using Equation 1 (Wu et al., 2020a), where C_i is the initial concentration of TCs antibiotics (ppm), and C_e is the equilibrium concentration of TCs antibiotics (ppm), respectively.

$$E(\%) = \frac{C_i - C_e}{C_i} \times 100\% \quad [1]$$

RESULTS AND DISCUSSION

Fourier Transform Infrared Spectroscopy (FT-IR)

Figure 2a illustrates the FTIR spectrum of the synthesised CDs. The broad peak observed at 3322 cm^{-1} corresponds to O–H stretching vibrations, indicating the presence of hydroxyl groups on the surface. A sharp peak at 1638 cm^{-1} is attributed to C=O stretching, suggesting carbonyl functional groups. A weaker peak appearing around 1538 cm^{-1} is assigned to N–H bending vibrations, indicating the incorporation of amine groups.

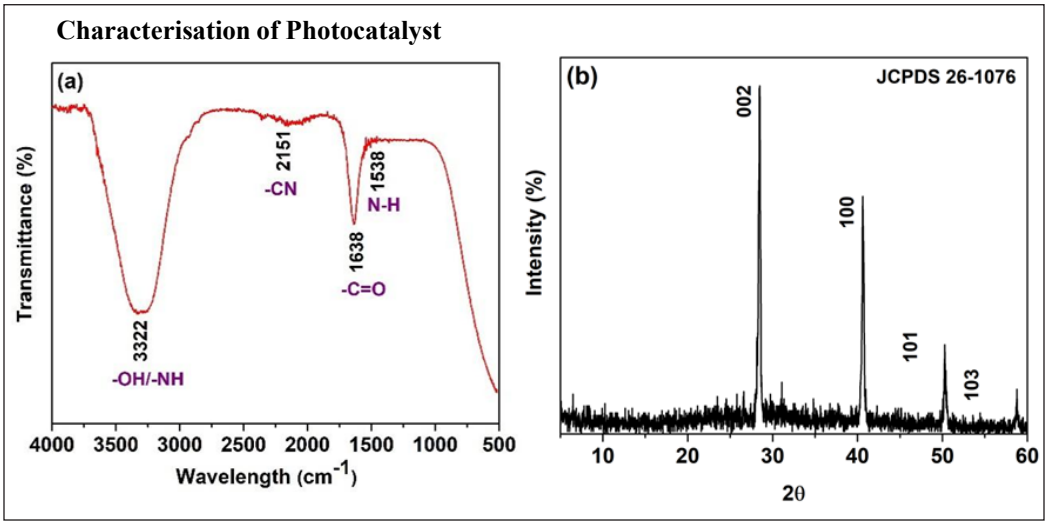


Figure 2. (a) FT-IR spectrum and (b) X-ray diffraction pattern of CDs

Additionally, a distinct peak at 2151 cm^{-1} is attributed to C≡N stretching, suggesting the presence of nitrogen-containing functionalities (Nair et al., 2020). The presence of these oxygen- and nitrogen-containing groups is typical for CDs synthesised via hydrothermal treatment and supports their hydrophilic nature and surface reactivity, as reported in previous studies (Ţucureanu et al., 2016).

X-ray Diffraction Analysis (XRD)

X-ray diffraction (XRD) analysis of the synthesised carbon dots (CDs), as depicted in Figure 2b, reveals a dominant peak at approximately 2θ = 28°, corresponding to the (002) plane of graphitic carbon, consistent with JCPDS No. 26-1076, indicating partial graphitisation. In addition to this primary peak, several distinct reflections are observed at around 2θ = 43°, 50.6°, and 58.8°, which can be assigned to the (100), (101), and (103) planes, respectively. These diffraction features suggest the presence of crystalline or semi-crystalline domains embedded within the carbon matrix. To validate the (002) assignment, Bragg's Law ($n\lambda = 2d \sin\theta$) was applied using Cu Kα radiation (λ = 0.15406 nm, n = 1, θ = 14°). The calculated interlayer spacing d = 0.319 nm, which agrees closely with the 0.32 nm lattice fringe spacing observed in HRTEM (Figure 3b). The absence of a broad hump commonly associated with amorphous carbon further supports the development of ordered structures during synthesis. These findings agree with previous studies on biomass-derived CDs synthesised via hydrothermal methods, which report similar XRD patterns indicative of polycrystalline characteristics and structural ordering (Jung et al., 2022; Mukherjee et al., 2021; N. Sharma et al., 2022).

Transmission Electron Microscopy (TEM) and Scanning Electron Microscope-energy Dispersive X-ray (SEM-EDX) Analysis

Figure 3a illustrates the morphology of the CDs from the TEM image. The synthesised CDs from EFB via the hydrothermal method were observed to have a spherical shape, consisting of a variety of sizes with slight aggregation, consistent with the findings stated by Xu et al. (2023) In Figure 3b, the lattice fringes of CDs are 0.32 nm, which corresponds well to the (002) plane of CDs in XRD analysis. This result is consistent with the work of Li et al. (2013). The size distribution of the CDs, measured to be less than 10 nm (Figure 3c), further corroborates their designation as nanomaterials, in agreement with previous studies (Liu et al., 2020). Table 1 presents the elemental composition of the synthesised CDs, determined by SEM-EDX. The values represent the average from multiple spot analyses to ensure accuracy and reduce operator bias. Carbon (30.4 wt%) and oxygen (46.1 wt%) were identified as the dominant elements, which is consistent with previous reports on carbon dots, where carbon and oxygen were similarly the predominant elements (Fawaz et al., 2023). The presence of minor elements such as K, Na, S, Ca, and Si can be ascribed to the inherent inorganic constituents of the biomass precursor. Potassium and sodium are biogenic elements inherently present in plant-derived biomass, typically absorbed from the soil during plant growth and retained during carbonisation or hydrothermal synthesis processes (Sahu & Sahoo, 2024). Oxygen is likely derived from oxygen-containing functional groups (e.g., –OH, –COOH, –C=O) introduced during the hydrothermal process. Sulphur detected in the CDs may be attributed to the natural sulphur content in the biomass, originating from intrinsic organic compounds and mineral uptake from soil and water, or retained within the ash content after hydrothermal treatment. Calcium and silicon are typically associated with residual mineral ash content within lignocellulosic biomass (Wang et al., 2022).

UV-Vis Diffuse Reflectance Spectroscopy (UV-DRS)

Analysis of the peaks in the optical absorption spectrum (Figure 4a) provides insights into the molecular transformations within the CDs. The sharp peak observed at 241 nm is attributed to the π-π^* transition in aromatic C-C and C=C bonds, while the peak at 287 nm corresponds to the n-π^* transition in C=O bonds (Mahat & Shamsudin, 2020). Additionally, a distinct tailing shape is observed in the wavelength range of 380 nm to 780 nm, corresponding to the visible light region. These findings are presented in Figure 4a, confirming the consistency of the synthesised CDs analysis with previous research (Abinaya et al., 2021). Moreover, Figure 1b shows the fluorescence behaviour of the CDs in solution under UV-light irradiation at a wavelength of 365 nm, where the CDs exhibited a green fluorescence. Conversely, the solution appeared as a light shade of brown under daylight, as shown in Figure 1a. This colouration agrees with the findings reported by Mahat and Shamsudin, (2020).

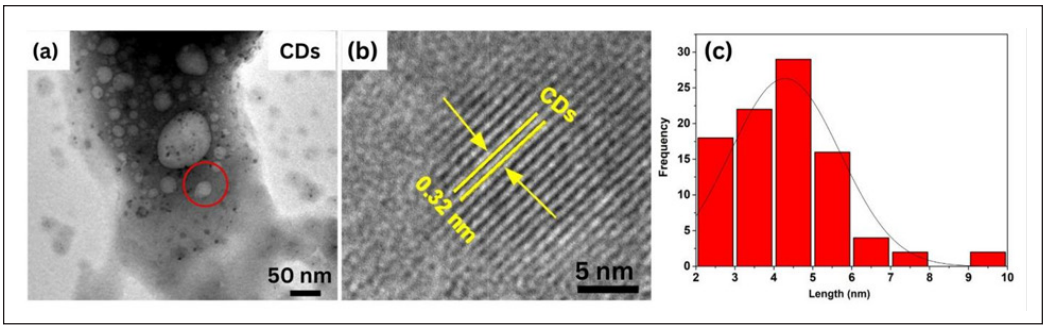


Figure 3. (a) TEM images; (b) HRTEM images; and (c) particle distribution of CDs

Table 1
Average elemental composition of CDs from multiple SEM-EDX spots

Element	C	O	K	Na	P	S	Ca	Cl	Si
Weight (%)	18.5	37.4	32.4	0.7	1.2	1.2	0.8	5.1	1.8
Atomic (%)	30.4	46.1	16.3	0.6	1.2	1.2	0.4	2.8	1.2

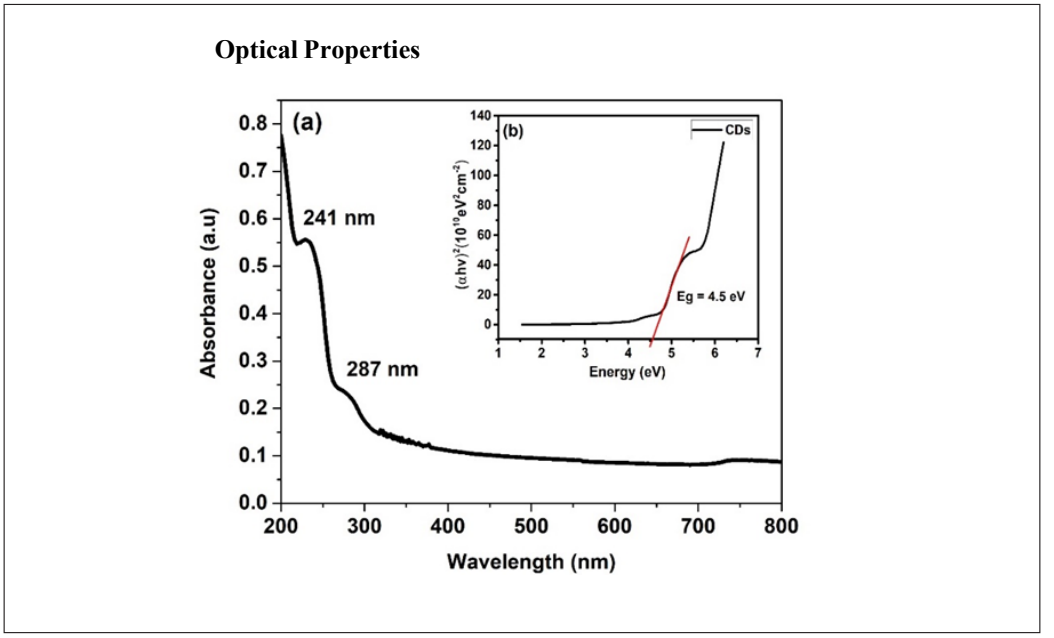


Figure 4. (a) UV-Vis diffuse reflectance of CDs and (b) Tauc plot of CDs for determination of the bandgap (inset)

The green fluorescence of CDs is primarily attributed to the presence of C=N bonds within their structure. These bonds are crucial as they influence the electronic structure, allowing CDs to emit green light when excited. This phenomenon is supported by previous studies showing that CDs with specific nitrogen concentrations exhibit significant photoluminescence due to C=N and C-N bonds (Hu et al., 2022). In addition, a wide band gap indicates that the material primarily responds to higher-energy photons, such as ultraviolet UV light, while a narrower band gap allows for the absorption of visible light. Therefore, understanding the band gap of a photocatalyst is vital for optimising its performance in applications such as pollutant degradation and wastewater treatment. The band gap of the photocatalyst was determined using Tauc's plot analysis as shown in Figure 4b. The plot of *hv* (photon energy) versus energy revealed a band gap value of 4.5 eV, indicating that the photocatalyst operates within the UV light range. This relatively large band gap suggests limited π-conjugation within the carbon core, likely influenced by quantum confinement effects and surface passivation. In contrast, most biomass-derived carbon dots typically exhibit band gaps in the range of 2.0-3.5 eV, depending on the source and synthesis conditions, e.g., CDs derived from chitosan or plant materials (Mintz et al., 2019). The higher band gap observed in our study may be attributed to specific factors, including the synthesis method and precursor composition, which could result in a more restricted π-conjugation and more prominent surface passivation. Additionally, quantum confinement effects due to the smaller particle size could further increase the band gap energy. As reported by Yu et al. (2018), both the carbon core and surface states contribute significantly to the optical behaviour of carbon dots, with surface states exhibiting a band gap of approximately 4.5 eV-narrower than the core's band gap (~5.0 eV). This supports our observation that optical transitions in the synthesised material are likely dominated by surface-state contributions. Similar findings have also been discussed by Mintz et al. (2019) and Zhou et al. (2018), who emphasised the role of surface functionalities in modulating the band structure and photophysical behaviour of carbon-based nanomaterials.

Effect of CD's Dosage

Figure 5a investigates the influence of photocatalyst (CDs) loading (0.5-2.5 g/L) on TCs degradation within 180 minutes for an initial TCs concentration of 15 ppm. An optimal CDs dosage of 1 g/L (red bar) is observed, corresponding to the highest TCs degradation efficiency (40.35%). This could be attributed to the increased availability of active sites on the catalyst surface at this concentration, facilitating a greater number of TCs molecules to undergo photocatalytic degradation into harmless byproducts like carbon dioxide and water. Our findings are aligned with previous reports who observed degradation rates of 83.5%, 86.2%, and 81.7% for CDs dosages of 0.5 g/L, 1 g/L, and 2 g/L, respectively (Saadati et al., 2016; Safari et al., 2015; Yan et al., 2015). It is also noteworthy that exceeding the

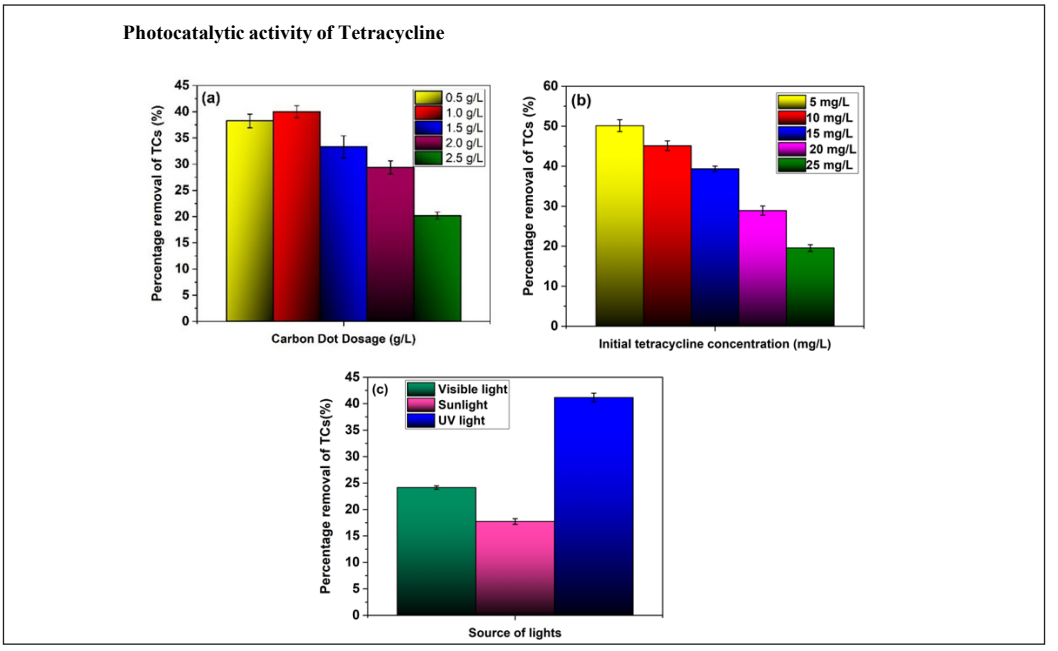


Figure 5. Percentage removal of TCs at different (a) dosages; (b) initial concentrations of tetracycline; and (c) sources of light

optimal CDs loading can hinder pollutant removal efficiency. A possible explanation for these decreases at higher CDs dosages could be the aggregation of photocatalysts, which restricts the accessibility of active sites and the generation of reactive species responsible for TCs degradation. Additionally, at high CDs loadings, the pollutant molecules might compete with excess CDs particles for available active sites on the catalyst surface, further hindering pollutant adsorption and degradation. Therefore, optimising catalyst loading and mitigating excessive aggregation are critical factors for maximising the photocatalytic performance in wastewater treatment applications. Similar observations were reported by those who identified an optimal dosage of 1.5 g/L, with degradation efficiency declining at higher loadings of 0.25-2.5 g/L (Azimi & Nezamzadeh-Ejhieh, 2015).

Effect of Concentration of TCs

Figure 5b depicts the TCs removal percentage by manipulating TCs concentration (5-25ppm) using 1 g/L CDs within 180 minutes. It was found that TCs removal efficiency was inverse of its concentration within the range of 5-25 ppm, with the highest removal efficiency recorded as 50.14 %. The correlation between the degradation pattern of TCs and its concentration could be attributed to the blocking effect of excessive TCs molecules,

which hinders the penetration of UV light in the solution (Wu et al., 2021b). In a similar work conducted by Saadati et al. (2016) the lowest TCs concentration (10 ppm) exhibited the highest degradation rate at 79.1 %. However, as TC concentration increased from 15 ppm to 20 ppm and 30 ppm, the percentage degradation of TCs decreased slightly. At lower TCs concentrations, the CDs did not experience surface saturation, thus enabling rapid photocatalytic production of superoxide and hydroxyl ions for pollutant degradation (Niu et al., 2013). Initial tetracycline concentrations (15-25 ppm) were selected based on previous laboratory-scale photocatalytic studies of antibiotic removal and to represent highly contaminated pharmaceutical effluent conditions. This range is commonly used in the literature to simulate realistic pollution scenarios and evaluate photocatalytic efficiency under such conditions.

Effect of Source of Light on the CDs

In the present work, UV-light, visible light and solar were selected to investigate the effect of the light source on the TCs removal (TCs concentration: 15 ppm, CDs dosage: 1 g/L). As illustrated in Figure 5c, the TC's degradation efficiency corresponding to the effect of light source followed: UV light (41.2%) > visible light > (24.15%) > sunlight (17.74%). This observation could be attributed to the band gap of the synthesised CDs (~4.5 eV), which primarily absorbs light in the UV region as shown in Figure 4 (inset). The relatively low photocatalytic degradation efficiency observed under UV irradiation may be attributed to the wide band gap (4.5 eV) of the synthesised carbon dots, which limits their light absorption to the UV range and may reduce the generation and separation efficiency of charge carriers. Additionally, surface-related recombination and the absence of co-catalysts or composite structures could also contribute to the limited activity. These findings are consistent with previous studies by Zhong et al. (2020) and M. Sharma et al. (2022), who also observed reduced photocatalytic degradation of sulfamethoxazole (SMX) and enrofloxacin (ENR) under visible light compared to UV light. Similarly, Khan et al. (2010) reported a significantly higher degradation efficiency (71.8%) under UV irradiation, compared to 40% under visible light. Interestingly, although sunlight contains both UV and visible light components, its degradation efficiency was lower than that under visible light alone. This counterintuitive result may be due to several contributing factors. Firstly, solar irradiance is naturally variable and dependent on atmospheric and geographic conditions, potentially causing fluctuations in light intensity during the reaction period (Chong et al., 2010).

Secondly, the proportion of UV photons in sunlight during the early afternoon (12:00-2:00 PM) could have been relatively low, with the spectral distribution shifting more toward longer wavelengths. In contrast, the visible light source used in this study

emitted strongly at 400 nm, providing a more consistent and effective excitation of the CDs. Thirdly, outdoor exposure under sunlight may introduce scattering, reflective losses, or elevated solution temperatures, all of which can negatively affect photocatalytic activity through reduced photon flux or enhanced recombination of charge carriers (Pelaez et al., 2012). Furthermore, CDs demonstrated some activity under visible light irradiation, possibly due to surface states or defect sites that can trap visible light photons (Akbar et al., 2021). This facilitates electron-hole pair generation despite the larger band gap. These defects can introduce intermediate energy levels within the band gap, reducing the energy required for electron excitation and enabling visible light absorption (Liu et al., 2020).

The reusability and stability of the synthesised carbon dots (CDs) were evaluated for their efficacy in tetracycline degradation through five consecutive cycles under identical reaction conditions (5 ppm tetracycline, 1 g/L CDs, 180 minutes irradiation time). Following each cycle, the reaction media were centrifuged to recover the CDs. The recovered CDs were then thoroughly washed with a 1:3 ethanol/water mixture (w/w or v/v) and oven-dried at 80 °C overnight (Mukherjee et al., 2021). As shown in Figure 6a, the tetracycline degradation efficiency exhibited a minimal decrease from 50.08% in the initial cycle to 44.01% after the fifth recycling cycle. Furthermore, X-ray diffraction (XRD) analysis was performed on the CDs before and after the five-cycle experiment to assess the reuse of the CDs' structural integrity. The XRD results demonstrated no significant changes in the crystalline structure of the CDs, indicating their high chemical stability throughout the degradation process (Figure 6c). Various scavenging experiments were conducted following the established protocol by previous literature (Niu et al., 2013; Trenczek-Zajac et al., 2022; Wu et al., 2020a) to elucidate the dominant reactive species involved in the photocatalytic degradation of tetracycline using CDs under UV light irradiation. The employed scavengers included ethylenediamine tetraacetic acid disodium salt (EDTA-2Na) for h^+, silver nitrate ($AgNO_3$) for e^-, p-benzoquinone (p-BQ) for $\bullet O_2^-$, and tert-butanol (BuOH) for •OH radicals. As shown in Figure 6b, the introduction of EDTA-2Na and $AgNO_3$ did not significantly affect tetracycline degradation. Conversely, the inclusion of BuOH and PBQ as scavengers for •OH and $\bullet O_2^-$ radicals, respectively, significantly suppressed tetracycline degradation by 38.56% and 46.73%, respectively. These findings strongly suggest that •OH and $\bullet O_2^-$ radicals play the most critical roles in tetracycline degradation mediated by the CDs as shown in Figure 7. Generally, the effectiveness of photocatalytic degradation relies on the generation of ROS, which is contingent upon the photocatalyst's ability to suppress exciton recombination. This suppression enabled the excitons to undergo alternate transitions, thereby facilitating the production of ROS. Here, the CDs photocatalyst absorbs UV light of energy greater than its bandgap, inducing photoexcitation. This process promotes the transition of electrons from the valence band to the conduction band, generating electron-hole pairs (e^-/h^+) within the CDs. The electrons in the conduction band can interact with

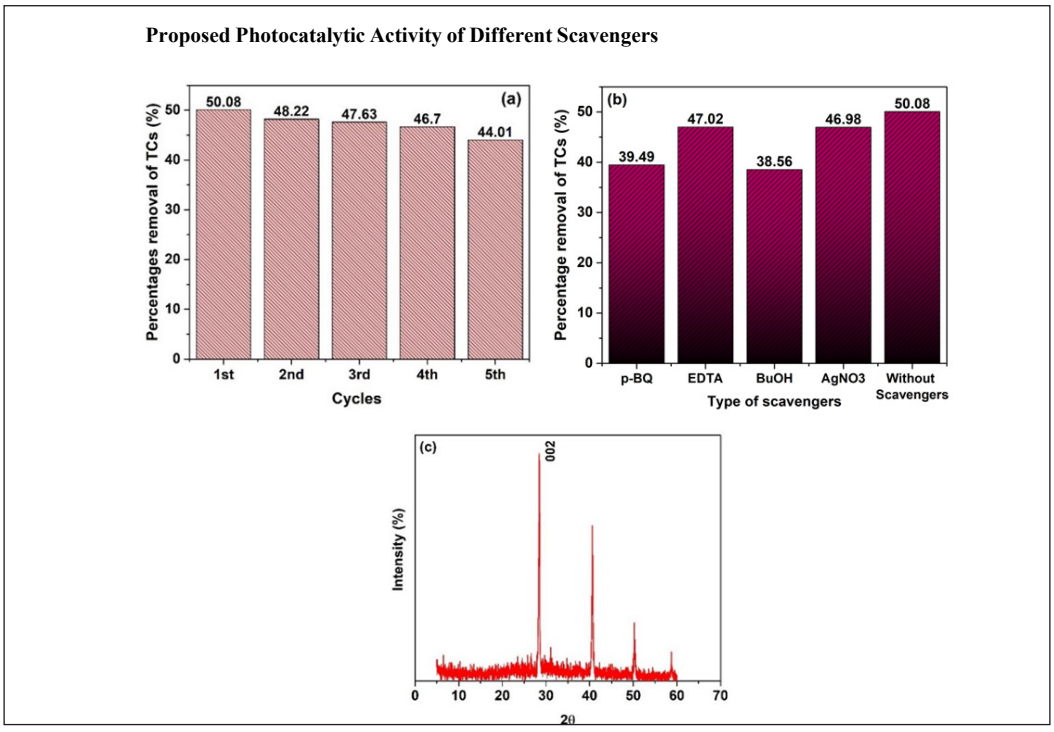


Figure 6. (a) Reusability of CDs up to five cycles; (b) scavenger test; and (c) XRD patterns of recycled photocatalyst

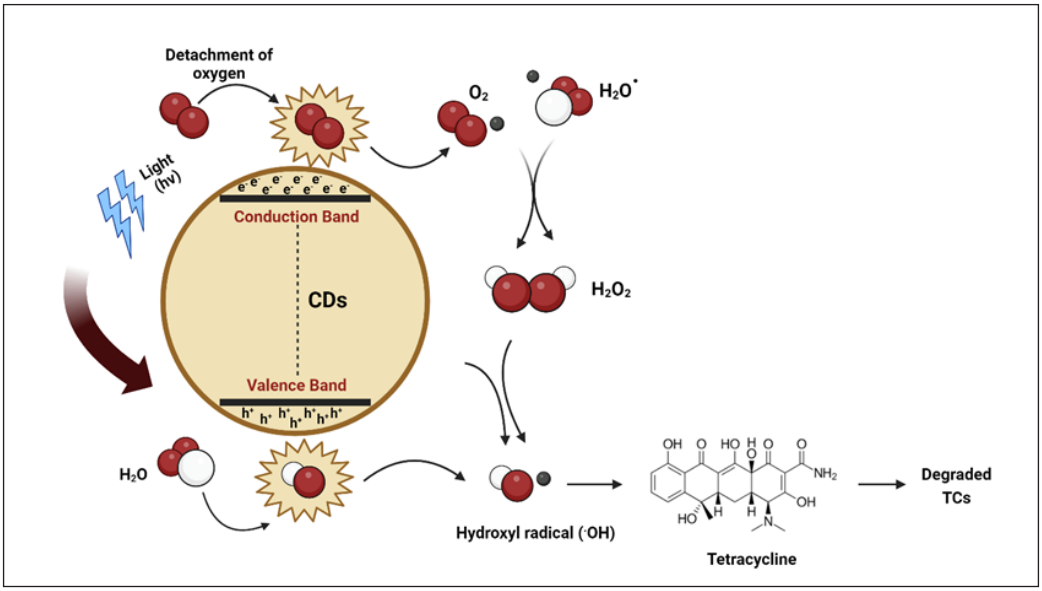


Figure 7. Proposed mechanism of degradation TCs by CDs

Table 2
The kinetic parameters at different concentrations of TCs degradation

Initial Concentration (mgL^{-1})	Reaction Rate, k (min^{-1})
5	0.0031
10	0.0028
15	0.0021
20	0.0014
25	0.01

oxygen (O_2) molecules to form superoxide radicals (O_2•-). Conversely, the holes in the valence band can react with water (H_2O) molecules to produce hydroxyl radicals (•OH) and hydrogen peroxide (H_2O_2). Notably, hydrogen peroxide (H_2O_2) can further decompose to generate additional hydroxyl radicals (•OH). The hydroxyl radicals (•OH), being highly reactive, attack the tetracycline (TCs) molecules, breaking them down into smaller, less harmful molecules. This phenomenon leads to the degradation of TCs into simpler, environmentally benign compounds, ultimately reducing its presence as a pollutant.

Kinetic study for the Photodegradation of TCs

The photodecomposition of TCs was further analysed using a pseudo-first-order kinetic model. The natural logarithm (ln) of the initial concentration of TCs (C_o) and the concentration remaining after a specific irradiation time (C_t) was plotted against irradiation time. The rate constant (k), representing the pseudo-first-order reaction rate coefficient, was then obtained from the slope of the resulting linear plot, and the data were depicted in Table 2. This approach provides an apparent first-order rate constant based on the straight slope of the ln (C_o/C_t) versus time relationship. Following the Langmuir-Hinshelwood (LH) kinetic model, the initial degradation rate (r_o) can be expressed by the Equation 2, 3, and 4:

$$-(\ln C - \ln C_o) = kt \quad [2]$$

$$r_o = -\frac{dC}{dt} = \frac{K_R K_{LH} C_o}{1 + K_{LH} C_o} = k_{app} C_o \quad [3]$$

$$\frac{1}{k_{app}} = \frac{1}{K_R K_{LH}} + \frac{C_o}{K_R} \quad [4]$$

where kr is the pseudo-first order Langmuir-Hinshelwood type rate coefficient, and k_{LH} is a pseudo-equilibrium constant related to monolayer adsorption on the catalyst surface (Equations 2 , 3, and 4). Our investigation into the normalised concentration versus 1/kapp plot suggests a two-step photodecomposition mechanism for TCs. The first step involves pollutant adsorption onto the active sites of CDs, followed by photochemical degradation consistent with the Langmuir-Hinshelwood (LH) kinetic model. The Langmuir-type correlation observed between the initial degradation rate and concentration strongly supports the occurrence of adsorption during the photocatalytic reaction. The Langmuir-type correlation observed between the initial degradation rate and concentration further supports this adsorption-controlled mechanism. Although the photocatalytic reaction takes place in a suspended system, the kinetic behaviour and adsorption parameters are consistent with the LH model, indicating that the degradation of tetracycline proceeds predominantly through surface-mediated mechanisms rather than bulk-phase reactions. Notably, the calculated reaction rate constant (k_r = 0.01654 mg L^{-1} min^{-1}) and the relatively high adsorption equilibrium constant, k_{LH} (1.7653 mg L^{-1}), indicate a strong interaction between TCs and the catalyst surface shown in Figure 8. These findings align with Syahin et al. (2019), who reported similar behaviour in the photocatalytic degradation of phenol. A comparative analysis presented in Table 3 reveals that the carbon dot photocatalyst synthesised in this study exhibited superior photocatalytic degradation of tetracycline compared to those reported in previous literature. These findings underscore the promising potential of the synthesised carbon dots in combating antibiotic contaminants present in water.

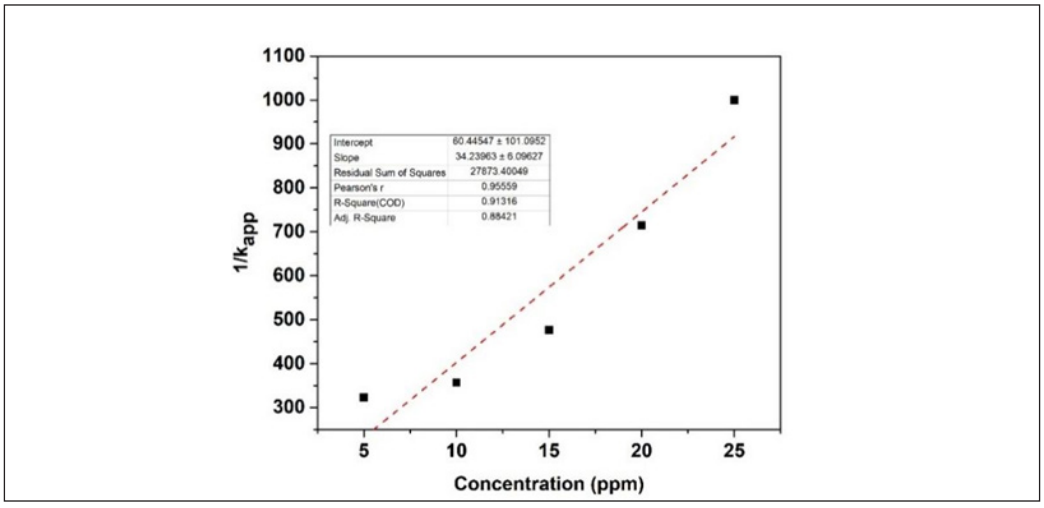


Figure 8. The relationship between $1/k_{app}$ and the initial concentration of TCs

Table 3
Comparison of previous literature on CDs based photocatalysts for photodegradation

Photocatalyst	Morphology	Source of Light	Degradation Efficiency	Fluorescence	References
Carbon dots	Spherical	Natural Sunlight	43 %, 10 hr for removal of methylene blue	Blue emission	(Jusuf et al., 2018)
Polymer-modified GQDs	Spherical	300 W, Xe lamp	45 %, 100 min for removal of methylene blue	Green emission	(Fan et al., 2016)
GQDs	Spherical	Natural Sunlight	45 %, 100 min for removal of methylene blue	-	(Kumar et al., 2017)
Carbon dots	-	Visible light	42.8 %, 8 days	Blue emission	(Liu et al., 2021)
$Bi_2O_2CO_3$/ NCQDs	Spherical	Near-infrared light	35 %, 360 min for removal of ciprofloxacin	-	(Hu et al., 2021)
Carbon dots	Spherical	UVB light	9.48 %, 60 min for decolourisation of the organic compound	-	(Ayu et al., 2023)
Carbon dots	-	Stimulated solar light	47 %, 4 hr for removal of antibiotics	Blue emission	(Louros et al., 2021)
Carbon dots	Spherical	UV light	50.14 %, 180 min for removal of tetracycline	Green emission	(Syahin et al., 2019)

CONCLUSION

In conclusion, this study successfully demonstrated the synthesis of carbon dots (CDs) from empty fruit bunch (EFB) waste via a simple one-pot hydrothermal method. The EFB-derived CDs exhibited efficient photocatalytic activity for tetracycline (TCs) degradation in wastewater treatment. Optimal TCs removal was achieved at a CDs dosage of 1 g/L and a TCs concentration of 5 ppm under UV light irradiation. Furthermore, preliminary studies indicate that the CDs exhibit good reusability and recyclability, suggesting their potential for long-term application in wastewater treatment. The photocatalytic mechanism primarily involves the generation of hydroxyl (•OH) and superoxide ($\bullet O_2$-) radicals under light irradiation, leading to the oxidation and mineralisation of TCs molecules. These findings highlight the potential of EFB as a sustainable precursor for photocatalyst production, contribute significantly to the field of environmental remediation and waste valorisation, offering a sustainable solution for both agricultural waste management and

water treatment. Future research efforts could focus on optimising the synthesis conditions to further enhance photocatalytic activity and explore the feasibility of immobilising CDs on suitable supports for practical wastewater treatment applications.

ACKNOWLEDGEMENT

This work was financially supported by the My RA (GPM/033/2022). The authors also extend their gratitude to the College of Engineering at Universiti Teknologi, MARA for their assistance in completing the research project.

LIST OF ABBREVIATIONS

AMR	:	Antimicrobial resistance
AOPs	:	Advanced oxidation processes
CDs	:	Carbon dots
CPO	:	Crude palm oil
DI	:	Deionised
EFB	:	Empty fruit bunch
FTIR	:	Fourier transform infrared spectroscopy
HRTEM	:	High resolution transmission electron microscopy
LH	:	Langmuir-Hinshelwood
PAHs	:	Polycyclic Aromatic Hydrocarbons
PL	:	Photoluminescence
ROS	:	Reactive oxygen species
SEM	:	Scanning electron microscopy
SEM-EDX	:	Scanning electron microscopy-energy dispersive X-ray
TEM	:	Transmission electron microscopy
TC	:	Tetracycline
TCs	:	Tetracyclines
UV	:	Ultraviolet
UV-DRS	:	Ultraviolet diffuse reflectance spectroscopy
UV–Vis	:	Ultraviolet-visible spectroscopy
XRD	:	X-ray diffraction
h^+	:	Hole
e^-	:	Electron
•OH	:	Hydroxyl radical
$O_2•^-$	:	Superoxide radical
H_2O_2	:	Hydrogen peroxide
EDTA-2Na	:	Ethylenediaminetetraacetic acid disodium salt
$AgNO_3$	:	Silver Nitrate
PBQ / p-BQ	:	p-Benzoquinone
BuOH	:	tert-Butanol

REFERENCES

Abejón, R., De Cazes, M., Belleville, M. P., & Sanchez-Marcano, J. (2015). Large-scale enzymatic membrane reactors for tetracycline degradation in WWTP effluents. *Water Research, 73*, 118-131.https://doi.org/10.1016/j.watres.2015.01.012

Abinaya, K., Rajkishore, S. K., Lakshmanan, A., Anandham, R., Dhananchezhiyan, P., & Praghadeesh, M. (2021). Synthesis and characterisation of carbon dots from coconut shell by optimising the hydrothermal carbonisation process. *Journal of Applied and Natural Science, 13*(4), 1151-1157.https://doi.org/10.31018/jans.v13i4.2916

Akbar, K., Moretti, E., & Vomiero, A. (2021). Carbon dots for photocatalytic degradation of aqueous pollutants: Recent advancements. *Advanced Optical Materials, 9*(17), Article 2100532.https://doi.org/10.1002/adom.202100532

Ayu, D. G., Gea, S., Andriayani, N., Telaumbanua, D. J., Piliang, A. F. R., Harahap, M., Yen, Z., Goei, R., & Tok, A. I. Y. (2023). Photocatalytic degradation of methylene blue using N-doped ZnO/carbon dot (N-ZnO/CD) nanocomposites derived from organic soybean. *ACS Omega, 8*(17), 14965-14984.https://doi.org/10.1021/acsomega.2c07546

Azimi, S., & Nezamzadeh-Ejhieh, A. (2015). Enhanced activity of clinoptilolite-supported hybridised PbS–CdS semiconductors for the photocatalytic degradation of a mixture of tetracycline and cephalexin aqueous solution. *Journal of Molecular Catalysis A: Chemical, 408*, 152-160.https://doi.org/10.1016/j.molcata.2015.07.017

Bai, X., Chen, W., Wang, B., Sun, T., Wu, B., & Wang, Y. (2022). Photocatalytic degradation of some typical antibiotics: Recent advances and future outlooks. *International Journal of Molecular Sciences, 23*(15), Article 8130.https://doi.org/10.3390/ijms23158130

Chong, M. N., Jin, B., Chow, C. W. K., & Saint, C. (2010). Recent developments in photocatalytic water treatment technology: A review. *Water Research, 44*(10), 2997-3027.https://doi.org/10.1016/j.watres.2010.02.039

Cui, L., Ren, X., Sun, M., Liu, H., & Xia, L. (2021). Carbon dots: Synthesis, properties and applications. *Nanomaterials, 11*(12), Article 3419. https://doi.org/10.3390/nano11123419

Egorova, M. N., Tomskaya, A. E., Kapitonov, A. N., Alekseev, A. A., & Smagulova, S. A. (2018). Hydrothermal synthesis of luminescent carbon dots from glucose and birch bark soot. *Journal of Structural Chemistry, 59*(4), 780-785. https://doi.org/10.1134/S0022476618040054

Fan, J., Li, D., & Wang, X. (2016). Effect of modified graphene quantum dots on photocatalytic degradation property. *Diamond and Related Materials, 69*, 81-85. https://doi.org/10.1016/j.diamond.2016.07.008

Fawaz, W., Hasian, J., & Alghoraibi, I. (2023). Synthesis and physicochemical characterisation of carbon quantum dots produced from folic acid. *Scientific Reports, 13*(1), Article 46084. https://doi.org/10.1038/s41598-023-46084-1

Gopal, G., Alex, S. A., Chandrasekaran, N., & Mukherjee, A. (2020). A review on tetracycline removal from aqueous systems by advanced treatment techniques. *RSC Advances, 10*(45), 27081-27095.https://doi.org/10.1039/D0RA04264A

Hak, C. H., Leong, K. H., Chin, Y. H., Saravanan, P., Tan, S. T., Chong, W. C., & Sim, L. C. (2020). Water hyacinth derived carbon quantum dots and g-C_3N_4 composites for sunlight driven photodegradation of 2,4-dichlorophenol. *SN Applied Sciences, 2*(6), Article 2840.https://doi.org/10.1007/s42452-020-2840-y

Hasanudin, Asri, W. R., Muthiarani, T. E., Bahrain, D., & Hadiah, F. (2023). Esterification of free fatty acid from sludge palm oil using zeolite sulfonated carbon from sugar cane catalysts. *Journal of Oil Palm Research, 35*(2), 268-280. https://doi.org/10.21894/jopr.2022.0051

Hu, C., Wang, K.-H., Chen, Y.-Y., Maniwa, M., Lin, K.-Y. A., Kawai, T., & Chen, W. (2022). Detection of Fe^{3+} and Hg^{2+} ions through photoluminescence quenching of carbon dots derived from urea and bitter tea oil residue. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 272*, 120963. https://doi.org/10.1016/j.saa.2022.120963

Hu, X., Zhao, H., Liang, Y., Chen, F., Li, J., & Chen, R. (2021). Broad-spectrum response NCQDs/$Bi_2O_2CO_3$ heterojunction nanosheets for ciprofloxacin photodegradation: Unravelling the unique roles of NCQDs upon different light irradiation. *Chemosphere, 264*, 128434.https://doi.org/10.1016/j.chemosphere.2020.128434

Jamaludin, N., Rashid, S. A., & Tan, T. (2018). Natural biomass as carbon sources for the synthesis of photoluminescent carbon dots. In *Synthesis, technology and applications of carbon nanomaterials* (pp. 109-134). Elsevier. https://doi.org/10.1016/B978-0-12-815757-2.00005-X

Jumardin, J., Maddu, A., Santoso, K., & Isnaeni, I. (2021). Synthesis of carbon dots (CDs) and determination of optical gap energy with Tauc plot method. *Jambura Physics Journal, 3*(2), 73-86. https://doi.org/10.34312/jpj.v3i2.11235

Jung, H., Sapner, V. S., Adhikari, A., Sathe, B. R., & Patel, R. (2022). Recent progress on carbon quantum dots based photocatalysis. *Frontiers in Chemistry, 10*, Article 881495. https://doi.org/10.3389/fchem.2022.881495

Jusuf, B. N., Sambudi, N. S., Isnaeni, & Samsuri, S. (2018). Microwave-assisted synthesis of carbon dots from eggshell membrane ashes using sodium hydroxide and their usage for degradation of methylene blue. *Journal of Environmental Chemical Engineering, 6*(6), 7426-7433. https://doi.org/10.1016/j.jece.2018.10.032

Khairol Anuar, N. K., Tan, H. L., Lim, Y. P., So'aib, M. S., & Abu Bakar, N. F. (2021). A review on multifunctional carbon dots synthesised from biomass waste: Design/fabrication, characterisation and applications. *Frontiers in Energy Research, 9*, Article 626549.https://doi.org/10.3389/fenrg.2021.626549

Khan, M. H., Bae, H., & Jung, J. Y. (2010). Tetracycline degradation by ozonation in the aqueous phase: Proposed degradation intermediates and pathway. *Journal of Hazardous Materials, 181*(1-3), 659-665. https://doi.org/10.1016/j.jhazmat.2010.05.063

Kumar, S., Ojha, A. K., Ahmed, B., Kumar, A., Das, J., & Materny, A. (2017). Tunable (violet to green) emission by high-yield graphene quantum dots and exploiting its unique properties towards sunlight-driven photocatalysis and supercapacitor electrode materials. *Materials Today Communications, 11*, 76-86. https://doi.org/10.1016/j.mtcomm.2017.02.009

Li, Y., Zhang, B. P., Zhao, J. X., Ge, Z. H., Zhao, X. K., & Zou, L. (2013). ZnO/carbon quantum dots heterostructure with enhanced photocatalytic properties. *Applied Surface Science, 279*, 367-373.https://doi.org/10.1016/j.apsusc.2013.04.114

Liu, J., Li, R., & Yang, B. (2020). Carbon dots: A new type of carbon-based nanomaterial with wide applications. *ACS Central Science, 6*(12), 2179-2195. https://doi.org/10.1021/acscentsci.0c01306

Liu, Y. Y., Yu, N. Y., Fang, W.-D., Tan, Q.-G., Ji, R., Yang, L.-Y., Wei, S., Zhang, X.-W., & Miao, A.-J. (2021). Photodegradation of carbon dots causes cytotoxicity. *Nature Communications, 12*(1), Article 812. https://doi.org/10.1038/s41467-021-21080-z

Louros, V. L., Ferreira, L. M., Silva, V. G., Silva, C. P., Martins, M. A., Otero, M., Esteves, V. I., & Lima, D. L. D. (2021). Photodegradation of aquaculture antibiotics using carbon dots–TiO_2 nanocomposites. *Toxics, 9*, Article 267. https://doi.org/10.3390/toxics9120330

Mahat, N. A., & Shamsudin, S. A. (2020). Blue luminescence carbon quantum dots derived from oil palm empty fruit bunch biomass. *IOP Conference Series: Materials Science and Engineering, 736*(5), Article 052001. https://doi.org/10.1088/1757-899X/736/5/052001

Mintz, K. J., Zhou, Y., & Leblanc, R. M. (2019). Recent development of carbon quantum dots regarding their optical properties, photoluminescence mechanism, and core structure. *Nanoscale, 11*(11), 4634-4652. https://doi.org/10.1039/C8NR10059D

Mukherjee, I., Cilamkoti, V., & Dutta, R. K. (2021). Sunlight-driven photocatalytic degradation of ciprofloxacin by carbon dots embedded in ZnO nanostructures. *ACS Applied Nano Materials, 4*(8), 7686-7697.https://doi.org/10.1021/acsanm.1c00883

Nair, S. S. P., Kottam, N., & Kumar S. G. P. (2020). Green synthesised luminescent carbon nanodots for the sensing application of Fe^{3+} ions. *Journal of Fluorescence, 30*(2), 357-363.https://doi.org/10.1007/s10895-020-02505-2 Niu, J., Ding, S., Zhang, L., Zhao, J., & Feng, C. (2013). Visible-light-mediated Sr-Bi_2O_3 photocatalysis of tetracycline: Kinetics, mechanisms and toxicity assessment. *Chemosphere, 93*(1), 1-8. https://doi.org/10.1016/j.chemosphere.2013.04.043

Pelaez, M., Nolan, N. T., Pillai, S. C., Seery, M. K., Falaras, P., Kontos, A. G., Dunlop, P. S. M., Hamilton, J. W. J., Byrne, J. A., O'Shea, K., Entezari, M. H., & Dionysiou, D. D. (2012). A review on the visible light active titanium dioxide photocatalysts for environmental applications. *Applied Catalysis B: Environmental, 125*, 331-349. https://doi.org/10.1016/j.apcatb.2012.05.036

Saadati, F., Keramati, N., & Ghazi, M. M. (2016). Influence of parameters on the photocatalytic degradation of tetracycline in wastewater: A review. *Critical Reviews in Environmental Science and Technology, 46*(8), 757-782. https://doi.org/10.1080/10643389.2016.1159093

Safari, G. H., Hoseini, M., Seyedsalehi, M., Kamani, H., Jaafari, J., & Mahvi, A. H. (2015). Photocatalytic degradation of tetracycline using nanosized titanium dioxide in aqueous solution. *International Journal of Environmental Science and Technology, 12*(2), 603-616. https://doi.org/10.1007/s13762-014-0706-9

Sahu, V., & Sahoo, S. K. (2024). Biogenic synthesis of carbon dots with inbuilt biological activity. *Next Nanotechnology, 5*, 100034. https://doi.org/10.1016/j.nxnano.2023.100034

Sawalha, S., Azzam, H., Bin Ali, R., Dweikat, H., & Anaya, K. (2021, October 12-14). Photodegradation of methylene blue by carbon nanodots synthesised from olive solid wastes [Conference paper]. *Proceedings of the 9th Jordan International Chemical Engineering Conference (JICHEC9), Jordan*.

Sharma, M., Mandal, M. K., Pandey, S., Kumar, R., & Dubey, K. K. (2022). Visible-light-driven photocatalytic degradation of tetracycline using heterostructured Cu_2O–TiO_2 nanotubes: Kinetics and toxicity evaluation of degraded products on cell lines. *ACS Omega, 7*(37), 33572-33586. https://doi.org/10.1021/acsomega.2c04576

Sharma, N., Sharma, I., & Bera, M. K. (2022). Microwave-assisted green synthesis of carbon quantum dots derived from *Calotropis gigantea* as a fluorescent probe for bioimaging. *Journal of Fluorescence, 32*(3), 1039-1049. https://doi.org/10.1007/s10895-022-02923-4

Sutanto, H., Alkian, I., Romanda, N., Lewa, I. W. L., Marhaendrajaya, I., & Triadyaksa, P. (2020). High green-emission carbon dots and its optical properties: Microwave power effect. *AIP Advances, 10*(5), Article 055013.https://doi.org/10.1063/5.0004595

Syahin, M., Zamri, F. A., & Sapawe, N. (2019). Kinetic study on photocatalytic degradation of phenol using green electrosynthesised TiO_2 nanoparticles. *Materials Today: Proceedings, 19*, 2214-2219.https://doi.org/10.1016/j.matpr.2019.11.131

Thiang, E. L., Lee, C. W., Takada, H., Seki, K., Takei, A., Suzuki, S., Wang, A., & Bong, C. W. (2021). Antibiotic residues from aquaculture farms and their ecological risks in Southeast Asia: A case study from Malaysia. *Ecosystem Health and Sustainability, 7*(1), Article 1926337. https://doi.org/10.1080/20964129.2021.1926337

Trenczek-Zając, A., Synowiec, M., Zakrzewska, K., Zazakowny, K., Kowalski, K., Dziedzic, A., & Radecka, M. (2022). Scavenger-supported photocatalytic evidence of an extended type I electronic structure of the TiO_2@Fe_2O_3 interface. *ACS Applied Materials & Interfaces, 14*(33), 38255-38269. https://doi.org/10.1021/acsami.2c06404

Ţucureanu, V., Matei, A., & Avram, A. M. (2016). FTIR spectroscopy for carbon family study. *Critical Reviews in Analytical Chemistry, 46*(6), 502-520. https://doi.org/10.1080/10408347.2016.1157013

Wang, W., Lemaire, R., Bensakhria, A., & Luart, D. (2022). Review on the catalytic effects of alkali and alkaline earth metals (AAEMs), including sodium, potassium, calcium and magnesium, on the pyrolysis of lignocellulosic biomass and on the co-pyrolysis of coal with biomass. *Journal of Analytical and Applied Pyrolysis, 163*, 105479. https://doi.org/10.1016/j.jaap.2022.105479

World Health Organisation. (2021). *Antimicrobial resistance*. https://www.who.int/groups/framework-development-stewardship-AMR

Wu, S., Hu, H., Lin, Y., Zhang, J., & Hu, Y. H. (2020a). Visible light photocatalytic degradation of tetracycline over TiO_2. *Chemical Engineering Journal, 382*, 122842. https://doi.org/10.1016/j.cej.2019.122842

Wu, S., Li, X., Tian, Y., Lin, Y., & Hu, Y. H. (2021b). Excellent photocatalytic degradation of tetracycline over black anatase-TiO_2 under visible light. *Chemical Engineering Journal, 406*, 126747.https://doi.org/10.1016/j.cej.2020.126747

Xu, J. J., Lu, Y. N., Tao, F. F., Liang, P. F., & Zhang, P. A. (2023). ZnO nanoparticles modified by carbon quantum dots for the photocatalytic removal of synthetic pigment pollutants. *ACS Omega, 8*(8), 7845-7857.https://doi.org/10.1021/acsomega.2c07591

Yan, M., Yan, Y., Wu, Y., Shi, W., & Hua, Y. (2015). Microwave-assisted synthesis of monoclinic-tetragonal $BiVO_4$ heterojunctions with enhanced visible-light-driven photocatalytic degradation of tetracycline. *RSC Advances, 5*(110), 90255-90264. https://doi.org/10.1039/C5RA13684A

Zhong, Q., Lin, Q., Huang, R., Fu, H., Zhang, X., Luo, H., & Xiao, R. (2020). Oxidative degradation of tetracycline using persulfate activated by N and Cu codoped biochar. *Chemical Engineering Journal, 380*, 122608.https://doi.org/10.1016/j.cej.2019.122608

Zhou, Y., Mintz, K. J., Oztan, C. Y., Hettiarachchi, S. D., Peng, Z., Seven, E. S., Liyanage, P. Y., De La Torre, S., Celik, E., & Leblanc, R. M. (2018). Embedding carbon dots in superabsorbent polymers for additive manufacturing. *Polymers, 10*(8), Article 921. https://doi.org/10.3390/polym10080921

Pertanika J. Sci. & Technol. 34 (2): 845 - 860 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Effect of β-cyclodextrin Concentration on the Antioxidant Stability in an Avocado Oil Emulsion-based System

Carissa Eilin Ardana[1], Anggun Puspitarini Siswanto[1], Mohamad Endy Yulianto[1], Arva Rafif Adi Prayoga[1], Elsan Febiyanti[2], and Hermawan Dwi Ariyanto[1]*

[1]Department of Industrial Technology, Vocational School, Universitas Diponegoro, Jalan Gubernur Mochtar, 50275 Semarang, Indonesia
[2]Graduate School of Chemical Engineering, Faculty of Engineering, Universitas Diponegoro, Jalan Prof. Jacub Rais, 50275 Semarang, Indonesia

ABSTRACT

Avocado oil is recognised as a valuable source of vegetable oil, rich in antioxidants and high levels of fatty acids, particularly oleic acid. Oleic acid is commonly utilised as a moisturising agent in cosmetics and as a natural alternative to synthetic chemicals in emulsion-based lotions. This study aimed to investigate the effect of β-cyclodextrin (β-CD) on the antioxidant stability within an avocado oil emulsion system. The emulsion was prepared using a combination of emulsifiers (Tween 80 and Span 80) with a specific HLB ratio and homogenised with a T 25 digital Ultra-Turrax at 15,000 rpm for 10 minutes, including 30-second intervals between homogenisation cycles. The stability of the emulsion was evaluated over 30 days at a constant temperature of 37°C by analysing oil droplet size, viscosity, creaming index, and antioxidant activity. The findings revealed that the stability of the antioxidant in the emulsion system was influenced by the mixed emulsifier's HLB value and the β-CD concentration. The most stable emulsion was achieved with an HLB value of 10.7 and a 1% β-CD concentration. This stabilised emulsion demonstrated significant antioxidant activity, with an IC_{50} value of 44.6 ppm, and showed no signs of phase separation upon visual and microscopic analysis.

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 08 June 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.11

E-mail addresses:
risaadra@gmail.com (Carissa Eilin Ardana)
anggun.siswanto@live.undip.ac.id (Anggun Puspitarini Siswanto)
endy_y@yahoo.com (Mohamad Endy Yulianto)
arvarafif@gmail.com (Arva Rafif Adi Prayoga)
elsanfebiyanti@gmail.com (Elsan Febiyanti)
hd.ariyanto@live.undip.ac.id (Hermawan Dwi Ariyanto)
* Corresponding author

Keywords: Avocado oil, β-cyclodextrin, emulsion, stability, surfactant

INTRODUCTION

The avocado tree, *Persea americana* Mill., is a tropical plant known globally for its fruit, which is often used as a culinary component, notably in Indonesia (Orhevba & Jinadu, 2011). The avocado fruit is not

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

only consumed as food, but it can also be used in several different preparations. It is well acknowledged by the community for its high content of antioxidants and nutrients, particularly fat (oil), which amounts to 9.8 g per 100 g of fruit flesh (Maitera et al., 2014). In addition, avocado fruit also has a significant number of fatty acids that may be used as a vegetable oil source. The fatty acid composition of this oil includes oleic acid (75.16%), palmitic acid (14.72%), palmitoleic acid (0.19%), stearic acid (0.87%), linolenic acid (3.37%), and linoleic acid (3.57%) (Alkhalaf et al., 2019). Oleic acid, mostly present in avocados, is classified as an unsaturated fatty acid. This fatty acid is exogenous, meaning it is obtained from external sources and is often not endogenously produced by the body. Consequently, oleic acid is often included as a hydrating component in cosmetic products. One of them is used as a chemical alternative in the production of lotions to promote the recycling of natural materials for skincare.

Lotion is a cosmetic product that consists of a blend of water, emollients, emulsifiers, active agents, thickeners, preservatives, and perfumes. The challenge in creating an emulsion lies in the disparity between the liquid and oil phases during the manufacturing process, which cannot be easily blended. To overcome this, the appropriate emulsifier is required to facilitate the formation of the emulsion. Hence, to inhibit the segregation of the oil phase and water phase, cyclodextrin is used as an emulsifying agent. Cyclodextrin has been extensively used in many sectors, including its application in enhancing the stability of perfume oil by protecting it from oxidation and unpleasant odour (Inoue et al., 2010; Ma et al., 2023; Paramita et al., 2022; Shimada et al., 1992). Furthermore, cyclodextrin has the capacity to enhance the stability of emulsions. Cyclodextrin is a cyclic oligosaccharide molecule that has a toroidal shape with a hydrophobic inner surface and a hydrophilic outer section (Xiao et al., 2021). Nevertheless, Hou et al. (2021) indicated that the emulsification efficacy of β-cyclodextrin (β-CD) was insufficient to impede phase separation. Therefore, it is crucial to use supplementary emulsifiers to create a stable emulsion (Hou et al., 2021; Huang et al., 2021; Putri & Ariyanto, 2022; Sahafi et al., 2021). To prevent phase separation in β-CD emulsification, one method is to use Tween 80 and Span 80 as emulsifying agents in the process of emulsion synthesis. Lv et al. (2014) have conducted a study on the use of Tween 80 and Span 80 as emulsifiers in a sunscreen composition including ethanol extracts of avocado leaves. The study findings indicated that the cream preparation's physical qualities, including elasticity, adhesiveness, viscosity, and pH, were influenced by Tween 80 and Span 80. Increasing the amount of Tween 80 leads to improved elasticities and pH levels (Faghmous et al., 2021). Furthermore, the use of Tween 80 and Span 80 in the sunscreen formulation successfully fulfils the desirable physical attributes of a high-quality cream, as determined by the Simplex Lattice Design technique (SLD) (Ghorbanzadeh et al., 2019). Thus, there has been limited research on the efficacy of using cyclodextrin as an emulsifier in emulsion systems.

Therefore, the objective of this research was to examine the impact of β-CD on the durability of the antioxidant in the emulsion system of avocado oil. An emulsion of avocado oil was created by combining a mixture of emulsifiers (Tween 80 and Span 80) in a certain ratio of hydrophilic-lipophilic balance (HLB). The mixture was then homogenised using a T 25 digital Ultra-Turrax at a speed of 15,000 rpm for 10 min, with a 30-second break between each homogenisation cycle. β-CD is a cyclodextrin with a large cavity that can accommodate many molecules. It is more stable and can be quickly isolated from the product without the need for organic solvents due to its extremely low solubility. Meanwhile, Tween 80 and Span 80 are emulsifying agents with non-ionic properties. Each agent has distinct characteristics. Specifically, Tween 80 has a high HLB value and exhibits hydrophobic tendencies, while Span 80 has a low HLB value and exhibits hydrophilic tendencies. Additionally, vitamin E (α-tocopherol acetate) is an antioxidant that plays a crucial role in safeguarding the skin against damage caused by free radicals. The effect of adding β-CD and vitamin E (α-tocopherol acetate) to this emulsion preparation will be determined to assess its impact on the outcomes of the emulsion system.

MATERIALS AND METHODS

Materials

Avocado oil was obtained from PT Darjeeling Semrani Aroma (Bandung, West Java, Indonesia). Beta-cyclodextrin (β-CD) was gifted from Cyclochem Co., Ltd. (Japan). Tween 80 and Span 80 were obtained from CV Indrasari (Semarang, Central Java, Indonesia). α-tocopherol acetate was obtained from Sentana Sempurna (Makassar, South Sulawesi, Indonesia). All chemicals used in this study were of analytical grade.

Methods

Preparation for Avocado Oil Emulsion

The aqueous phase solution was created by dissolving the combined emulsifier with the appropriate hydrophilic-lipophilic balance (HLB) ratio in 200 mL of distilled water at a temperature of 60°C. 140 mL of avocado oil was mixed using a high-speed magnetic stirrer. An oil-in-water emulsion was created by adding the oil phase to water while continuously swirling at a high speed. To achieve emulsification, the solution was homogenised using a T 25 digital Ultra-Turrax (IKA Co. Ltd., China) at 15,000 rpm for 10 min, with a 30-second break after every minute of homogenisation (Hou et al., 2021). The formulation of avocado oil in an emulsion-based system with a specific HLB value is shown in Table 1.

Table 1
Formulation of avocado oil in an emulsion-based formulation with a specific HLB value

Composition	HLB Value	Concentration of β-CD
Tween 80 (87 mL); Span 80 (58 mL)	8.5	0 wt.%
Tween 80 (87 mL); Span 80 (58 mL)	8.5	0.5 wt.%
Tween 80 (87 mL); Span 80 (58 mL)	8.5	1 wt.%
Tween 80 (58 mL); Span 80 (87 mL)	10.7	0 wt.%
Tween 80 (58 mL); Span 80 (87 mL)	10.7	0.5 wt.%
Tween 80 (58 mL); Span 80 (87 mL)	10.7	1 wt.%
Tween 80 (116 mL); Span 80 (29 mL)	12.8	0 wt.%
Tween 80 (116 mL); Span 80 (29 mL)	12.8	0.5 wt.%
Tween 80 (116 mL); Span 80 (29 mL)	12.8	1 wt.%

Emulsion Stability

The stability of O/W emulsion was investigated at a constant temperature of 37°C for 30 days. Viscosity, pH, creaming index value, droplet size, antioxidant activity and physical morphology of the emulsion were used to measure the stability of the emulsion system.

Viscosity Measurement

The viscosity of the emulsion was measured using a Brookfield digital viscometer (Model DV-I + Programmable). Viscosity was evaluated using a No. 64 spindle for approximately 80 mL of sample and then expressed in Cp (centipoise) (Ma et al., 2023).

pH Analysis

Analysis of the degree of acid or pH is carried out based on the Indonesian National Standard method or SNI 06-2413-1991 using the Mettler Toledo Sevencompact S210 pH meter to determine the pH of the emulsion preparation, so that the quality standard can be known to the skin.

Creaming Index Measurement

The creaming index of emulsions was examined by conducting a gravity separation test at a temperature of 37°C for a period of 30 days. Exactly 80 mL of each oil-in-water emulsion was deposited in a glass tube and kept at a temperature of 37°C. The creaming index of the emulsion was tested throughout the storage period. 10 mL of emulsion was placed in a centrifuge tube and then subjected to centrifugation at a speed of 4500 revolutions per minute for a duration of 3 min (Putri & Ariyanto, 2022). The cream's volume was measured, and the creaming index % was computed using Equation 1:

$$Creaming\ index = \frac{HS}{HE} \times 100\% \quad [1]$$

where *HS* is the height of the precipitate layer, and *HE* is the total height of the emulsion.

Droplet Size Measurement

The particle size of the oil droplet was observed using an Olympus optical microscope and then measured using ImageJ software. The droplet diameter data were analysed with Origin software. Where a drop of emulsion was placed on a prepared glass slide and examined under a light microscope with 100x magnification (Xiao et al., 2021).

Antioxidant Activity Measurement

Antioxidant activity can be determined using the DPPH method. A 2 mg quantity of 2,2-diphenyl-1-picrylhydrazyl (DPPH) powder was dissolved in methanol in a 100 mL volumetric flask until the solution reached the limit mark and was agitated to achieve homogeneity. This resulted in a DPPH solution with a concentration of 0.05 mM. The blank solution was prepared by adding 2 mL of methanol to 2 mL of a 0.05 mM DPPH solution and then covering it with aluminium foil. Next, the sample should be agitated and placed in an enclosed space without light for a duration of 30 min. Subsequently, use a UV-Vis spectrophotometer to quantify the solution at a wavelength of 517 nm. By mixing 5 mg of emulsion preparation with 10 mL of methanol until it became uniform, a concentration of 500 ppm was achieved. The solution was prepared using varying concentrations of 10, 20, and 30 parts per million (ppm) by transferring 0.1 mL, 0.2 mL, and 0.3 mL, respectively, into a 10 mL volumetric flask, and then filling the remaining volume with methanol to get a total volume of 10 mL. For each test, 2 mL of the solution was pipetted and then mixed with an equal amount of DPPH. The resulting mixture was then transferred to a test tube and homogenised. The solution was subjected to incubation at a temperature of 37°C for a duration of 30 min, and the measurement of absorbance was performed at a specific wavelength of 517 nm (Hou et al., 2021). The metric used to quantify antioxidant activity was the effective concentration (EC_{50}) or inhibitory concentration (IC_{50}).

$$Q = \frac{A_1 - A_2}{A_1} \times 100\% \quad [2]$$

where Q is the percentage of inhibition, A_1 is the absorbance of the control, and A_2 is the absorbance of the sample.

Characterisation of Emulsion Morphology

The morphological characterisation of the emulsion was assessed using two methods: eye inspection and microscopic examination. The emulsion's phase separation was visually seen using a high-speed digital camera with a resolution of 64 megapixels, namely the Samsung A52S 5G model. An Olympus optical microscope was used for the purpose of microscopy observation. An emulsion droplet was applied onto a prepared glass slide and viewed using a light microscope at a magnification of 100x.

RESULTS AND DISCUSSION

Effect of β-CD Concentration and HLB Value on the Viscosity of Emulsion Preparation

Viscosity and emulsion stability are tightly interconnected. Table 2 shows the viscosity measurements obtained from a Brookfield Model DV-I Viscometer, using spindle no. 64 and a speed of 4 rpm. The measurements were taken on day 0 and day 30. The investigation revealed that on the first day of preparation (day 0), the viscosity of each sample rose as the β-CD content increased. The sample without β-CD treatment had the lowest recorded value. Following a storage period of 30 days, the viscosity of each sample exhibited a decline, even though each sample underwent a distinct treatment. As a result, water was absorbed by each sample during the storage phase in the incubator. The use of 1% β-CD concentration in the samples showed a beneficial impact on enhancing viscosity. Moreover, it is demonstrated that an HLB of 10.7 is also efficient in preserving the stability of the emulsion. β-CD has a molecular structure that exhibits hydrophilicity on its outer surface and hydrophobicity inside its cavity (Ciobanu et al., 2013). This characteristic makes β-CD suitable for use as an emulsifier in the formulation of emulsions.

Table 2
Viscosity of emulsion at various concentrations of β-CD; 0% β-CD; 0.5% β-CD; 1% β-CD

HLB Value	Concentration of β-CD	Viscosity	
		0-day Storage (cP)	30-day Storage (cP)
8.5	0 wt.%	3050 ± 0.100	2150 ± 0.076
	0.5 wt.%	3200 ± 0.100	2816.7 ± 0.101
	1 wt.%	3350 ± 0.050	2700 ± 0.104
10.7	0 wt.%	26217 ± 0.882	25566.7 ± 1.540
	0.5 wt.%	28980 ± 0.053	25900 ± 1.532
	1 wt.%	33800 ± 0.557	31600 ± 0.671
12.8	0 wt.%	3050 ± 0.161	2966.7 ± 0.060
	0.5 wt.%	3550 ± 0.180	3316.7 ± 0.159
	1 wt.%	4400 ± 0.132	4166.7 ± 0.130

According to research, the addition of β-CD is expected to increase the viscosity of the emulsion preparation (Kibici & Kahveci, 2019).

The results indicate that among all the samples, HLB 10.7 achieves the highest viscosity, which is influenced by the variation in emulsifier composition. According to the table, the HLB 10.7 sample, containing 40% Tween 80 and 60% Span 80, shows a higher viscosity compared to the HLB 8.5 sample (60% Tween 80 and 40% Span 80) and the HLB 12.8 sample (80% Tween 80 and 20% Span 80). The lowest viscosity is observed in the sample with an HLB of 8.5. This demonstrates that the HLB value significantly affects the emulsion's stability during storage. The findings suggest that increasing the proportion of Span 80 relative to Tween 80 can enhance the stability of the emulsion (Suardana et al., 2020). Increasing the concentration of surfactants typically results in a higher apparent viscosity of emulsions. This is because surfactant molecules form structured films at the oil-water interface, which contributes to an increase in viscosity. Sun et al. (2017) also found that increasing the organic/inorganic alkali surfactants raised the apparent viscosity of the O/W emulsion. This increase in apparent viscosity was attributed to the enhanced ionisation of the interfacial active components and the compression of the diffused double layer, with the competition between these factors determining whether the apparent viscosity increased or decreased.

Effect of β-CD Concentration and HLB Value on pH of Emulsion Preparation

The pH value of emulsions may be examined to determine their acidity. This analysis was conducted at regular intervals throughout a 30-day storage period. The findings clearly indicated that the pH values obtained from each sample fall within the acceptable range of 7.08-7.62, as specified by the quality criteria for emulsion, which are outlined in SNI 16-4399-1999. This is due to the component ingredients in emulsion treatments often having a neutral pH range of 6-7. The pH of the emulsion preparation's component ingredients comprises Tween 80 with a pH range of 6-8 and Span 80 with a pH below 8 (Jiang et al., 2024; Lv et al., 2014). The pH of avocado oil is around 7.5 (Flores et al., 2019; Krumreich et al., 2018), whereas β-CD has a pH ranging from 5 to 8 (Paramita et al., 2022). Distilled water has a pH between 6.7 and 7 (Khorolskyi & Kryvoruchko, 2021).

Figure 1 shows the decline in acidity, as shown by the pH values of all samples, after a 30-day storage period. This decline suggested that the emulsion preparation composition underwent structural deterioration (Singh et al., 2019). Increasing the concentration of β-CD may lead to a reduction in the acidity level of the emulsion preparation (Inoue et al., 2010). The results from each sample indicated that the fluctuation of HLB values in emulsion preparations does not have a significant impact on the acidity evaluated by the pH value.

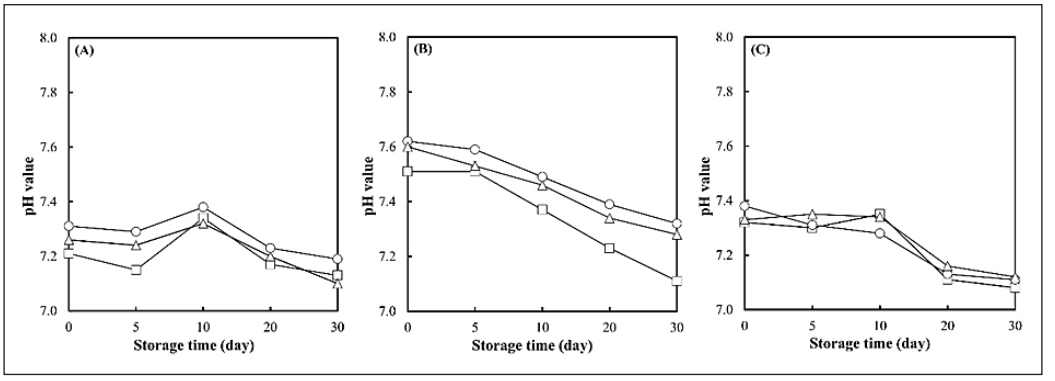


Figure 1. Effect of HLB values on the pH emulsion stability: (a) HLB 8.5; (b) HLB 10.7; (c) HLB 12.8 at different β-CD concentrations: (□) 0 wt.% β-CD; (○) 0.5 wt.% β-CD; (Δ) 1 wt.% β-CD

Effect of β-CD Concentration and HLB value on Creaming Index of Emulsion Preparation

The creaming index value is a measure of the stability of an emulsion, indicating its resistance to separating into distinct layers and its capacity to bind with other particles. The creaming index value of all samples was assessed after 30 days of storage. The findings indicated that the sample with the greatest creaming index percentage was the sample with a hydrophilic-lipophilic balance (HLB) of 8.5, consisting of 60% Tween 80 and 40% Span 80. This was followed by the sample with an HLB of 12.8, including 80% Tween 80 and 20% Span 80. The creaming index suggests that the emulsion preparation is rather unstable. It is proven that the higher the percentage of creaming index in emulsion preparations, the smaller the stability (Whitby et al., 2009).

Emulsions with an HLB of 10.7 consistently exhibited a creaming index value of 0%. This demonstrated that the sample has a high level of stability. The results indicated that there was no difference at the conclusion of the storage period, namely, after 30 days of storage. The addition of a higher concentration of Span 80 leads to the formation of a stable emulsion. This was due to the Span 80 acting as an emulsifier, possessing hydrophobic characteristics that facilitate the binding of the oil phase during the emulsification process (Inoue et al., 2010). An emulsion with high stability will be formed when a larger amount of the oil phase is bound by Span 80. Additionally, the ratio of the cream and clear layer created during the centrifugation process will drop. To reduce the creaming index, the proportion of creaming index created should be decreased (Netto & Jose, 2018; Onuki et al., 2016).

Figure 2 demonstrates that the concentration of β-CD has a favourable impact on the stability of the emulsion. The sample treated with a 1% concentration of β-CD exhibited the lowest creaming index value in comparison to the sample without β-CD and the sample treated with a 0.5% concentration of β-CD. To summarise, the ideal concentration for

emulsion formation that demonstrates stability was 1% β-CD with a hydrophilic-lipophilic balance (HLB) of 10.7. There was no significant change in the creaming index of HLB 8.5 and HLB 12.8 samples over the storage period. This reduction is attributed to the occurrence of same-sex coalescence or colloquy, which leads to the separation or creaming of the oil phase and water phase.

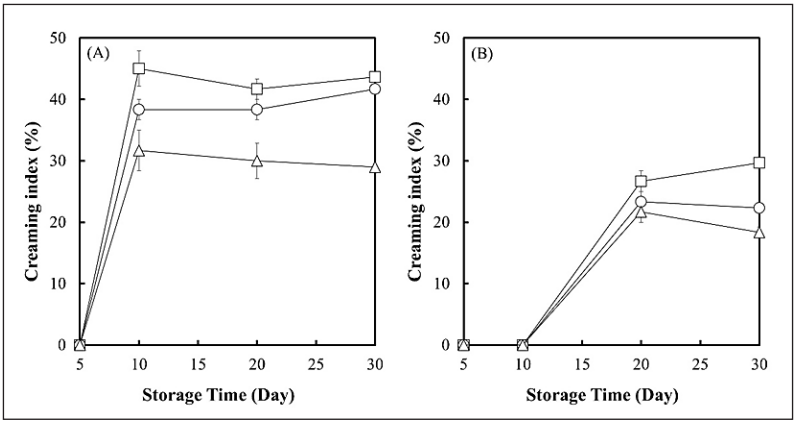


Figure 2. Influence of varying HLB values on the creaming index of the emulsion system: (a) HLB 8.5; (b) HLB 12.8 at different β-CD concentrations: (□) 0 wt.% β-CD; (○) 0.5 wt.% β-CD; (Δ) 1 wt.% β-CD

Effect of β-CD Concentration and HLB Value on Droplet Size of Emulsion Preparation

The droplet size of the emulsions was assessed using ImageJ software. On day 0 of the storage period, each sample was examined under an Olympus optical microscope at 100x magnification (Figure 3). Subsequently, the droplet diameter was evaluated to determine the frequency of visible droplet sizes. Observations reveal that samples with HLB 8.5 (60% Tween 80; 40% Span 80) and HLB 12.8 (80% Tween 80; 20% Span 80) exhibit reduced particle sizes in comparison to samples with HLB 10.7 (40% Tween 80; 60% Span 80). This result is consistent with the study conducted by Hou et al. (2021), which shows that higher concentrations of Tween 80 might result in a reduction in particle size. Nevertheless, the HLB 8.5 sample exhibits a little inaccuracy, whereas the HLB 12.8 sample contradicts the earlier evidence. The HLB 12.8 sample has a greater particle size compared to the sample with HLB 8.5. There was a noticeable impact on the size of particles when the concentration of β-CD was altered. This demonstrated the properties of β-CD inclusions that might hinder the merging and/or enlargement of oil droplets (Hou et al., 2021; Inoue et al., 2010; Shimada et al., 1992).

Latest studies have shown that the use of α-tocopherol acetate in emulsion preparation leads to a reduction in particle size. The reason for this is that α-tocopherol acetate has high oil solubility, which may induce turbulence or heightened particle movement, as well as

reduce oil surface tension (Sahafi et al., 2021). These circumstances facilitate the diffusion of Tween 80 into the oil particles, leading to the formation of very tiny oil particles. This may facilitate the generation of tiny emulsion particles with enhanced stability against separation (Ahmad et al., 2020).

Effect of Concentration and HLB Value on Antioxidant Activity of Emulsion Preparations

The antioxidant activity of each emulsion stored for 30 days was analysed through a UV-VIS spectrophotometer, observing the absorbance value, and then the IC50 was calculated. IC_{50} (Inhibitory Concentration) is the concentration of the sample solution that can reduce the activity of DPPH by 50%, or is a number that shows the concentration of the sample (ppm) which can inhibit the oxidation process by 50%. It is known that the smaller the IC_{50} value, the higher the antioxidant activity (Sahafi et al., 2021).

Figure 4 indicates that the sample treated with a concentration of 1% β-CD had the lowest IC_{50} value at each storage period, reaching its highest point after 30 days of storage. This demonstrated the efficacy of β-CD in preserving the antioxidant content of the emulsion formulation. The given data indicated that the HLB value did not have a noticeable impact on preserving the antioxidant content. Nevertheless, the sample with HLB 10.7 exhibits a minor discrepancy that contradicts the previously established evidence. It is evident that the sample with an HLB of 10.7 exhibits the lowest IC_{50} value in comparison to the samples with HLB values of 8.5 and 12.8. Incorporating a higher amount of Span 80 led to a rise in emulsion viscosity, preventing the formation of creaming index. Moreover, it effectively inhibited the oxidation process in emulsion preparations, therefore preserving the antioxidant content.

Visual Observation of Emulsion Preparation

The process of visually observing the development of an emulsion showed the occurrence of emulsion instability, which is a physical phenomenon that arises during the preparation of the emulsion. The physical emulsion was visually examined and photographed using a Samsung A52S 5G 64 MP camera over a period of 30 days at a constant temperature of 37°C. The findings obtained by observing the storage time are shown in Figure 5.

Figure 5 illustrates that the sample with an HLB of 8.5, without the addition of β-CD concentration, exhibits the most significant phase separation after 30 days of storage. The solution and emulsifying agent do not mix, resulting in an increase in surface tension between the particles. This causes the dispersed phase to separate from the dispersing phase, leading to an unstable emulsion. Observations indicate that there was no occurrence of phase separation in the sample with an HLB value of 10.7 throughout the 30-day storage period. This demonstrated that the ratio of Tween 80 to Span 80 in terms of volume has an impact on the stability of the resulting emulsion.

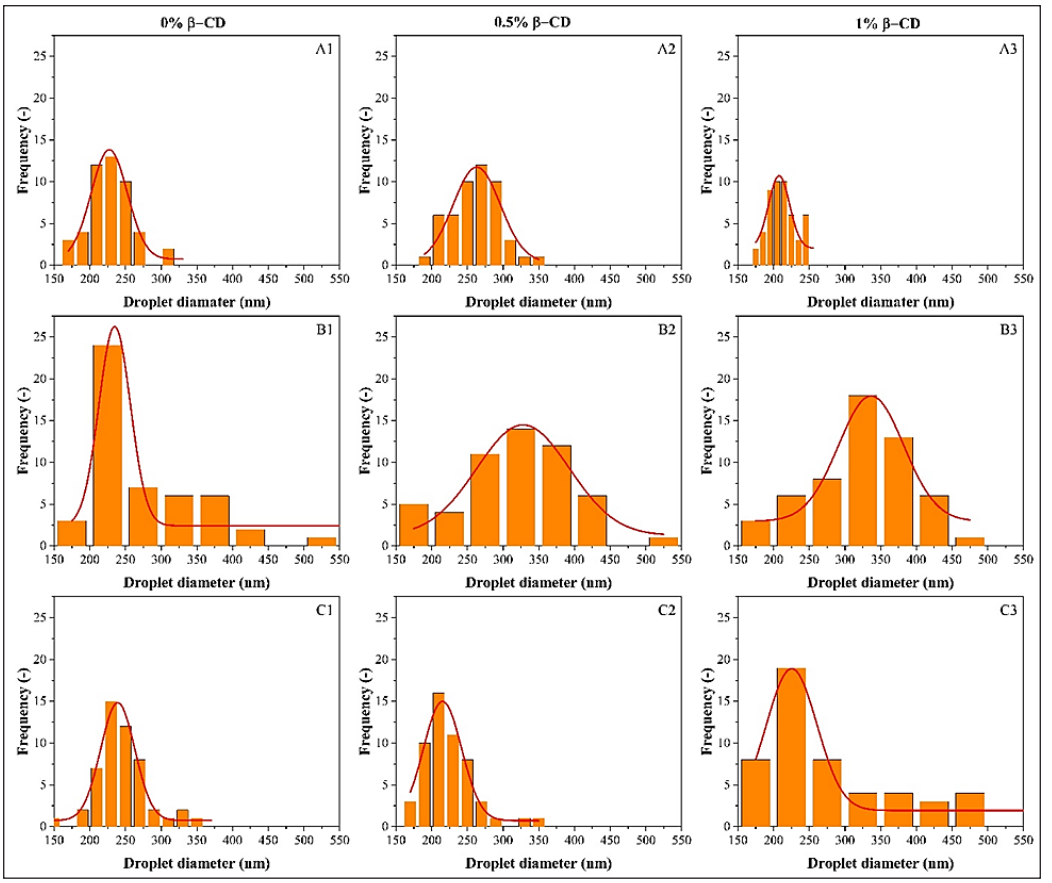


Figure 3. Effect of varying hydrophilic-lipophilic balance (HLB) values on emulsion droplet size: (a) HLB 8.5; (b) HLB 10.7; and (c) HLB 12.8, at different β-cyclodextrin (β-CD) concentrations: (1) 0 wt%; (2) 0.5 wt%; and (3) 1 wt%.

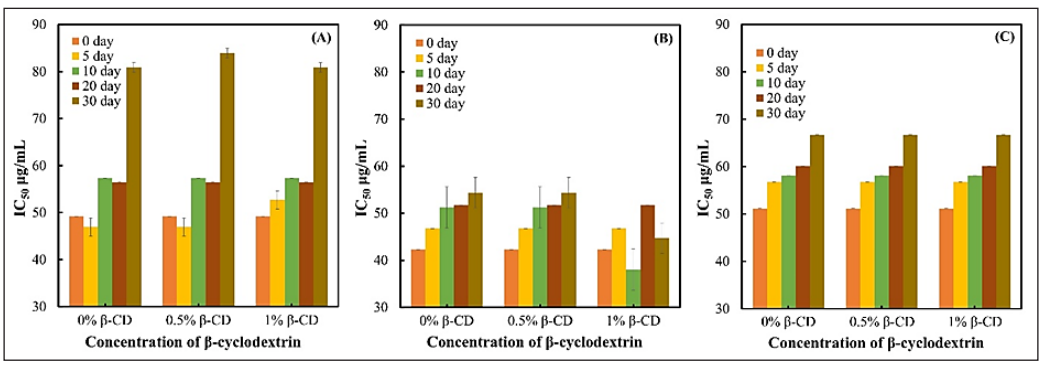


Figure 4. Effect of HLB values on the antioxidant stability of emulsion: (a) HLB 8.5; (b) HLB 10.7; (c) HLB 12

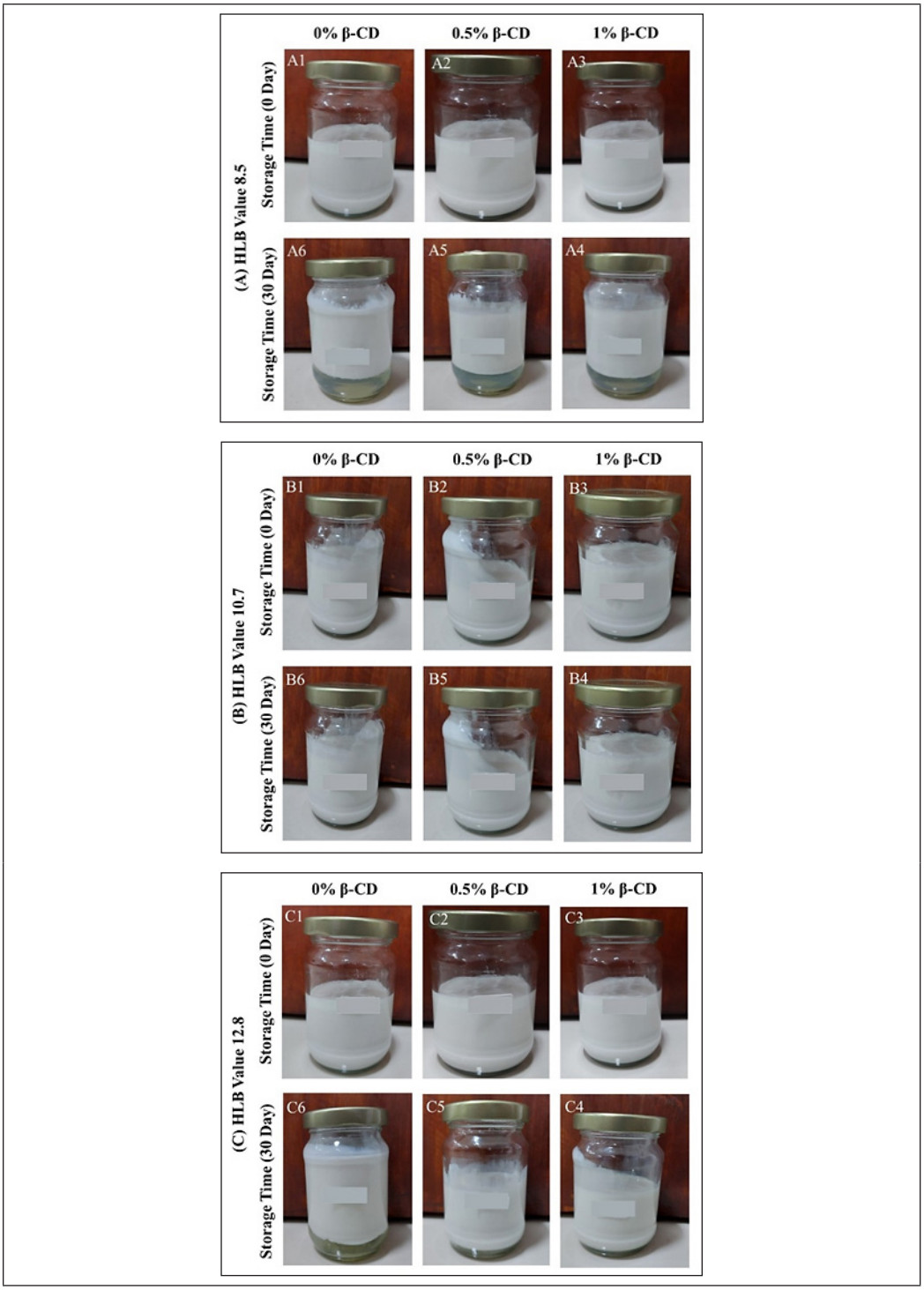


Figure 5. Visual observation of the emulsion system at different HLB values: (a) HLB 8.5; (b) HLB 10.7; (c) HLB 12.8

Microscope Observation of Emulsion Preparation

The emulsion was examined using an Olympus Optical Microscope at a magnification of 100x at 37°C. Microscopic investigations may detect the motion of dispersed phase droplets that lead to creaming and coalescence, which are types of emulsion instability seen after 30 days of storage. Figure 6 shows the findings of microscopic examinations conducted on day 0 and day 30 at regular intervals. Over time, the droplets of the dispersed phase inside the dispersing phase diminish. The phase that was scattered goes in an upward direction and creates a layer called creaming (Putri & Ariyanto, 2022).

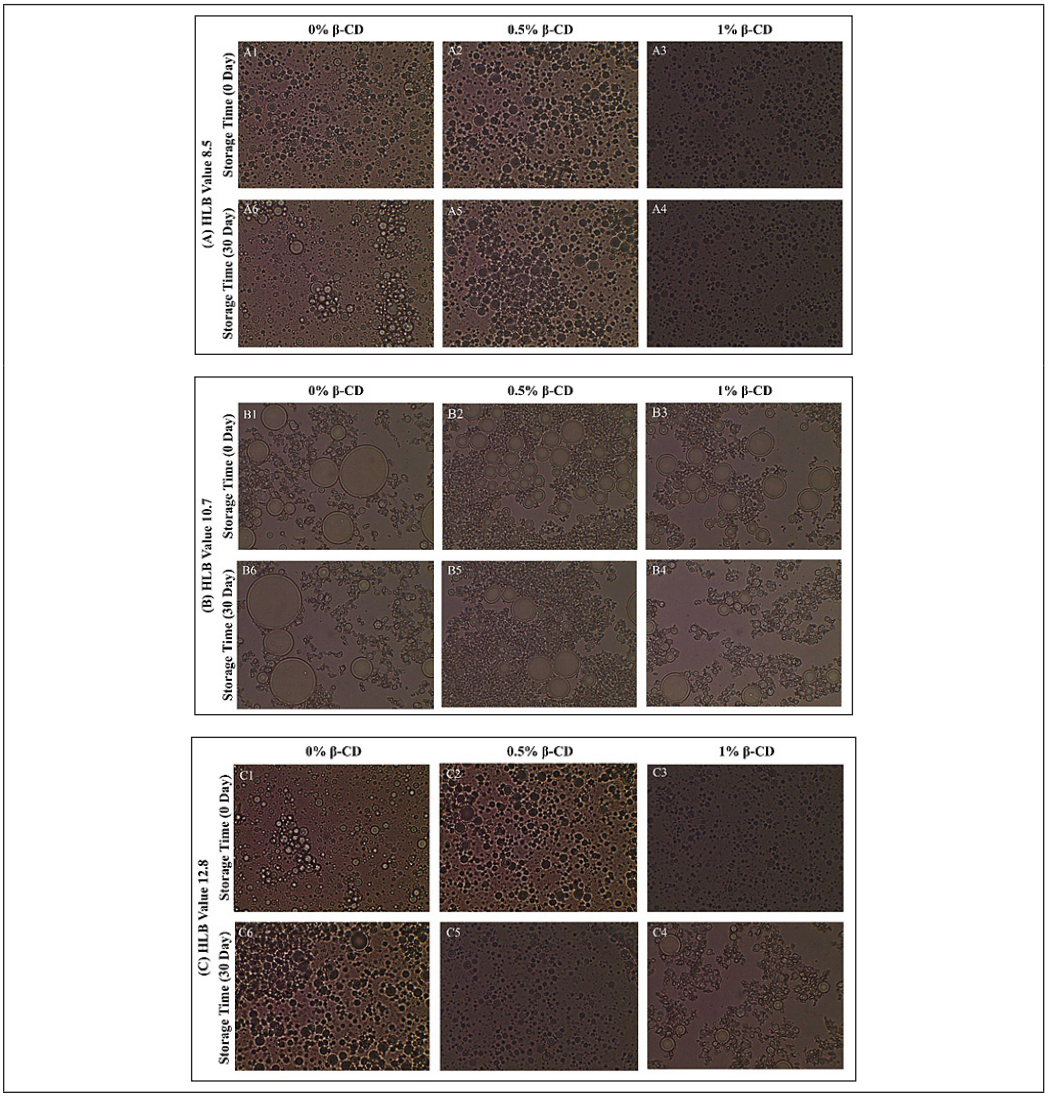


Figure 6. Effect of different HLB values on the morphology properties of emulsion: (a) HLB 8.5; (b) HLB 10.7; (c) HLB 12.8

CONCLUSION

This study investigated the influence of α-tocopherol acetate and β-cyclodextrin (β-CD) on the stability of emulsions in avocado oil lotion. By conducting meticulous experiments, it was shown that the incorporation of these ingredients greatly improved the stability of the emulsion. In this study, the emulsification process of avocado oil lotion preparation with the addition of α-tocopherol acetate and β-CD had a significant effect on emulsion stability. The emulsion with the highest stability was the sample with an HLB value of 10.7 with 1% β-CD concentration. This sample has 0% creaming index, 33800 ± 0.557 cP viscosity, and the lowest IC_{50} value, which proves its high antioxidant activity. Visual and microscopic observation showed no phase separation.

ACKNOWLEDGEMENT

The authors are thankful to Cyclochem Co. Ltd., Japan, for their kindness and support of cyclodextrin for the research. This study was supported in part by the Innovative Applied Research Program Vocational School, Universitas Diponegoro (Grant No: 26/UN7.5.13/SK/2022).

LIST OF ABBREVIATIONS

β-CD	:	Beta cyclodextrin
DPPH	:	2,2-diphenyl-1-picrylhydrazyl
HLB	:	Hydrophilic-lipophilic balance
IC_{50}	:	Half maximal inhibitory concentration

REFERENCES

Ahmad, H. N., Sri, P. N., & Sitanggang, A. B. (2020). Stability of herbal honey with emulsifier and antioxidant using spontaneous emulsification method. *Jurnal Litbang Industri, 10*(2), 111-120. https://doi.org/10.24960/jli.v10i2.6114.111-120

Alkhalaf, M. I., Alansari, W. S., Ibrahim, E. A., & Elhalwagy, M. E. (2019). Antioxidant, anti-inflammatory and anti-cancer activities of avocado (*Persea americana*) fruit and seed extract. *Journal of King Saud University-Science, 31*(4), 1358-1362. https://doi.org/10.1016/j.jksus.2018.10.010

Ciobanu, A., Landy, D., & Fourmentin, S. (2013). Complexation efficiency of cyclodextrins for volatile flavour compounds. *Food Research International, 53*(1), 110-114. https://doi.org/10.1016/j.foodres.2013.03.048

Faghmous, N., Bouzid, D., Boumaza, M., Touati, A., & Boyron, O. (2021). Optimisation of chitosan-coated W/O/W multiple emulsion stabilised with Span 80 and Tween 80 using Box–Behnken design. *Journal of Dispersion Science and Technology, 42*(10), 1566-1578. https://doi.org/10.1080/01932691.2020.1774387

Flores, M., Saravia, C., Vergara, C. E., Avila, F., Valdés, H., & Ortiz-Viedma, J. (2019). Avocado oil: Characteristics, properties, and applications. *Molecules, 24*(11), Article 2172. https://doi.org/10.3390/molecules24112172

Ghorbanzadeh, M., Farhadian, N., Golmohammadzadeh, S., Karimi, M., & Ebrahimi, M. (2019). Formulation, clinical and histopathological assessment of microemulsion-based hydrogel for UV protection of skin. *Colloids and Surfaces B: Biointerfaces, 179*, 393-404. https://doi.org/10.1016/j.colsurfb.2019.04.015

Hou, K., Xu, Y., Cen, K., Gao, C., Feng, X., & Tang, X. (2021). Nanoemulsion of cinnamon essential oil co-emulsified with hydroxypropyl-β-cyclodextrin and Tween-80: Antibacterial activity, stability and slow-release performance. *Food Bioscience, 43*, Article 101232. https://doi.org/10.1016/j.fbio.2021.101232

Huang, K., Liu, R., Zhang, Y., & Guan, X. (2021). Characteristics of two cedarwood essential oil emulsions and their antioxidant and antibacterial activities. *Food Chemistry, 346*, Article 128970. https://doi.org/10.1016/j.foodchem.2020.128970

Inoue, M., Hashizaki, K., Taguchi, H., & Saito, Y. (2010). Emulsifying ability of β-cyclodextrins for common oils. *Journal of Dispersion Science and Technology, 31*(12), 1648-1651. https://doi.org/10.1080/01932690903297058

Jiang, J., Wang, Z., Yan, J., Wang, C., Diao, D., Zhang, Y., & Zhang, L. (2024). Dual-responsive emulsion system: Unravelling pH and host-guest interactions for emulsion stability and enhanced oil recovery. *Colloids and Surfaces A: Physicochemical and Engineering Aspects, 684*, Article 133253. https://doi.org/10.1016/j.colsurfa.2024.133253

Khorolskyi, O. V., Kryvoruchko, A. V. (2021). Non-trivial behaviour of the acid–base balance of pure water near the temperature of its dynamic phase transition. *Ukrainian Journal of Physics, 66*(11), 972-977. https://doi.org/10.15407/ujpe66.11.972

Kibici, D., & Kahveci, D. (2019). Effect of emulsifier type, maltodextrin, and β-cyclodextrin on physical and oxidative stability of oil-in-water emulsions. *Journal of Food Science, 84*(6), 1273-1280. https://doi.org/10.1111/1750-3841.14619

Krumreich, F. D., Borges, C. D., Mendonça, C. R. B., Jansen-Alves, C., & Zambiazi, R. C. (2018). Bioactive compounds and quality parameters of avocado oil obtained by different processes. *Food Chemistry, 257*, 376-381.https://doi.org/10.1016/j.foodchem.2018.03.048

Lv, G., Wang, F., Cai, W., & Zhang, X. (2014). Characterization of the addition of lipophilic Span 80 to the hydrophilic Tween 80-stabilized emulsions. *Colloids and Surfaces A: Physicochemical and Engineering Aspects, 447*, 8-13. https://doi.org/10.1016/j.colsurfa.2014.01.066

Ma, J., Fan, J., Xia, Y., Kou, X., Ke, Q., & Zhao, Y. (2023). Preparation of aromatic β-cyclodextrin nano/microcapsules and corresponding aromatic textiles: A review. *Carbohydrate Polymers, 308*, Article 120661. https://doi.org/10.1016/j.carbpol.2023.120661

Maitera, O. N., Osemeahon, S. A., & Barnabas, H. L. (2014). Proximate and elemental analysis of avocado fruit obtained from Taraba state, Nigeria. *Indian Journal of Science and Technology, 2*(2), 67-73.

Netto, G. (MPharm), & Jose, J. (2018). Development, characterisation, and evaluation of sunscreen cream containing solid lipid nanoparticles of silymarin. *Journal of Cosmetic Dermatology, 17*(6), 1073-1083. https://doi.org/10.1111/jocd.12470

Onuki, Y., Kida, C., Funatani, C., Hayashi, Y., & Takayama, K. (2016). MRI as a promising tool for evaluation of the stability of cosmetic emulsions. *International Journal of Cosmetic Science, 38*(3), 272-278.https://doi.org/10.1111/ics.12290

Orhevba, B. A., & Jinadu, A. O. (2011). Determination of physico-chemical properties and nutritional contents of avocado pear (*Persea americana* M.). *Academic Research International, 1*(3), 372-381.

Paramita, V., Fitri Novia, S., Dwi Ariyanto, H., Pramudono, B., Yoshii, H., Kusumayanti, H., & Amalia, R. (2022). The influence of α-cyclodextrin on the stability and fatty acids of medium-chain triglycerides high oil load nanoemulsion. *Materials Today: Proceedings, 63*, S312-S317. https://doi.org/10.1016/j.matpr.2022.03.156

Putri, A. I. E., & Ariyanto, H. D. (2022). Effect of hydrophilic-lipophilic balance (HLB) value on the stability of cosmetic lotion based on walnut oil (*Canarium indicum* L.) oil-in-water emulsion. *Journal of Vocational Studies on Applied Research, 4*(2), 53-60. https://doi.org/10.14710/jvsar.v4i2.15376

Sahafi, S. M., Goli, S. A. H., Kadivar, M., Varshosaz, J., & Shirvani, A. (2021). Pomegranate seed oil nanoemulsion enriched by α-tocopherol: The effect of environmental stresses and long-term storage on its physicochemical properties and oxidation stability. *Food Chemistry, 345*, Article 128759.https://doi.org/10.1016/j.foodchem.2020.128759

Shimada, K., Kawano, K., Ishii, J., & Nakamura, T. (1992). Structure of inclusion complexes of cyclodextrins with triglyceride at the vegetable oil/water interface. *Journal of Food Science, 57*(3), 655-656. https://doi.org/10.1111/j.1365-2621.1992.tb08063.x

Singh, S., Lohani, A., Mishra, A. K., & Verma, A. (2019). Formulation and evaluation of carrot seed oil-based cosmetic emulsions. *Journal of Cosmetic and Laser Therapy, 21*(2), 99-107. https://doi.org/10.1080/14764172.2018.1469769

Sun, N., Jiang, H., Wang, X., & Jin, K. (2017). Research on factors affecting heavy oil-in-water emulsion rheology and pressure drop. *Journal of Dispersion Science and Technology, 39*(3), 411-422.https://doi.org/10.1080/01932691.2017.1324795

Whitby, C. P., Fornasiero, D., & Ralston, J. (2009). Effect of adding anionic surfactant on the stability of Pickering emulsions. *Journal of Colloid and Interface Science, 329*(1), 173-181. https://doi.org/10.1016/j.jcis.2008.09.056

Xiao, Z., Zhang, Y., Niu, Y., Ke, Q., & Kou, X. (2021). Cyclodextrins as carriers for volatile aroma compounds: A review. *Carbohydrate Polymers, 269*, Article 118292. https://doi.org/10.1016/j.carbpol.2021.118

Pertanika J. Sci. & Technol. 34 (2): 861 - 887 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Response Surface Optimisation of a *Strobilanthes crispus*-Loaded Red Palm Oil Nanoemulsion with Enhanced Physicochemical Stability

Ee Lynn Fong[1], Chui Fung Loke[2], Chin Ping Tan[3], Tze Ven Poh[4], Bee Hiong Tan[5], and Oi Ming Lai[1,6]*

[1]*Natural Medicines and Products Research Laboratory (NaturMeds), Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Department of Bioscience, Faculty of Applied Sciences, Tunku Abdul Rahman University of Management and Technology, 53300 Setapak, Kuala Lumpur, Malaysia*
[3]*Department of Food Technology, Faculty of Food Science and Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[4]*Department of Computer Science and Embedded System, Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, 53300 Kuala Lumpur, Malaysia*
[5]*Aurora Health Sdn Bhd, MTDC-UPM, 43400 Serdang, Selangor, Malaysia*
[6]*Department of Bioprocess Technology, Faculty of Biotechnology & Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

ABSTRACT

Plant-derived lipophilic bioactive compounds often show poor dispersion and instability in a water system; this limits their use in topical and pharmaceutical formulations. This study aimed to develop and optimise an oil-in-water nanoemulsion containing *Strobilanthes crispus* extract dissolved in red palm oil using Response Surface Methodology (RSM). A three-factor Box-Behnken design was applied to evaluate the effects of Tween® 80 concentration, glycerol concentration, and homogenisation pressure on droplet size, polydispersity index (PDI), and creaming index. The results show that the surfactant concentration was the most important factor affecting the droplet size and overall stability. The optimised nanoemulsion formulation comprised 5.06% (w/w) Tween® 80, 10.4% (w/w) glycerol, and a homogenisation pressure of 617.1 bar. Experimental validation produced a nanoemulsion with a droplet size of 151.8 ± 1.3 nm, PDI of 0.127 ± 0.021, and no visible creaming. The regression models showed strong predictive ability for droplet size (R^2 = 0.984) and creaming index (R^2 = 0.999), but moderate prediction for PDI (R^2 = 0.7516). Stability studies under different pH (2-6),

ARTICLE INFO

Article history:
Received: 21 July 2025
Accepted: 25 March 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.12

E-mail addresses:
gs65477@student.upm.edu.my (Ee Lynn Fong)
lokecf@tarc.edu.my (Chui Fung Loke)
tancp@upm.edu.my (Chin Ping Tan)
pohtv@tarc.edu.my (Tze Ven Poh)
laurel@aurorahealthco.com (Bee Hiong Tan)
omlai@upm.edu.my (Oi Ming Lai)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

temperature (40-60°C), ionic strength (0.5-2.5mM), and centrifugal conditions (1000-5000rpm), as well as four-month storage, showed that the nanoemulsion remained physically stable without phase separation. This stability is mainly attributed to steric stabilisation from the non-ionic surfactant. Overall, the study demonstrates that Response Surface Methodology is an effective tool for optimising nanoemulsion formulations for the delivery of lipophilic plant bioactives.

Keywords: Box-Behnken design, lipophilic bioactive, nanoemulsion, physicochemical stability, red palm oil, response surface methodology, *Strobilanthes crispus*

INTRODUCTION

Nanoemulsions are kinetically stable colloidal dispersions consisting of two immiscible liquids, usually stabilised by a surfactant, with droplet diameters typically in the range of 20-200nm. Their small droplet size gives nanoemulsions a larger surface area, which enhances the dispersion of hydrophobic compounds in aqueous systems and improves their bioavailability (Preeti et al., 2023). Oil-in-water (O/W) nanoemulsions are widely used in pharmaceutical and cosmetic applications as they allow lipophilic bioactives to be effectively incorporated into water-based formulations (Shakeel et al., 2021). Nanoemulsions also exhibit enhanced physical stability by minimising instability mechanisms such as creaming, flocculation, and coalescence, which contribute to improved shelf-life (Marhamati et al., 2021).

The development and stability of nanoemulsions are influenced by both formulation components and processing parameters. Surfactants reduce interfacial tension and stability of droplets, while co-solvents like glycerol affect the system's viscosity and density, influencing the droplet movement and stability. High-pressure homogenisations play a key role in reducing droplet size and ensuring a homogeneous dispersion (Zhou et al., 2022). Response Surface Methodology (RSM) is often used to evaluate the interaction between these variables and identify the optimal formulation conditions (Podder & Mukherjee, 2024).

Strobilanthes cripus (*S.crispus*) is a medicinal plant widely used in Southeast Asia and is known for its antioxidant, anti-inflammatory, and antiproliferative properties (Zhu et al., 2022). Studies have shown that *S.crispus* bioactive compounds can help regulate keratinocyte proliferation and reduce skin inflammation, which is useful for topical applications (Imam et al., 2023). Red palm oil can also be used for skin applications as it is rich in tocotrienols and carotenoids, which have strong antioxidant properties and help protect the skin from oxidative damage (Tan et al., 2021). Since these compounds are lipophilic, they are unlikely to dissolve in water-based formulations, so a suitable delivery system is required to ensure their stability and bioactivity.

In this study, an oil-in-water nanoemulsion containing red palm oil and *S.crispus* extract was developed and optimised using a Box–Behnken design within Response

Surface Methodology. The effects of formulation and processing variables on droplet size, polydispersity index, and physical stability were evaluated. The optimised nanoemulsion was then tested under various storage and stress conditions, and the results showed it remained stable over time. These findings highlight that nanoemulsion design can not only stabilise lipophilic bioactives but also enhance their delivery and efficacy as natural skin-protective ingredients.

MATERIAL AND METHODS

Preparation of Oil-in-Water Coarse Emulsion

An oil-in-water (O/W) nanoemulsion system was developed. The oil phase was red palm oil with *Strobilanthes crispus* extract, and the water phase contained Tween® 80 and glycerol.

The formulation included 20% (w/w) red palm oil (Nutrolein®, Golden Palm Oil) with 0.008% (w/w) *S. crispus* extract from TKC Herbal Nursery Sdn. Bhd. (Seremban, Malaysia). Tween® 80 (15% w/w; Fisher Scientific, Loughborough, UK) was used as the surfactant, and glycerol (30% w/w; USP grade, InterMed Sdn. Bhd., Kuala Lumpur, Malaysia) was added to adjust the thickness and density of the water phase. Citric acid (0.1% w/w; Shandong Ensign Industry Co. Ltd., China) was added to improve surfactant action at the oil-water interface, enhancing stability and reducing droplet aggregation, while also acting as a preservative. Deionised water was added to make 100% of the formulation.

For the coarse emulsion, the oil and water phases were first mixed and stirred together. The mixture was then pre-emulsified using a high-speed homogeniser (Silverson L4R, Buckinghamshire, UK) at 5000-8000 rpm for 5-10 minutes at room temperature until a uniform coarse emulsion was formed.

Optimisation of High-pressure Homogenisation Cycles

The coarse emulsion from the pre-emulsification step was further processed using a high-pressure homogeniser (Panda 2 K, Niro Soavi, Lübeck, Germany) to produce the nanoemulsion.

A preliminary study was carried out to study the effect of homogenisation cycles on nanoemulsion droplet size and stability. The coarse emulsion was homogenised at a constant pressure of 600 bar for up to eight cycles. Samples were collected after each cycle (1-8 cycles) and analysed for the droplet size, polydispersity index (PDI), and visual stability to determine the optimal number of cycles required for nanoemulsion formation. High-pressure homogenisation is effective in reducing droplet size and producing uniform nanoemulsions as it generates strong shear, cavitation, and turbulence forces during processing (Chandel et al., 2025; Rauber et al., 2025)

The emulsion temperature was monitored during homogenisation and kept below 50°C (±5°C) to minimise potential degradation of heat-sensitive bioactive compounds. All experiments were performed in triplicate.

Preparation of Nanoemulsion

After the preliminary study was conducted to determine optimum homogenisation conditions and parameter range, a Box-Behnken design was used to evaluate the effect of homogenisation pressure, surfactant concentration, and glycerol concentration. RSM was applied to optimise these variables and analyse their interactions (Kumar & Reji, 2023).

The coarse emulsion prepared under optimised pre-emulsification was then processed using a high-pressure homogeniser to produce the nanoemulsion. Homogenisation was run with the optimum number of cycles determined from a preliminary study, with the pressure range 500 to 700 bar based on the experimental design. The process helps reduce droplet size through strong shear, turbulence and cavitation force.

The nanoemulsion was allowed to equilibrate at room temperature (24.0 ± 1.0°C) before the physicochemical characteristics study to ensure stabilisation of the droplet system.

Optimisation of Nanoemulsion Formulation using Response Surface Methodology

The nanoemulsion formulation was optimised using a Box-Behnken design (BBD) under Response Surface Methodology (RSM) to evaluate the effects of surfactant concentration (Tween® 80, 5-15% w/w), glycerol concentration (10-30% w/w), and homogenisation pressure (500-700 bar). Each variable was studied at three coded levels (-1,0,+1), corresponding to actual value presented in Table 1. These ranges were selected based on preliminary screening and literature to ensure adequate interfacial stabilisation, appropriate viscosity, and effective droplet size reduction (Chong et al., 2022; Lüdtke et al., 2022).

The experimental matrix of the Box–Behnken design (BBD), including the combinations of surfactant concentration (X_1), glycerol concentration (X_2), and homogenisation pressure (X_3), is presented in Table 2.

The oil-to-surfactant (O/S) ratio is a key determinant of droplet formation and emulsion stability, as it governs the balance between dispersed phase volume and interfacial coverage by surfactants. In this study, the oil phase was fixed at 20% (w/w) based on preliminary optimisation, where this concentration provided sufficient solubilisation of *Strobilanthes crispus* extract while maintaining formulation stability. Lower oil content resulted in insufficient loading, whereas higher oil content reduced stability due to droplet aggregation. This range is also consistent with reported stable nanoemulsion systems (20-30% oil fraction) (Sarheed et al., 2020; Luu et al., 2025). Red palm oil was selected as the oil phase due to its high content of carotenoids and tocotrienols, which contribute to antioxidant activity and support topical applications (Chong et al., 2018).

The *Strobilanthes crispus* extract concentration (0.008% w/w) was selected based on preliminary studies demonstrating effective antioxidant and anti-inflammatory activity for topical application, while ensuring complete solubility and formulation stability. This is supported by Loke et al. (2019), who reported strong antioxidative activity of *S. crispus* extract at low concentrations.

Optimisation was therefore achieved by varying surfactant concentration, which indirectly adjusted the O/S ratio while evaluating the effects of glycerol concentration and homogenisation pressure. Droplet size (Y_1), polydispersity index (PDI, Y_2), and creaming index (Y_3) were used as indicators of nanoemulsion quality and stability (Mukherjee & Baruah, 2025). The BBD consisted of 15 experimental runs, and the data were fitted to a second-order polynomial model to determine optimal formulation conditions.

Table 1
Coded levels of independent variables used in response surface methodology (RSM)

Independent Variables	Symbols	Coded levels		
		-1	0	1
Surfactant concentration (%)	X_1	5	10	15
Glycerol concentration (%)	X_2	10	20	30
Homogenisation Pressure (bar)	X_3	500	600	700

Table 2
Matrix of the Box-Behnken design (BBD)

Run	X_1: Surfactant Concentration (% w/w)	X_2: Glycerol Concentration (% w/w)	X_3: Homogenisation Pressure (bar)
1	15	20	700
2	5	30	600
3	15	20	500
4	5	20	700
5	10	20	600
6	10	30	500
7	15	30	600
8	10	10	700
9	10	20	600
10	5	10	600
11	10	30	700
12	5	20	500
13	10	20	600
14	10	10	500
15	15	10	600

Stability Studies of Emulsion

Evaluation of Droplet Size and Polydispersity Index

The mean droplet size and polydispersity index (PDI) of both coarse emulsion and nanoemulsion were measured using dynamic light scattering (DLS) with Zetasizer Nano-ZS (Malvern Instruments Ltd., Worcestershire, UK). The sample was diluted 10-fold

with deionised water to minimise multiple scattering effects and to prevent the interference from high particle concentration, which can affect the accuracy of DLS measurement (Rodriguez-Loya et al., 2023). Data collection and analysis were carried out using Malvern Zetasizer Software version 7.11.

Evaluation of Creaming Index

The coarse emulsion and nanoemulsion were monitored for the formation of a creaming layer to evaluate stability, following the method described by Mohammed et al. (2020). After 24 hours of incubation at room temperature, the height of the creamed layer (HL) and the total height of the emulsion (HE) were measured. The creaming index was then calculated using the formulation Equation 1:

$$CI = \frac{HL}{HE} \times 100 \quad [1]$$

Short-term Stability Studies of Nanoemulsion

Short-term stability refers to accelerated stability studies in which different environmental stresses are applied to the sample, following the guidelines of the International Conference on Harmonisation (ICH). Freshly prepared nanoemulsion samples of 5mL were used for testing for different experimental conditions, which are summarised in Table 3. All samples were then incubated at room temperature (25°C) and stored for 48 hours, after which droplet size, PDI, and creaming index measurements were taken.

Long-term Stability Studies of Nanoemulsion

For long-term stability studies, 10 mL of newly prepared nanoemulsion was aliquoted into 15 mL polypropene centrifuge tubes and stored in the dark at 5°C, 25°C, and 40°C for four months.

Table 3
Experimental conditions for accelerated stability testing for nanoemulsion

Parameter	Conditions
pH adjustment	Samples were prepared with varying pH ranging from 2 to 6 using 1 N hydrochloric acid and 1 N sodium hydroxide solutions.
Heat treatment	Samples were incubated at temperatures ranging from 40°C to 60°C for 30 minutes in a water bath.
Ionic Strength	Samples were prepared with varying ionic strengths, ranging from 0.5 mM to 2 mM, by adding a 5 mM sodium chloride solution.
Centrifugation	Samples were centrifuged at speeds from 1000 to 5000 rpm for 30 minutes at room temperature.

Droplet size, PDI, and creaming index were measured at predetermined time points of 0, 2, 3, and 4 months.

Statistical Analysis

Experimental design and response surface methodology (RSM) analysis were performed using Design Expert version 13 (Stat-Ease, Inc., Minneapolis, Minnesota, USA). Measurements were carried out in triplicate and represented as mean ± standard deviation. One-way ANOVA was performed using Minitab Statistical Software version 21.2 (Minitab, LLC, Pennsylvania, USA), and Tukey's post hoc test was subsequently applied. Statistical significance was kept at $P < 0.05$ with a 95% confidence interval.

RESULTS AND DISCUSSION

Influence of Homogenisation Speed and Time on Coarse Emulsion Formation

High-speed homogenisation plays a critical role in producing a stable coarse emulsion by reducing droplet size, enhancing uniformity, and improving overall stability. Increasing the speed intensifies shear forces, resulting in smaller droplet sizes due to more efficient disruption of oil droplets within the aqueous phase. As shown in Figure 1, the droplet size decreased significantly as the speed increased. Additionally, extending homogenisation time from 5 to 10 minutes further reduced droplet size, indicating that prolonged shear allows more complete droplet breakup.

The result for the polydispersity index (PDI) shows that the higher homogenisation speeds generally produce lower PDI, as summarised in Table 4. This also indicates that the emulsion is characterised by a narrower droplet size distribution and greater uniformity. However, at the speed of 8000 rpm for 10 minutes, a slight and unexpected increase in PDI. This result may be explained by the droplet's recoalescence due to excessive shear energy, which has a similar observation reported by Xu et al. (2018). As a result, a homogenising speed of 8000 rpm for 5 minutes was determined to be optimal for coarse emulsion conditions, producing fine droplets while minimising potential degradation of heat-sensitive bioactives, such as the phenolic compound in *S.crispus* extract (Chua et al., 2019).

Effect of High-pressure Homogenisation Cycle on Nanoemulsion Properties

High-pressure homogenisation affected the nanoemulsion properties, especially droplet size, polydispersity index (PDI), and physical stability. Increasing the number of homogenisation cycles reduced droplet size through shear, turbulence, and cavitation forces (Singh et al., 2017). Too many cycles, however, can cause overprocessing and lead to droplet aggregation because of more frequent collisions (Zhou et al., 2023). Therefore, it is important to choose the right number of cycles to balance size reduction and stability.

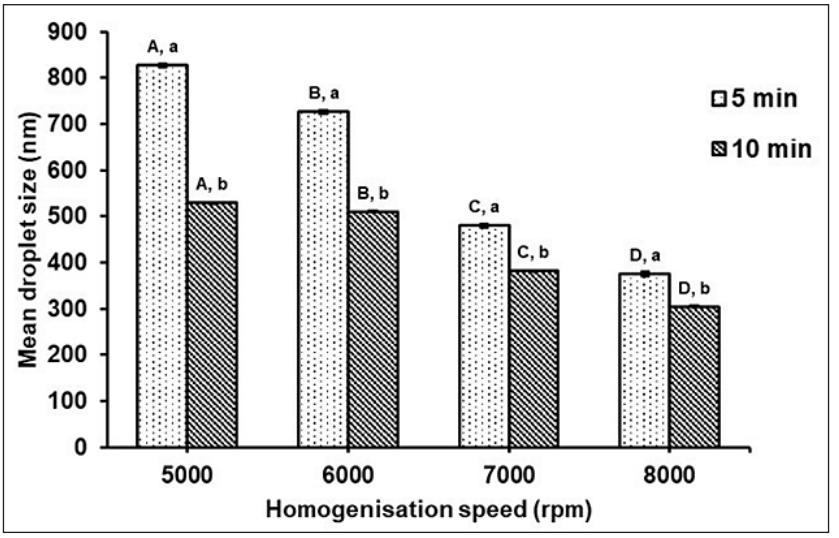


Figure 1. Effect of homogenisation speed and time on the mean droplet size of coarse emulsion
Note. Different uppercase letters indicate significant differences at the same homogenisation speed, whereas different lowercase letters indicate significant differences at the same homogenisation time

Table 4
Effect of homogenisation speed and time on the polydispersity index (PDI) of the coarse emulsion

Homogenisation Speed (rpm)	Homogenisation Time (min)	
	5	10
5000	0.440 ± 0.015A, a	0.452 ± 0.009A, a
6000	0.496 ± 0.006B, a	0.459 ± 0.008A, b
7000	0.429 ± 0.001A, a	0.421 ± 0.001A, b
8000	0.303 ± 0.017C, a	0.345 ± 0.015B, a

Note. Different uppercase superscript letters in the same column and lowercase superscript letters in the same row denote statistical significance ($P < 0.05$)

The mean droplet size decreased from 197.9 nm to 128.1 nm after eight cycles, showing effective droplet dispersion (Figure 2). This matches previous studies, where more cycles improved droplet break-up due to stronger shear and cavitation (Preiss et al., 2021). The PDI and creaming index at different homogenisation cycle are shown in Table 5. The PDI decreased with increasing cycles, showing better size uniformity (Santos et al., 2022). Samples processed with 1-6 cycles showed visible creaming, which indicates poor physical stability, likely caused by density differences (Chow et al., 2024). From cycle 7 onwards, the creaming index was 0.0, showing that the nanoemulsion was physically stable (Galvão et al., 2018).

The optimum number of cycles was chosen based on droplet size in the nanoscale range (100-200 nm), PDI below 0.2, and no creaming. The temperature was kept below 50°C (±5°C) to prevent heat damage. High temperatures can cause surfactant dehydration, destabilise the interface, and degrade heat-sensitive bioactive compounds (Iacob-Tudose et al., 2021). Cycles 7 and 8 both produced stable nanoemulsions. Cycle 7 was chosen as the best condition because it achieved the desired results with less processing. Extra cycles may slightly improve droplet dispersion but can also increase heat exposure (Zhou et al., 2023). Cycle 7 provides efficient emulsification while reducing unnecessary processing. These conditions were used in the following RSM optimisation.

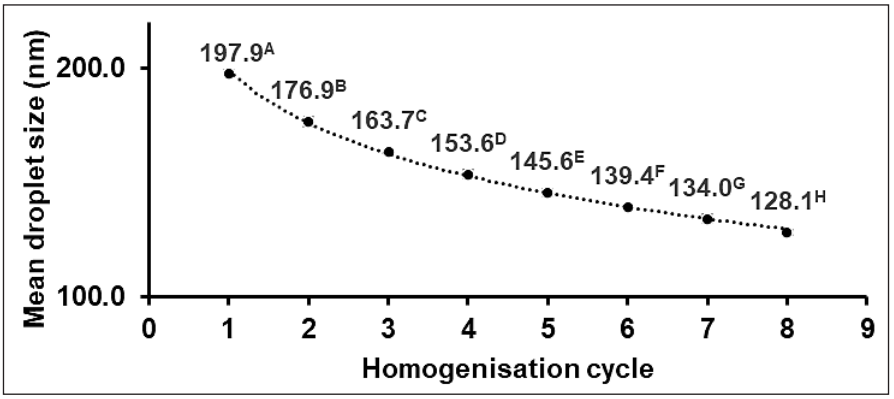


Figure 2. Mean droplet size of nanoemulsion at different high-pressure homogenisation cycles

Table 5
Polydispersity index and creaming index of nanoemulsion processed at different high-pressure homogenisation cycles

Homogenisation Cycle	PDI	Creaming Index
1	0.187 ± 0.009^{B}	42.42
2	0.147 ± 0.009^{A}	30.30
3	0.156 ± 0.017^{AB}	24.24
4	0.150 ± 0.005^{AB}	15.15
5	0.147 ± 0.007^{A}	9.09
6	0.153 ± 0.005^{AB}	6.06
7	0.135 ± 0.005^{A}	0.00
8	0.143 ± 0.005^{A}	0.00

Note. Superscript letters denote significant differences within columns ($p < 0.05$)

Optimisation of Nanoemulsion using Response Surface Methodology (RSM)

Response Surface Methodology (RSM) is a statistical method used to model and optimise processes with multiple variables and responses (Bakhaidar et al., 2022). In this study, a Box–Behnken design (BBD) was used to examine the effects of surfactant concentration, glycerol concentration, and homogenisation pressure on nanoemulsion properties. A total of 15 experimental runs were carried out to explore the formulation efficiently.

Figures 3, 4, and 5 show predicted versus actual response plots. The data points are closely clustered along the diagonal, which indicates good agreement between predicted and experimental values and confirms that the models are reliable. The normal probability plot for PDI showed a slight deviation from linearity, suggesting higher variability for this response. This may be due to the sensitivity of PDI to small changes in droplet size and processing conditions.

The coefficients of determination (R^2 and adjusted R^2) in Tables 6 to 8 also support good model fitting for droplet size and creaming index. Analysis of variance (ANOVA) showed that the quadratic models were statistically significant ($p < 0.05$), and the high F-values indicate strong effects of the formulation variables on the responses (Pongsumpun et al., 2020). The regression equations in Table 6,7 and 8 were used to predict the optimal formulation conditions for the nanoemulsion.

The regression equation for droplet size, expressed in terms of the actual factors, is as follows in Equation 2:

$$Y_1 = 216.88571 - 0.209571X_1 + 0.047750X_2 - 0.038875X_3 - 0.004820X_1X_2 - 0.002170X_1X_3 + 0.013086X_1^2 \quad [2]$$

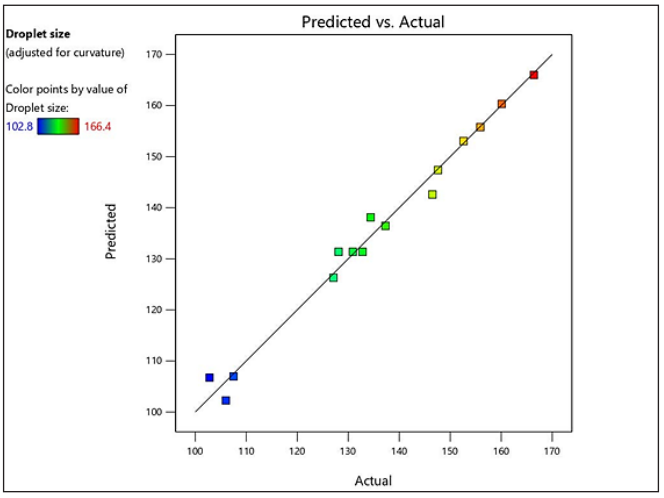


Figure 3. Predicted vs actual plot for droplet size of nanoemulsion

Table 6
ANOVA summaries and regression coefficients for droplet size, PDI and creaming index responses (Y_1)

Variable	Droplet Size (Y_1)		
	Mean Square	**F-value**	**P-value**
Intercept	890.54	96.72	<0.0001
Main effect			
X_1	2346.13	254.80	<0.0001
X_2	746.91	81.12	<0.0001
X_3	1737.55	188.71	<0.0001
Interaction effect			
X_1X_2	145.20	15.77	0.0041
X_1X_3	117.72	12.79	0.0072
Quadratic effect			
X_1^2	249.72	27.12	0.0008
Lack of fit	10.41	1.86	0.3897
R^2		0.9864	
Adjusted R^2		0.9762	

Note. X_1 = surfactant concentration (% w/w); X_2 = glycerol concentration (% w/w); X_3 = homogenisation pressure (bar)

The regression equation for the polydispersity index, expressed in terms of the actual factors, is as follows in Equation 3:

$$Y_2 = 0.115167 + 0.000535X_1 - 0.000155X_2 + 0.000024X_3 \quad [3]$$

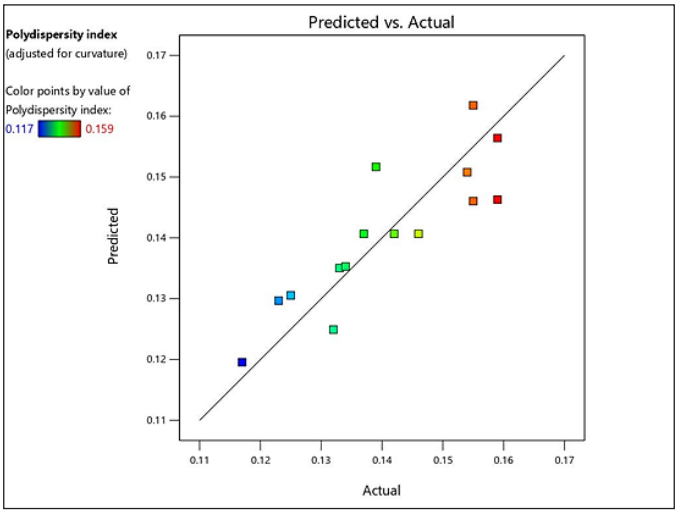


Figure 4. Predicted vs actual plot for polydispersity index of nanoemulsion

Table 7
ANOVA of the regression coefficient of the fitted quadratic equations for the polydispersity index (Y_2)

Variable	Polydispersity index (Y_2)		
	Mean Square	**F-value**	**P-value**
Intercept	0.0007	11.10	0.0012
Main effect			
X_1	0.0014	24.35	0.0004
X_2	0.0005	8.17	0.0155
X_3	0.0000	0.7677	0.3997
Lack of fit	0.0001	3.31	0.2535
R^2		0.7516	
Adjusted R^2		0.6839	

Note. X_1 = surfactant concentration (% w/w); X_2 = glycerol concentration (% w/w); X_3 = homogenisation pressure (bar)

The regression equation for the creaming index, expressed in terms of the actual factors, is as follows in Equation 4:

$$Y_3 = 87.91500 + 1.07710X_1 - 0.482975X_2 - 0.265213X_3 - 0.002904X_1X_2 - 0.001598X_1X_3 + 0.000714X_2X_3 + 0.003246X_1^2 + 0.000640X_2^2 + 0.000197X_3^2 \quad [4]$$

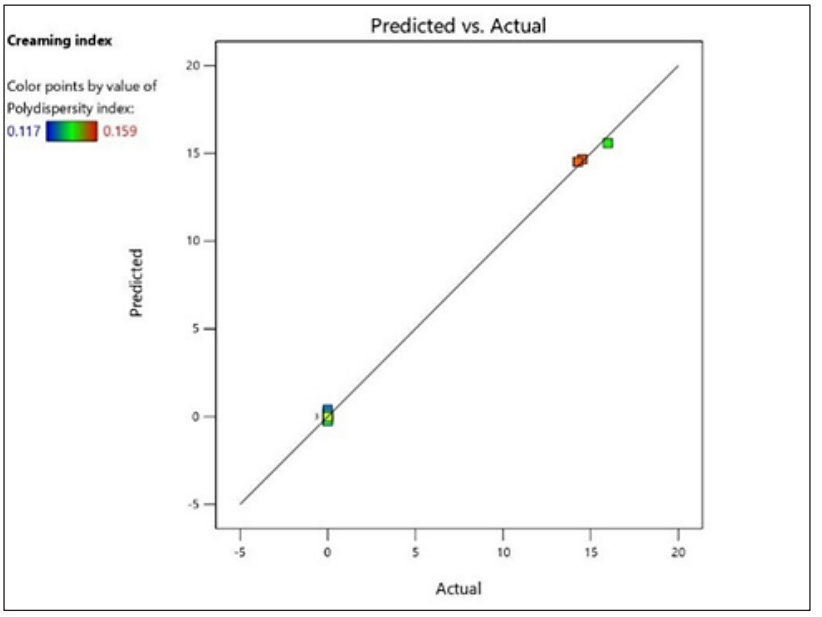


Figure 5. Predicted vs actual plot for creaming index

Table 8
ANOVA of the regression coefficient of the fitted quadratic equations for creaming index (Y_3)

Variable	Creaming Index (Y_3)		
	Mean Square	**F-value**	**P-value**
Intercept	59.51	465.06	<0.0001
Main effect			
X_1	116.25	908.77	<0.0001
X_2	103.61	809.72	<0.0001
X_3	114.36	893.93	<0.0001
Interaction effect			
X_1X_2	52.71	411.92	<0.0001
X_1X_3	63.84	498.93	<0.0001
X_2X_3	50.91	397.86	<0.0001
Quadratic effect			
X_1^2	15.20	1118.77	0.0001
X_2^2	9.47	73.99	0.0004
X_3^2	14.27	111.56	0.0001
Lack of fit	0.2133		
R^2		0.9988	
Adjusted R^2		0.9967	

Note. X_1 = surfactant concentration (% w/w); X_2 = glycerol concentration (% w/w); X_3 = homogenisation pressure (bar)

Influence of Independent Variables on Droplet Size

The results in Table 6 show that the model explained 98.64% of the variation in droplet size ($R^2 = 0.9864$; adjusted $R^2 = 0.9762$), indicating strong predictive ability and minimal overfitting. All main factors (X_1, X_2, and X_3) significantly affected droplet size ($p < 0.05$), with surfactant concentration having the strongest effect, followed by homogenisation pressure and glycerol concentration. This highlights that stabilising the oil-water interface is crucial for controlling droplet breakup and preventing recoalescence during high-pressure homogenisation. The optimised droplet size of 151.8 nm is smaller than many plant-based nanoemulsions, showing efficient droplet disruption and stabilisation under the selected conditions (Gupta et al., 2016; Singh et al., 2017).

It should be noted that droplet size measurements were performed under diluted conditions to ensure accuracy in dynamic light scattering (DLS) analysis. Although dilution may influence inter-droplet interactions, the same dilution protocol was consistently applied to all samples. Therefore, the results remain reliable for comparative evaluation of formulation and processing effects.

A significant interaction between surfactant concentration and homogenisation pressure (X_1X_3, $p < 0.05$) suggests that droplet size reduction depends on both interfacial

coverage and mechanical energy. At low surfactant levels, increasing pressure alone does not reduce droplet size effectively because the interface is not fully stabilised, which can lead to droplet merging. At higher surfactant concentrations, there is enough coverage for the applied mechanical energy to break droplets into smaller, stable sizes. The quadratic effect of surfactant concentration (X_1^2) also indicates that there is an optimal surfactant level, beyond which further increases do not significantly reduce droplet size.

The response surface plot in Figure 6a-c illustrates the combined effect of surfactant concentration, glycerol concentration and homogenisation pressure on droplet size. Increasing surfactant concentration reduces droplet size by lowering interfacial tension and stabilising the droplets, thereby preventing coalescence. Similarly, increasing homogenisation pressure decreases droplet size. This explains that higher homogenisation pressure give stronger shear, turbulence, and cavitation forces break droplets more effectively (Gawin-Mikołajewicz et al., 2023; Kaur et al., 2024; Zhou et al., 2023). However, excessive mechanical energy does not always improve droplet size and may lead to temporary aggregation due to more frequent droplet collisions.

In contrast, glycerol concentration has a comparatively small effect on droplet size and does not show a significant interaction with homogenisation pressure ($p > 0.05$). This suggests that glycerol mainly affects the viscosity and density of the water phase, rather than directly contributing to droplets disruption. Higher viscosity can slow droplet movement and reduce creaming, but it has a limited effect on droplet formation during homogenisation (Gupta et al., 2016). This explains why glycerol had a smaller impact compared to surfactant and pressure.

The lack-of-fit test was not significant ($p > 0.05$), confirming that the model accurately represents the data within the studied range. Overall, these results show that successful nanoemulsion formation requires a balance between interfacial stabilisation and mechanical energy, with surfactant concentration being the most critical factor for achieving small, uniform droplets.

Influence of Independent Variables on Polydispersity Index (PDI)

The analysis in Table 7 shows that surfactant concentration (X_1) and glycerol concentration (X_2) significantly affected droplet size distribution ($p < 0.05$). Surfactant had the strongest effect ($F = 24.35$, $p = 0.0004$), followed by glycerol ($F = 8.17$, $p = 0.0155$). Homogenisation pressure (X_3) was not significant ($p = 0.3997$), suggesting that mechanical energy has less influence on droplet uniformity than surfactant coverage and the properties of the continuous phase (Singh et al., 2017). The lack-of-fit test was not significant ($p = 0.2535$), indicating the model fits the data well.

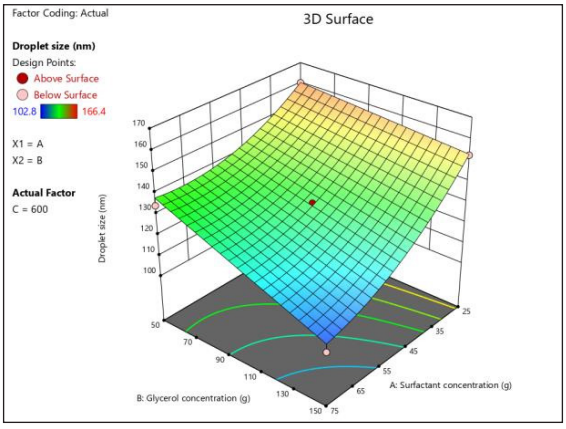


Figure 6a. Response surface and contour plots illustrating the interaction between surfactant concentration and glycerol concentration on the droplet size of the nanoemulsion

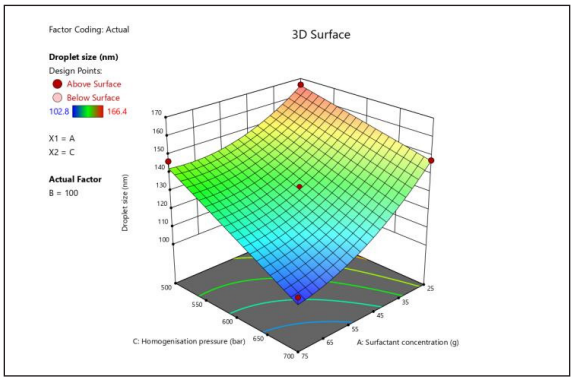


Figure 6b. Response surface and contour plots illustrating the interaction between surfactant concentration and homogenisation pressure on the droplet size of the nanoemulsion

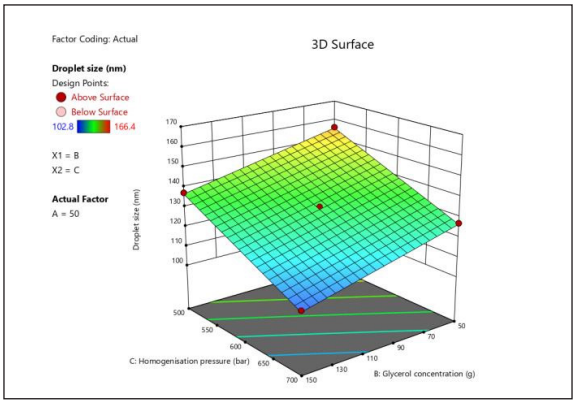


Figure 6c. Response surface and contour plots illustrating the interaction between glycerol concentration and homogenisation pressure on the droplet size of the nanoemulsion

The model's R^2 (0.7516) and adjusted R^2 (0.6839) show moderate predictability. This is lower than for droplet size because PDI is more sensitive to small changes in droplet distribution and processing conditions. PDI reflects the spread of droplet sizes, so minor variations in homogenisation efficiency, local concentration differences, or temporary droplet merging can affect it (Kaur et al., 2024). Other factors not included in the model, such as slight temperature changes or interfacial dynamics, may also contribute to variability (Zhou et al., 2023). Additionally, small differences in shear during homogenisation and the sensitivity of DLS to a few larger droplets can increase variation.

The response surface plots in Figure 7a–b illustrate the combined effects of surfactant concentration, glycerol concentration, and homogenisation pressure on the polydispersity index (PDI). Lower PDI values indicate more uniform droplets. In Figure 7a, increasing surfactant concentration reduced PDI by improving interfacial coverage and limiting droplet coalescence (Singh et al., 2017). Glycerol concentration shows a comparatively small effect, with a slight reduction in PDI at higher levels, likely due to increased viscosity of the continuous phase, which limits droplet movement and reduces aggregation (Gupta et al., 2016).

In Figure 7b, increasing both surfactant concentration and homogenisation pressure further decreases PDI, as stronger mechanical forces promote more efficient droplet breakup. No significant interactions were observed between glycerol concentration and homogenisation pressure ($p > 0.05$), suggesting that each factor mainly affects PDI independently.

Future studies could include additional variables, such as temperature control or surfactant combinations, to improve model reliability for predicting PDI.

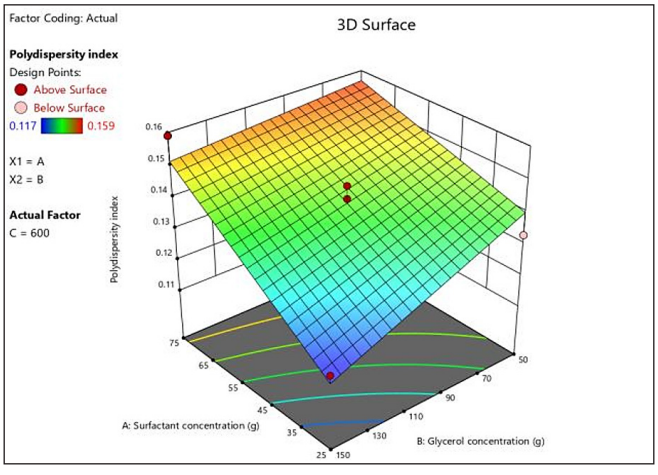


Figure 7a. Response surface and contour plots illustrating the interaction between surfactant concentration and glycerol concentration on the PDI of the nanoemulsion

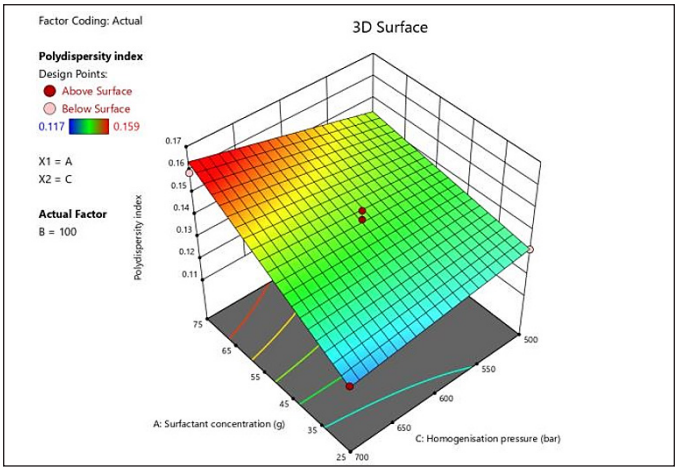


Figure 7b. Response surface and contour plots illustrating the interaction between surfactant concentration and homogenisation pressure on the PDI of the nanoemulsion

Influence of Independent Variables on Creaming Index

Table 8 shows the ANOVA results for the creaming index (Y_3). All three formulation variables, surfactant concentration (X_1), glycerol concentration (X_2), and homogenisation pressure (X_3), significantly affected nanoemulsion stability ($P < 0.0001$). The high F-values indicate that these factors strongly reduce creaming. All interaction terms (X_1X_2, X_1X_3, and X_2X_3) were also significant ($P < 0.0001$), showing that the combination of formulation and processing parameters plays an important role in stabilising the emulsion.

The quadratic effect of surfactant concentration (X_1^2) was especially strong ($F = 1118.77$, $P = 0.0001$), indicating a non-linear relationship between surfactant level and creaming. The model fit the data very well, with $R^2 = 0.9988$ and adjusted $R^2 = 0.9967$, and the lack-of-fit test was not significant ($P > 0.05$), confirming that the model reliably represents the experimental results. Overall, these results show that the optimised nanoemulsion is highly resistant to creaming and physically stable.

Figure 8 shows the response surface plots for the creaming index. Figure 8a demonstrates the interaction between surfactant and glycerol concentrations. Minimal creaming was observed when both variables were within an optimal range. Too much surfactant, especially at low glycerol levels, increased creaming, likely due to changes in interfacial structure or higher viscosity that limited droplet mobility. This highlights the need to balance surfactant and continuous phase modifiers to maintain stability.

Figure 8b shows the interaction between homogenisation pressure and surfactant concentration. At low surfactant levels, higher pressure improved stability by reducing droplet size and creaming. At high surfactant levels, low pressure was not enough to reduce droplet size, resulting in more creaming. This shows that mechanical energy must work together with surfactant coverage to achieve optimal nanoemulsion formation.

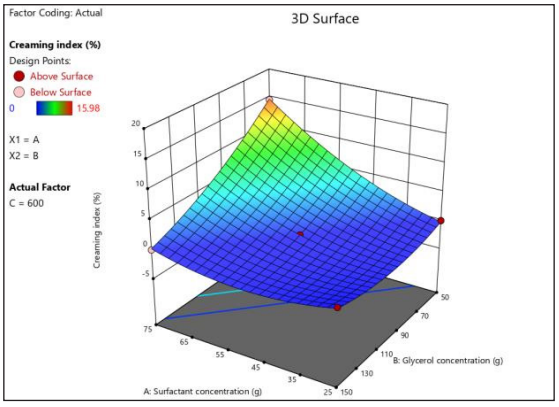


Figure 8a. Response surface and contour plots illustrating the interaction between surfactant concentration and glycerol concentration on the droplet size of the nanoemulsion

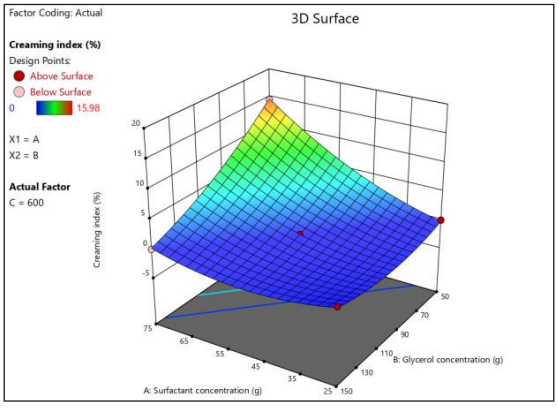


Figure 8b. Response surface and contour plots illustrating the interaction between surfactant concentration and homogenisation pressure on the droplet size of the nanoemulsion

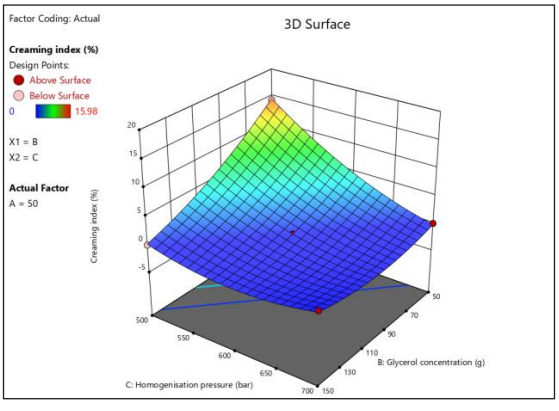


Figure 8c. Response surface and contour plots illustrating the interaction between glycerol concentration and homogenisation pressure on the droplet size of the nanoemulsion

Figure 8c shows that increasing glycerol concentration and homogenisation pressure decreased the creaming index, improving stability. This is mainly due to increased viscosity and density of the water phase, which slows droplet movement and reduces gravitational separation. This agrees with previous studies showing that higher continuous phase viscosity improves emulsion stability (Gupta et al., 2016; Singh et al., 2017). Figure 8a response surface and contour plots illustrating the interaction between surfactant concentration and glycerol concentration on the droplet size of the nanoemulsion

Optimisation Formulation of Oil-in-Water Nanoemulsion

The optimisation process aimed to develop a stable and cost-effective nanoemulsion with minimal surfactant and glycerol concentrations and homogenisation pressure while achieving a small droplet size, low polydispersity index (PDI), and absence of creaming. Based on model prediction, the optimum formulation consisted of 5.06% (w/w) surfactant and 10.4% (w/w) glycerol in the aqueous phase, with a homogenisation pressure of 617.1 bar. Under these conditions, the model predicted a droplet size of 158.2 nm, a PDI of 0.135, and a creaming index of 0.0%.

Experimental validation showed close agreement with the predicted values, producing a nanoemulsion with a droplet size of 151.8 ± 1.3 nm, a PDI of 0.127 ± 0.021, and no observable creaming. The strong correlation between predicted and experimental results confirms the accuracy and reliability of the response surface optimisation model. This agreement indicates that the selected formulation and processing parameters provided adequate interfacial stabilisation and efficient droplet disruption during homogenisation, resulting in a physically stable nanoemulsion system.

The results highlight the role of surfactant concentration in providing interfacial stabilisation, which prevents droplet coalescence during high-pressure homogenisation (Kumar et al., 2021; Singh et al., 2017). Glycerol also contributed to stability by increasing the viscosity and density of the continuous phase, slowing droplet movement and reducing creaming (Gupta et al., 2016). The chosen homogenisation pressure provided sufficient mechanical energy for droplet disruption without causing excessive thermal stress, protecting heat-sensitive bio- actives such as carotenoids and phenolics (Zhou et al., 2023).

Overall, the optimisation demonstrates that a balance among surfactant coverage, continuous-phase properties, and mechanical energy input is critical for producing a physically stable nanoemulsion with uniform droplet size. This approach also ensures cost-effectiveness by avoiding excessive surfactant or energy input while achieving desirable physicochemical characteristics, which is consistent with previous findings in plant-based nanoemulsion systems (Chong et al., 2018; Singh et al., 2017).

Accelerated Stability Studies

Nanoemulsions are thermodynamically unstable systems and therefore require stability evaluation to ensure formulation robustness (De Oca-Ávalos et al., 2017). Conventional stability testing under normal storage conditions is useful but often time-consuming. Accelerated stress testing provides an alternative approach by exposing formulations to extreme conditions, such as pH variation and elevated temperatures, to rapidly assess stability behaviour and identify potential instability mechanisms (Ali et al., 2013).

Stability of Nanoemulsion in Response to pH

pH can induce instability due to changes in the physicochemical properties of the emulsifier and the interfacial structure (Ren et al., 2021). In this study, the nanoemulsion was tested across the pH range of 2 to 6. The data obtained indicated there were no significant changes in droplet size or polydispersity index (PDI), indicating strong resistance to pH fluctuations, as shown in Table 9. These results also showed that the nanoemulsion is structurally stable, which is likely to maintain its quality during storage and use. This stability minimises the risk of emulsion separation or degradation over time, highlighting that the formulation is stable for long-term storage.

This stability is likely due to the protective interfacial layer formed by Tween® 80 and the increased viscosity from glycerol, which limits droplet movement and aggregation (Gupta et al., 2016; Kaur et al., 2024). Maintaining droplet uniformity under varying pH also helps preserve sensitive bioactive compounds like carotenoids and phenolics (Zhou et al., 2023). These results suggest that the nanoemulsion is physically and chemically stable, suitable for storage and applications in environments with varying pH.

Stability of Nanoemulsion in Response to Temperature

Temperature is a highly relevant parameter in nanoemulsion stability because it can influence droplet size, coalescence, and overall physicochemical behaviour. According to

Table 9
Effect of pH values (pH 2 to 6) on mean droplet size, polydispersity index, and cream layer of nanoemulsion

pH values	Mean Droplet Size (nm)	PDI	Presence of Creaming Layer
Control	151.8 ± 1.3^{A}	0.127 ± 0.021^{a}	Absent
2	151.2 ± 0.7^{A}	0.114 ± 0.022^{a}	Absent
3	150.0 ± 1.7^{A}	0.123 ± 0.001^{a}	Absent
4	151.7 ± 0.6^{A}	0.113 ± 0.018^{a}	Absent
5	149.5 ± 1.5^{A}	0.116 ± 0.016^{a}	Absent
6	149.5 ± 2.0^{A}	0.134 ± 0.004^{a}	Absent

Note. Different superscript letters in the same column denote statistical significance ($P < 0.05$), and the control is the untreated optimised nanoemulsion

Fu et al. (2022), elevated temperatures enhance molecular motion and the frequency of droplet collisions, which can promote coalescence and lead to larger droplet sizes. However, as shown in Table 10, the optimised nanoemulsion showed no significant changes in droplet size, PDI, or creaming when exposed to higher temperatures. This thermal stability may be attributed to the effective interfacial stabilisation provided by the surfactant and the viscosity-enhancing effect of glycerol, which reduces droplet mobility and resists coalescence. These findings indicate that the formulation is robust under thermal stress, supporting its potential for long-term storage and practical application.

Stability of Nanoemulsions in Response to Ionic Strength

Ionic strength affects nanoemulsion stability by influencing droplet interactions and emulsifier behaviour. Higher salt concentrations generally reduce electrostatic repulsion, which can promote droplet aggregation and phase separation (Li et al., 2022). In this study, the optimised nanoemulsion remained stable with no significant change in droplet size up to 1.5 mM ionic strength. Interestingly, at 2.0 mM, a slight decrease in droplet size was observed (Table 11), while the PDI remained unchanged, indicating that droplet uniformity was maintained.

This behaviour contrasts with previous reports where higher ionic strength led to droplet growth and instability (Bai & McClements, 2016; Onaizi, 2022; Qian et al., 2011). The slight reduction in droplet size observed here may be due to enhanced surfactant packing at the oil-water interface. Electrolytes can compress the electrical double layer and promote closer packing of surfactant molecules at the oil-water interface, resulting in a more compact droplet structure (Adjonu et al., 2023; Hunter et al., 2020). These results suggest that the nanoemulsion is resilient to moderate changes in ionic strength, which is beneficial for practical applications in different aqueous environments.

Table 10
Effects of heat treatment (30 to 60ºC) on mean droplet size, polydispersity index, and the cream layer of nanoemulsion

Temperature (ºC)	Mean Droplet Size (nm)	PDI	Presence of Creaming Layer
Control	151.8 ± 1.3^{A}	0.127 ± 0.021^{a}	Absent
40	151.3 ± 0.7^{A}	0.111 ± 0.016^{a}	Absent
45	151.7 ± 0.5^{A}	0.121 ± 0.015^{a}	Absent
50	151.0 ± 1.6^{A}	0.116 ± 0.019^{a}	Absent
55	151.4 ± 1.3^{A}	0.105 ± 0.007^{a}	Absent
60	152.2 ± 0.9^{A}	0.108 ± 0.005^{a}	Absent

Note. Different superscript letters in the same column denote statistical significance ($P < 0.05$), and the control is the untreated optimised nanoemulsion

Table 11
Effects of ionic strength (0.5 to 2.5mM) on mean droplet size, polydispersity index, and the cream layer of nanoemulsion

Ionic Strength (mM)	Mean Droplet Size (nm)	PDI	Presence of Creaming Layer
Control	151.8 ± 1.3^{A}	0.127 ± 0.021^{a}	Absent
0.5	150.7 ± 1.0^{A}	0.124 ± 0.014^{a}	Absent
1.0	149.4 ± 0.6^{AB}	0.114 ± 0.004^{a}	Absent
1.5	149.8 ± 1.1^{AB}	0.113 ± 0.018^{a}	Absent
2.0	146.8 ± 0.7^{B}	0.127 ± 0.012^{a}	Absent
2.5	147.5 ± 1.7^{B}	0.105 ± 0.005^{a}	Absent

Note. Different superscript letters in the same column denote statistical significance ($P < 0.05$), and the control is the untreated optimised nanoemulsion

Stability of Nanoemulsion in Response to Centrifugal Speed

Centrifugation is a widely used accelerated method to evaluate nanoemulsion stability by simulating long-term storage effects such as creaming, sedimentation, and coalescence (Preeti et al., 2023; Wu et al., 2020). Applying an increased gravitational force provides rapid insights into a formulation's structural integrity under stress.

In this study, centrifugal speeds up to 3000 rpm had no significant impact on mean droplet size. At 4000 and 5000 rpm, a slight but significant decrease in droplet size was observed (Table 12), while the PDI remained stable across all conditions, indicating consistent droplet uniformity. No creaming or phase separation was detected, highlighting the nanoemulsion's resilience. The observed size reduction may be linked to improved packing or minor reorganisation under stress.

The small droplet size (~150 nm) contributed to strong resistance against coalescence and gravitational separation, supporting the kinetic stability of the nanoemulsion (Li et al., 2025). These findings suggest that the formulation can maintain its physicochemical properties under mechanical stress, which is relevant for handling, transportation, and practical storage conditions.

Table 12
Effects of centrifugal speed (1000 to 5000 rpm) on mean droplet size, polydispersity index, and cream layer of nanoemulsion

Centrifugal Speed (rpm)	Mean Droplet Size (nm)	PDI	Presence of Creaming Layer
Control	151.8 ± 1.3^{A}	0.127 ± 0.021^{a}	Absent
1000	152.1 ± 0.9^{A}	0.135 ± 0.015^{a}	Absent
2000	150.1 ± 1.9^{AB}	0.119 ± 0.009^{a}	Absent
3000	151.9 ± 0.9^{A}	0.121 ± 0.006^{a}	Absent
4000	148.4 ± 0.6^{BC}	0.127 ± 0.020^{a}	Absent
5000	146.2 ± 0.3^{C}	0.121 ± 0.019^{a}	Absent

Note. Different superscript letters in the same column denote statistical significance ($P < 0.05$), and the control is the untreated optimised nanoemulsion

Table 13
Mean of droplet size and polydispersity index for long-term stability study

Month	0.0	2.0	3.0	4.0
Storage at 5°C				
Droplet size (nm)	151.8 ± 1.3$^{A, a}$	152.7 ± 1.4 ^{A,a}	152.1 ± 0.9 ^{A,a}	150.1 ± 0.6 ^{B,a}
PDI	0.127 ± 1.3 ^{A,a}	0.130 ± 0.001 ^{A,a}	0.102 ± 0.005 ^{A,a}	0.127 ± 0.005 ^{A,a}
Storage at 25°C				
Droplet size (nm)	151.8 ± 1.3 ^{A,a}	152.8 ± 1.5 ^{A,a}	152.1 ± 1.0 ^{A,a}	152.7 ± 0.5 ^{A,a}
PDI	0.127 ± 0.021^{A,a}	0.118 ± 0.007 ^{A,a}	0.121 ± 0.009 ^{A,a}	0.098 ± 0.001 ^{B,a}
Storage at 40°C				
Droplet size (nm)	151.8 ± 1.3 ^{A,a}	152.9 ± 1.8 ^{A,a}	152.9 ± 0.6 ^{A,a}	151.5 ± 1.3 ^{AB,a}
PDI	0.127 ±0.021 ^{A,a}	0.134 ± 0.012 ^{A,a}	0.107 ± 0.005 ^{A,a}	0.125 ± 0.013 ^{A,a}

Note. Different superscript letters in the same column denote statistical significance ($P < 0.05$)

Long-term Stability Study

Table 13 shows that the storage temperature and duration affect the mean droplet size of the nanoemulsion. The nanoemulsion had a consistent droplet size at the start (zeroth month). After three months, the droplet size remained stable across all temperatures. However, the droplet size at 25°C was noticeably smaller by the fourth month than at the other temperatures. The table also shows the PDI over the four months at 5°C, 25°C, and 40°C. No significant changes in PDI were seen during the first three months at any temperature. However, after four months, the PDI at 25°C was significantly lower than at the other temperatures. These results and the mean droplet size data suggest that the nanoemulsion remained physically stable for four months, confirming the stability of the RSM-optimised formulation, as there were no signs of instability, such as Ostwald ripening, coalescence, flocculation, or creaming (Kampa et al., 2022).

CONCLUSION

In this study, a red palm oil-based oil-in-water nanoemulsion containing *Strobilanthes crispus* extract was successfully optimised using response surface methodology. The optimum formulation consisted of 5.06% (w/w) Tween® 80, 10.4% (w/w) glycerol, and a homogenisation pressure of 617.1 bar, producing a nanoemulsion with a droplet size of 151.8 ± 1.3 nm, PDI of 0.127 ± 0.021, and no observable creaming.

Accelerated and four-month storage stability studies confirmed consistent physicochemical stability across varying pH, ionic strength, temperature, and centrifugation conditions. These findings indicate that the developed nanoemulsion is a stable delivery system for lipophilic bioactive compounds.

Further studies such as biological activity evaluation, skin penetration, and advanced characterisation (e.g., zeta potential and interfacial tension) can be explored in future work.

ACKNOWLEDGEMENT

This study was supported by the Technology Development Fund 2 (TED 2), Ministry of Science, Technology and Innovation (Grant No. TEF09211087). The authors would like to express their sincere gratitude to Aurora Health Sdn. Bhd., Lipidware Sdn. Bhd., and Universiti Putra Malaysia for their technical support and assistance throughout this study.

REFERENCES

Adjonu, R., Doran, G., Torley, P., & Agboola, S. (2023). Stability of whey protein bioactive peptide-stabilised nanoemulsions: Effect of pH, ions, heating and freeze-thawing. *International Journal of Food Science and Technology, 58*(4), 1787-1794. https://doi.org/10.1111/ijfs.16292

Ali, M. S., Alam, M. S., Alam, N., Anwer, T., & Safhi, M. M. A. (2013). Accelerated stability testing of a clobetasol propionate-loaded nanoemulsion as per ICH guidelines. *Scientia Pharmaceutica, 81*(4), 1089-1100.https://doi.org/10.3797/scipharm.1210-02

Bai, L., & McClements, D. J. (2016). Formation and stabilisation of nanoemulsions using biosurfactants: Rhamnolipids. *Journal of Colloid and Interface Science, 479*, 71-79. https://doi.org/10.1016/j.jcis.2016.06.047

Bakhaidar, R. B., Naveen, N. R., Basim, P., Murshid, S. S., Kurakula, M., Alamoudi, A. J., Bukhary, D. M., Jali, A. M., Majrashi, M. A., Alshehri, S., Alissa, M., & Ahmed, R. A. (2022). Response surface methodology (RSM)-powered formulation development, optimisation and evaluation of thiolated-based mucoadhesive nanocrystals for local delivery of simvastatin. *Polymers, 14*(23), 5184. https://doi.org/10.3390/polym14235184

Chandel, A., Xavier, J. R., & Chauhan, O. P. (2025). Applications of high-pressure homogenisation in the food industry for ensuring quality and safety. *Journal of Food Process Engineering, 48*(4). https://doi.org/10.1111/jfpe.70105

Chong, W. T., Tan, C. P., Cheah, Y. K., & Lai, O. M. (2022). In vitro and in vivo evaluations of tocotrienol-rich nanoemulsified system on skin wound healing. *PLoS ONE, 17*(5), e0267381. https://doi.org/10.1371/journal.pone.0202771

Chong, W., Tan, C., Cheah, Y., Lajis, A. F. B., Dian, N. L. H. M., Kanagaratnam, S., & Lai, O. (2018). Optimisation of process parameters in preparation of tocotrienol-rich red palm oil-based nanoemulsion stabilised by Tween 80-Span 80 using response surface methodology. *PLoS ONE, 13*(8), e0202771. https://doi.org/10.1371/journal.pone.0202771

Chow, P. S., Lim, R. T. Y., Cyriac, F., Shah, J. C., Badruddoza, A. Z. M., Yeoh, T., Yagnik, C. K., Tee, X. Y., Wong, A. B. H., Chia, V. D., & Wang, G. (2024). The effect of process parameters on the microstructure, stability, and sensorial properties of an emulsion cream formulation. *Pharmaceutics, 16*(6), 773. https://doi.org/10.3390/pharmaceutics16060773

Chua, L. Y. W., Chua, B. L., Figiel, A., Chong, C. H., Wojdyło, A., Szumny, A., & Choong, T. S. Y. (2019). Antioxidant activity and volatile and phytosterol contents of *Strobilanthes crispus* dehydrated using conventional and vacuum microwave drying methods. *Molecules, 24*(7), 1397. https://doi.org/10.3390/molecules24071397

De Oca-Ávalos, J. M. M., Candal, R. J., & Herrera, M. L. (2017). Nanoemulsions: Stability and physical properties. *Current Opinion in Food Science, 16*, 1-6. https://doi.org/10.1016/j.cofs.2017.06.003

Fu, Y., Xiao, S., Liu, S., Chang, Y., Ma, R., Zhang, Z., & He, J. (2022). Atomistic insights into the droplet size evolution during self-microemulsification. *Langmuir, 38*(10), 3129-3138. https://doi.org/10.1021/acs.langmuir.1c03099

Galvão, K. C. S., Vicente, A. A., & Sobral, P. D. A. (2018). Development, characterisation, and stability of O/W pepper nano-emulsions produced by high-pressure homogenisation. *Food and Bioprocess Technology, 11*, 355-367. https://doi.org/10.1007/s11947-017-2016-y

Gawin-Mikołajewicz, A., Nawrot, U., Malec, K. H., Krajewska, K., Nartowski, K. P., & Karolewicz, B. L. (2023). The effect of high-pressure homogenisation conditions on the physicochemical properties and stability of designed fluconazole-loaded ocular nano-emulsions. *Pharmaceutics, 16*(1), 11. https://doi.org/10.3390/pharmaceutics16010011

Hunter, S. J., Cornel, E. J., Mykhaylyk, O. O., & Armes, S. P. (2020). Effect of salt on the formation and stability of water-in-oil Pickering nano-emulsions stabilised by diblock copolymer nanoparticles. *Langmuir, 36*(51), 15523-15535.https://doi.org/10.1021/acs.langmuir.0c02742

Iacob-Tudose, E. T., Mamaliga, I., & Iosub, A. V. (2021). TES nano-emulsions: A review of thermophysical properties and their impact on system design. *Nanomaterials, 11*(12), 3415. https://doi.org/10.3390/nano11123415

Imam, S. S. (2023). Topical formulation constituted with transferosomes for the treatment of non-melanoma skin cancer. *Asian Journal of Pharmaceutical and Clinical Research*, 27-32. https://doi.org/10.22159/ajpcr.2023.v16i5.47033

Kampa, J., Koidis, A., Ghawi, S. K., Frazier, R. A., & Rodriguez-Garcia, J. (2022). Optimisation of the physicochemical stability of extra virgin olive oil-in-water nano-emulsion: Processing parameters and stabiliser type. *European Food Research and Technology, 248*(11), 2765-2777. https://doi.org/10.1007/s00217-022-04088-7

Kaur, G., et al. (2024). Nanoemulsion delivery systems. *Colloids and Surfaces B, 233*, 113456. https://doi.org/10.1016/B978-0-12-804306-6.00009-X

Kumar, M., Bishnoi, R. S., Shukla, A. K., & Jain, C. P. (2019). Techniques for formulation of nanoemulsion drug delivery system: A review. *Preventive Nutrition and Food Science, 24*(3), 225-234. https://doi.org/10.18231/j.ijmr.2022.042

Kumar, R., & Reji, M. (2023). Response surface methodology (RSM): An overview to analyse multivariate data. *Indian Journal of Microbiology Research, 9*(4), 241-248. https://doi.org/10.18231/j.ijmr.2022.042

Li, B., Ju, M., Dou, X., Li, N., Zhang, W., Sun, Z., Yu, K., Wang, J., & Wang, Z. (2022). Assessing nanoparticle-surfactant-salt synergistic effects on droplet-droplet electrocoalescence by molecular dynamics simulations. *Journal of Molecular Liquids, 367*, 120570. https://doi.org/10.1016/j.molliq.2022.120570

Li, Z., Kong, Y., Ma, Z., Zhao, X., Liu, J., Yang, X., Wang, L., Zhou, B., Xing, Y., & Gui, X. (2025). New insights into the role of interfacial components in oil-in-water nanoemulsion collector for low-rank coal flotation. *Physics of Fluids, 37*(9). https://doi.org/10.1063/5.0285726

Loke, C., Tan, S., Lim, S., Law, C., Yue, C., Poh, T., Saad, W., Ismail, S., & Yusoff, K. (2019). Antioxidative and photocytotoxic effects of standardised *Clinacanthus nutans* and *Strobilanthes crispus* extracts toward HepG2 liver cells. *Pharmacognosy Magazine, 15*(65), 613. https://doi.org/10.4103/pm.pm_84_19

Lüdtke, F. L., Stahl, M. A., Grimaldi, R., Forte, M. B. S., Gigante, M. L., & Ribeiro, A. P. B. (2022). Optimisation of high-pressure homogenisation conditions to produce nanostructured lipid carriers using natural and synthetic emulsifiers. *Food Research International, 160*, 111746. https://doi.org/10.1016/j.foodres.2022.111746

Luu, N. A., To, T. T., Tran, N. H., Vu, G. M., Ngo, N. Y. L., Tran, M. P. L., Nguyen, N. M., Dinh, T. N. N., Dam, A. N., Lieu, N. N., & Nguyen, D. Q. (2025). Optimisation of plant oil-based nano-emulsions prepared via phase inversion temperature and evaluation of their antioxidant capacity. *Oil Crop Science, 10*(3), 194-204. https://doi.org/10.1016/j.ocsci.2025.02.004

Marhamati, M., Ranjbar, G., & Rezaie, M. (2021). Effects of emulsifiers on the physicochemical stability of oil-in-water nano-emulsions: A critical review. *Journal of Molecular Liquids, 340*, 117218. https://doi.org/10.1016/j.molliq.2021.117218

Mohammed, N. K., Muhialdin, B. J., & Meor Hussin, A. S. (2020). Characterisation of nanoemulsion of *Nigella sativa* oil and its application in ice cream. *Food Science & Nutrition, 8*(6), 2608-2618. https://doi.org/10.1002/fsn3.1500

Mukherjee, P., & Baruah, K. N. (2025). Characterisation of nano-emulsion: Particle size, polydispersity index, zeta potential, thermal stability, and kinetic stability. In *Methods and protocols in food science* (pp. 273-288). https://doi.org/10.1007/978-1-0716-4634-2_17

Onaizi, S. A. (2022). Effect of salinity on the characteristics, pH-triggered demulsification and rheology of crude oil/water nano-emulsions. *Separation and Purification Technology, 281*, 119956. https://doi.org/10.1016/j.seppur.2021.119956

Podder, S., & Mukherjee, S. (2024). Response surface methodology (RSM) as a tool in pharmaceutical formulation development. *Asian Journal of Pharmaceutical and Clinical Research*, 18-25. https://doi.org/10.22159/ajpcr.2024v17i11.52149

Pongsumpun, P., Iwamoto, S., & Siripatrawan, U. (2020). Response surface methodology for optimisation of cinnamon essential oil nanoemulsion with improved stability and antifungal activity. *Ultrasonics Sonochemistry, 60*, 104604. https://doi.org/10.1016/j.ultsonch.2019.05.021

Preeti, Sambhakar, S., Malik, R., Bhatia, S., Harrasi, A. A., Rani, C., Saharan, R., Kumar, S., Geeta, & Sehrawat, R. (2023). Nano-emulsion: An emerging novel technology for improving the bioavailability of drugs. *Scientifica, 2023*, 1-25. https://doi.org/10.1155/2023/6640103

Preiss, F. J., Hetz, M., & Karbstein, H. P. (2021). Does cavitation affect droplet breakup in high-pressure homogenisation? Insights into local effects. *Chemie Ingenieur Technik, 94*(3), 374-384. https://doi.org/10.1002/cite.202100104

Qian, C., Decker, E. A., Xiao, H., & McClements, D. J. (2011). Physical and chemical stability of β-carotene-enriched nano-emulsions: Influence of pH, ionic strength, temperature, and emulsifier type. *Food Chemistry, 132*(3), 1221-1229. https://doi.org/10.1016/j.foodchem.2011.11.091

Rauber, G. S., Argenta, D. F., Caon, T., & Poláková, K. (2025). High-pressure homogenisation techniques for the production of nanomaterials. In *Smart nanomaterials technology* (pp. 135-163). https://doi.org/10.1007/978-981-96-9176-0_8

Ren, G., Li, B., Ren, L., Lu, D., Zhang, P., Tian, L., Di, W., Shao, W., He, J., & Sun, D. (2021). pH-responsive nano-emulsions based on a dynamic covalent surfactant. *Nanomaterials, 11*(6), 1390. https://doi.org/10.3390/nano11061390

Rodriguez-Loya, J., Lerma, M., & Gardea-Torresdey, J. L. (2023). Dynamic light scattering and its application to control nanoparticle aggregation in colloidal systems: A review. *Micromachines, 15*(1), 24. https://doi.org/10.3390/mi15010024

Sarheed, O., Dibi, M., & Ramesh, K. V. R. N. S. (2020). Studies on the effect of oil and surfactant on the formation of alginate-based O/W lidocaine nanocarriers using nanoemulsion template. *Pharmaceutics, 12*(12), 1223. https://doi.org/10.3390/pharmaceutics12121223

Shakeel, F., Salem-Bekhit, M. M., Haq, N., & Alshehri, S. (2021). Nanoemulsification improves the pharmaceutical properties and bioactivities of niaouli essential oil (*Melaleuca quinquenervia* L.). *Molecules, 26*(16), 4750. https://doi.org/10.3390/molecules26164750

Singh, Y., Meher, J. G., Raval, K., Khan, F. A., Chaurasia, M., Jain, N. K., & Chourasia, M. K. (2017). Nano-emulsion: Concepts, development and applications in drug delivery. *Journal of Controlled Release, 252*, 28-49. https://doi.org/10.1016/j.jconrel.2017.03.008

Tan, C. H., Lee, C. J., Tan, S. N., Poon, D. T. S., Chong, C. Y. E., & Pui, L. P. (2021). Red palm oil: A review on processing, health benefits and its application in food. *Journal of Oleo Science, 70*(9), 1201-1210. https://doi.org/10.5650/jos.ess21108

Wu, H., Wang, C., Wang, J., Chen, J., & Le, Y. (2020). Engineering of long-term stable transparent nanoemulsion using high-gravity rotating packed bed for oral drug delivery. *International Journal of Nanomedicine, 15*, 2391-2402. https://doi.org/10.2147/IJN.S238788

Xu, Y., Wang, G., Wang, X., Yu, J., Wang, J., Zhang, Z., & Li, R. (2018). Effects of homogenisation on the molecular flexibility and emulsifying properties of soy protein isolate. *Food Science and Biotechnology, 27*(5), 1293-1299. https://doi.org/10.1007/s10068-018-0361-x

Zhou, H., Qin, D., Vu, G., & McClements, D. J. (2023). Impact of operating parameters on the production of nano-emulsions using a high-pressure homogeniser with flow pattern and back pressure control. *Colloids and Interfaces, 7*(1), 21. https://doi.org/10.3390/colloids7010021

Zhou, L., Zhang, W., Wang, J., Zhang, R., & Zhang, J. (2022). Comparison of oil-in-water emulsions prepared by ultrasound, high-pressure homogenisation and high-speed homogenisation. *Ultrasonics Sonochemistry, 82*, 105885. https://doi.org/10.1016/j.ultsonch.2021.105885

Zhu, X., Xu, Y., Sun, D., Li, H., & Chen, L. (2022). The genus *Strobilanthes*: Phytochemistry and pharmacology. *TMR Modern Herbal Medicine, 5*(3), 15. https://doi.org/10.53388/MHM2022B0701001

Pertanika J. Sci. & Technol. 34 (2): 889 - 911 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

OD-SAM: Automated Zero-shot Segment Anything Model for Optic Disc Segmentation

Bhuvaneswari S* and Subashini P

Department of Computer Science, Centre for Machine Learning and Intelligence, Avinashilingam Institute for Home Science and Higher Education for Women, 641043 Coimbatore, Tamil Nadu, India

ABSTRACT

The severe stage of diabetic retinopathy (DR) is often identified by the presence of hard exudates on ophthalmologic examination. For precisely detecting exudates, Optic Disc (OD) segmentation is significant due to the high similarity between exudates and the OD. The primary objective of this OD-SAM study is to examine the feasibility of a zero-shot framework for OD segmentation, intending to improve subsequent hard exudate detection in retinopathy screening. The proposed work integrates automatic OD localisation using the peak-end thresholding approach, prompt-based optic disc segmentation using the segment anything model (SAM), a multi-criteria decision-making (MCDM) approach for optimal OD mask selection, and ellipse fitting to smooth the disc boundary. Subsequently, the OD is removed to evaluate its impact on YOLO-based hard exudate detection. The efficiency of this proposed framework is tested on the IDRiD dataset. It achieved an 86.2% overlap and a 90.7% dice coefficient. The results show the effectiveness of the proposed approach for OD segmentation and highlight improvements in exudate detection after OD removal. This study can support lesion analysis and grading in an automated diabetic retinopathy screening and severity assessment.

Keywords: Deep learning, diabetic retinopathy, hard exudate, localisation, optic disc, segment anything model, segmentation

ARTICLE INFO

Article history:
Received: 21 July 2025
Accepted: 17 March 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.13

E-mail addresses:
bhuvaneswari_cs@avinuty.ac.in (Bhuvaneswari S)
subashini_cs@avinuty.ac.in (Subashini P)
*Corresponding author

INTRODUCTION

Diabetes can affect various parts of the body and cause serious complications, including diabetic retinopathy (Halder et al., 2020), which affects the retina of the eye. This condition weakens the light-sensing cells and breaks the connections between retinal cells, damaging blood vessels and

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

breaking down the protective layer. This serious condition leads to vision impairment (Antonetti et al., 2021). However, with timely diagnosis and treatment, the chance of developing vision loss can be reduced (Valverde et al., 2016). Retinopathy screening using a retinal fundus image is a time-saving and cost-effective approach (Teoh et al., 2023). The severity of retinopathy is classified as non-proliferative (NPDR) or proliferative diabetic retinopathy (PDR) (Usman Akram et al., 2014), depending on the extent of retinal lesions, microaneurysms, haemorrhages, hard and soft exudates, and cotton wool spots. Exudates and cotton-wool spots are identified as bright lesions, whereas haemorrhages and microaneurysms are identified as dark lesions (Kumar et al., 2020). Therefore, the type and location of the lesions play an important role in grading.

Hard exudates are considered an important indicator among the factors used to assess disease severity (Liu et al., 2025). The International Clinical Diabetic Retinopathy (ICDR) scale also includes a factor of hard exudates to indicate the severity of the disease (Cleland, 2023). An automated screening system should be developed that can effectively identify hard exudates to help diagnose the disease and initiate treatment at the right time (Joshi & Karule, 2018). Since the brightness of OD is almost identical to that of hard exudate, it is very important to remove OD before detecting exudate (Theera-Umpon et al., 2019). Figure 1 shows that OD can be misidentified due to its similar appearance while identifying exudates. Therefore, removing the OD is very helpful in increasing the accuracy of hard exudate identification (Fu et al., 2023).

There are numerous techniques for segmenting OD from fundus images. However, OD is not easily identifiable due to its low contrast relative to the retinal region, its variability in appearance, and its blood vessels, making it extremely difficult to automatically determine the disc's boundaries at the pixel level (Motevali et al., 2021). Image segmentation is crucial for identifying regions and object edges within an image. However, effective segmentation requires suitable image preprocessing, which simplifies the data, removes unwanted data,

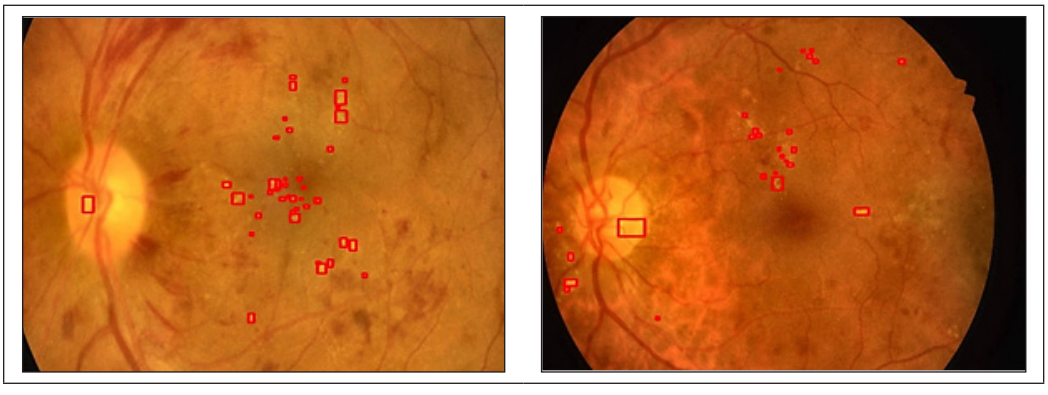


Figure 1. False positive prediction of hard exudate

retrieves important information, and creates an annotated dataset, which is crucial for deep learning-based image segmentation (Chen et al., 2024). Nevertheless, the effectiveness of supervised learning models is currently constrained by the limited availability of manually labelled image data (Hasan et al., 2021).

Inspired by the zero-shot segmentation method, in the present work, the SAM (Kirillov et al., 2023) is proposed to segment the OD from fundus images to address the lack of an annotated dataset and improve OD region segmentation. The procedure for the work carried out is given in Figure 2.

The following are the study's primary contributions:

- The segment anything model is proposed in this study to overcome the limitations of sparsely represented datasets and improve OD segmentation. It performs based on the vision transformer for segmenting the regions.
- SAM was originally developed for problems in real-world image segmentation. The multi-criteria decision-making (MCDM) technique has been used to apply this SAM approach to medical image segmentation. The goal of this adaptation is to address the challenges of both over- and under-segmentation.
- The proposed work integrates the shape-based ellipse fitting (Stojmenovic & Nayak, 2007) approach to enhance the OD boundary. Ellipse fitting helps smooth jagged boundary edges caused by imperfections in the segmentation process.

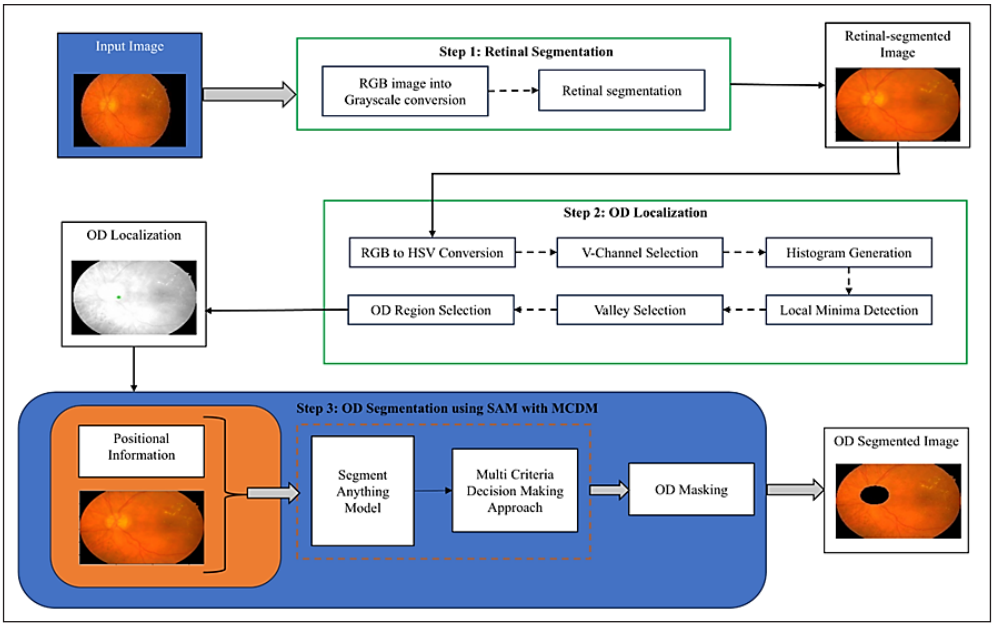


Figure 2. The general pipeline of the work

This paper is structured as follows. Section 2 contains the literature based on OD segmentation. The materials and methods used for this study are given in Section 3. Section 4 presents the experimental results obtained from this study and the discussions. Section 5 contains the conclusion.

RELATED WORK

This section describes works that support OD segmentation using traditional image processing approaches and deep learning techniques.

Traditional Image Processing Approaches for OD Segmentation

In their study, Dharmawan et al. (2020) proposed a new optic disc localiser method. The template-based, map-based vessel density and maximum entropy-based methods have been used to find the centre of the optic disc. Additionally, the modified Dolph-Chebyshev type I function is designed to detect boundary conditions. Finally, circular Hough transform, polar transform, and B-spline approximation are used to obtain the final OD boundary. Testing the proposed approach on the MESSIDOR and DRIVE datasets yielded overlap accuracies of 89.5% and 87.3%, respectively. The proposed method is robust to variations in image quality. But the fixed template match may not adapt well to significant variance in OD appearance. The paper (Ramani & Shanthamalar, 2020) presents an improved image-processing algorithm for automated optic disc segmentation to detect glaucoma early. For this purpose, an improved region-based pixel density calculation method for localisation, and circular Hough transform with Hough peak and red channel superpixel segmentation methods have been proposed to segment the optic disc. The model achieved segmentation accuracies of 99.2% to 99.73% across datasets. The proposed method is tested across multiple datasets and is computationally fast. Even though it is less flexible compared to DL methods. A study by Dietter et al. (2019) proposed a new paradigm to detect and segment the optic disc, as it is a key indicator in retinal image analysis and retinal disease detection. To deal with the strong technical artefacts during OD detection, a two-stage methodology has been proposed. First, the vessel's orientation and brightness are determined to locate the vessel. The modified score function was then used to detect the best OD border. Experimental results show that the best segmentation accuracy is across different datasets. The proposed method is robust against artefacts, but the segmentation performs poorly on distorted vessel structures. Abdullah et al. (2020) presented a new approach for optical disc localisation and segmentation. In their approach, the median filter and Otsu have been used to accurately determine the OD location. The red channel has then been used to remove the veins, and morphological operations have been applied to reduce noise. From the resulting optic disc region, fuzzy clustering, the mean-shift algorithm, the active contour model, and ellipse fitting were used to segment the disc.

The methods on the DRIVE, DIRATEDB1, DRIONS-DB, and STARE datasets achieved segmentation accuracies of 84.3%, 84.56%, 81.01%, and 78.35%, respectively. The proposed method helps handle noise and structural irregularities. Due to severe pathologies and poor image contrast, segmentation accuracy was reduced. Since OD is an indicator for diagnosing the macula and is very important in many analyses of retinal disease, Xue et al. (2022) have proposed a new hybrid level set model to segment the optic disc. For this purpose, the area-based term has been used to detect the average pixel value difference between the inside and outside of a contour. The shape-based term has also been used to identify the prior shape model and the contour. Experimental results showed that the proposed method achieved an overlap accuracy of 92.75% and 81.79% on the DRISHTI-GS and TMUEH datasets, respectively. This approach is robust to handle parapapillary atrophy (PPA) and bright artefacts. The proposed method heavily depends on the preprocessing quality and non-deep learning methods.

Deep Learning Approaches for OD Segmentation

Semi-supervised and transfer learning are proposed by Bengani et al. (2020) for automated OD segmentation in glaucoma detection. The unlabelled retinal images were used for convolutional autoencoder training, and the trained network was converted to a segmentation network. This network was then trained on retinal images using the OD ground-truth image, leveraging transfer learning. The segmentation study on the DRISHTI GS1 and RIM-ONE datasets reported a Jaccard score of 93.14% and 88.24%, respectively. The proposed method is effective with limited labelled data. But it requires a high convolutional autoencoder (CAE) training. Septiarini et al. (2023) have proposed to automatically segment the optic disc needed to diagnose glaucoma. To do this, they used two convolutional neural network architectures: the single-shot multibox detector and MobileNetV2 to segment the OD area. Before this, preprocessing such as augmentation, normalisation, and resizing was applied. Experimental results indicated that the proposed method is robust in handling diverse image quality and pathology. It achieved an overlap accuracy of 97.63% on the private dataset and 97.12% on the REFUGE dataset, but it suffers from over-segmentation with PPA and requires high computational time for large datasets. Sevastopolsky (2017) presented the modified U-Net convolutional model for automated optic disc and cup segmentation for optic nerve head examination. This modified model has fewer filters, allowing it to be lightweight. The approach achieved an overlap accuracy of 89% and 89% on datasets such as DRIONS-DB and RIM-ONE v.3. It was tested across multiple public datasets. However, the optic cup is highly sensitive, posing a challenge and limitation for this study. In a study by Zhang et al. (2021), a transferable attention U-Net model was proposed for optic disc and cup segmentation in cup-to-disc ratio-based glaucoma detection. Two discriminators and attention modules have been

applied to locate and extract invariant features from the datasets. The method used an attention mechanism and an adversarial training method to address the domain adaptation problem. Experimental results reported a higher dice score across datasets. The proposed approach is robust in handling different camera resolutions, contrasts, and field-of-view, even though it requires a labelled source domain and more complex adversarial training.

Yin et al. (2021) have proposed a level-set-based deep learning model to perform OD and OC segmentation with weak boundaries. This combination uses a level set loss function to achieve smooth, segmented boundaries and consistent regions. To reduce ambiguity, joint region-based CNNs, which assume the disc and cup are approximately elliptical, and M-Net knowledge, which shows that the optic cup is always inside the OD, are used to improve segmentation. It reports superior performance consistently. The usage of two separate computationally intensive networks is a limitation of the study.

From the literature study, it has been observed that the majority of optic disc segmentation studies have largely focused on glaucoma detection because the diagnosis of glaucoma is particularly dependent on the unique characteristics of the OD region. OD segmentation is also vital for diabetic retinopathy tasks to prevent false positives when recognising exudates. Thus, the current study seeks to bridge this gap in hard exudate detection in diabetic retinopathy, particularly in the context of OD localisation and segmentation. Furthermore, this work presents an effective solution to eliminate reliance on templates and reduce the need for large, annotated datasets, which are commonly encountered in existing image processing studies and in most deep learning techniques for OD segmentation.

MATERIALS AND PROPOSED METHODOLOGY

This section gives a detailed description of the dataset and the proposed method. SAM is a prompt-driven model. Therefore, it requires prompts to segment objects (Mazurowski et al., 2023). Thus, the study primarily focuses on identifying the optic disc to obtain the location data using traditional image processing techniques. Later, the SAM model uses the obtained location data and the input image to separate the OD from the retinal images.

Dataset Collection

The study obtained 81 colour fundus photographs from the publicly available IDRiD benchmark dataset (Porwal et al., 2018). For the segmentation study, the dataset gives 54 images for training purposes and 27 images for testing purposes with their ground-truth images. Most other popular datasets for diagnosing diabetic retinopathy usually give each image a severity grade or image-level annotation but lack precise annotation information (Chincholi & Koestler, 2023). However, the IDRiD database provides both ground-truth images and grading information, all of which were assessed by medical professionals.

Retinal Segmentation

Before OD localisation, the retinal region is segmented from the fundus images because images from the IDRiD database contain unwanted background information. Thus, the bright retinal region is segmented from the background to assist the model in concentrating on the important information in the foreground. Then, the segmented retinal input image (I_{BGR}) is converted into the HSV colour space (I_{HSV}) to isolate the Value (V) channel image for further analysis, as brightness information is more viable in the V channel (Nagpal et al., 2024). The splitting can be expressed as shown in Equation 1: -

$$(I_H,\ I_S,\ I_V) = \text{split}(I_{HSV}) \quad [1]$$

Optic Disc Localisation using Peak-end Thresholding Approach

Following retinal segmentation, the peak-end thresholding method is used on v-channel images for locating the optic disc. The goal of the OD localisation is to provide a reference point for the OD, which is then used to subdivide the OD area with the SAM. Several studies investigate the use of retinal vessels for optic disc localisation. However, vessel segmentation-based methods can lead to increased computational complexity in the OD localisation process (Kamble et al., 2017). For OD localisation, this work used a peak-end thresholding technique that combined connected component analysis with an inverted histogram. Figure 3 shows the OD localisation approach.

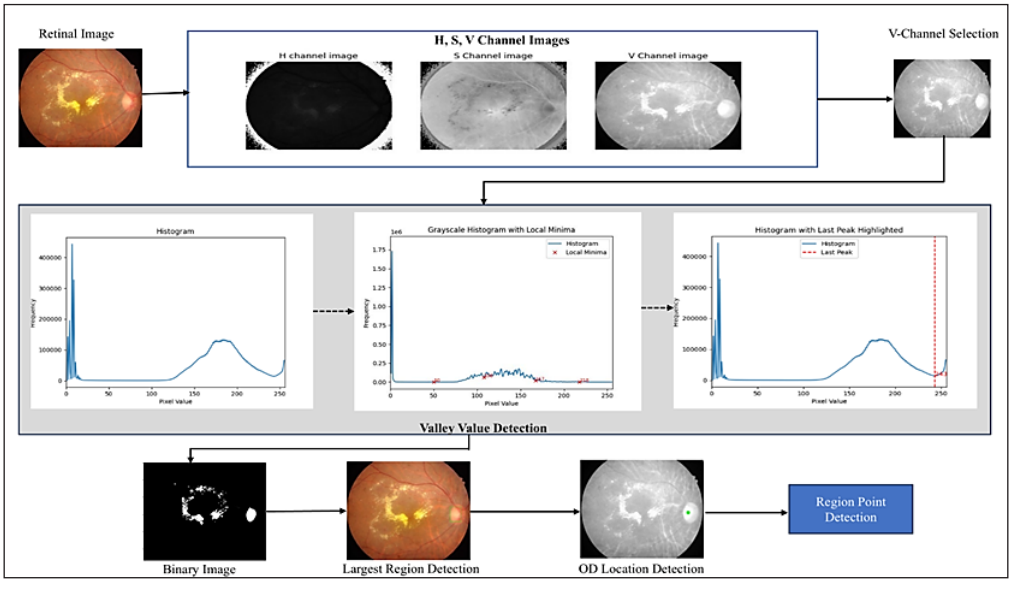


Figure 3. OD localisation

The OD is generally regarded as a bright area within the image. Histogram inversion and local minima detection techniques help identify the bright area associated with the optic disc in retinal images. To achieve this, a technique that involves detecting local minima in the histogram of the image is employed. The histogram inversion transforms the bright region of the OD into a local minimum, allowing the positions in the inverted histogram to be located. The detected minima are separated by a minimum distance to avoid false detections of closely spaced features. Once the local minima are identified, the value corresponding to the most significant minimum, typically the last detected peak, is used for further processing. Subsequently, this threshold value is employed in binary segmentation to isolate the bright OD region.

This approach, however, can sometimes yield false positives, as it may also capture high-intensity exudates. To accurately locate the OD region and distinguish it from other bright features, the largest connected component analysis is used. Because typically OD exhibits a larger area compared to other features. This approach effectively identifies the OD by assessing the size and shape of the connected components. Then, the positional information of an OD is obtained from the largest component region. In some situations, misvocalisation is occasionally possible due to large exudate clusters.

Proposed Optic Disc Segmentation using SAM

This work utilised a zero-shot learning SAM method for OD segmentation. Typically, the SAM prompt encoder, which provides positional encodings, requires real-time interactions. However, in this study, the process is automated through the prior OD localisation approach to eliminate the need for manual intervention (Guan et al., 2025). The optic disc usually appears bright, but this is not constant throughout. Due to factors such as differences in illumination, blood vessel darkness, or the fundamental anatomic structure of the optic disc, certain regions may look brighter than others. Effectively recognising and segmenting the optic disc might be difficult due to the uneven brightness in image processing. Therefore, a zero-shot learning SAM is proposed to segment OD from retinal images, ensuring effective region detection and segmentation, even when some regions are less bright than others.

Network Architecture

This study used the promptable (Misera et al., 2026) segmentation network (SAM), built on transformer-based vision models. The vision transformer serves as a backbone network in the SAM to extract features from the image. The model generates feature maps at multiple scales using a Feature Pyramid Network (FPN) to identify objects at different levels and their boundaries. The transformer-based design uses contextual information from the input image to improve the segmentation result. To do this, the SAM model is integrated with a prompt encoder that effectively extracts and combines the input image's spatial and

semantic elements to process positional encoding along with free-form text and encoding specific elements, a Masked Auto-encoder (MAE), and an image encoder that transforms the input image into image embeddings.

Enhanced transformer decoder blocks help to focus on significant areas within the inputs via self-attention and cross-attention mechanisms. To enhance segmentation precision, the model combines information from both the prompts and the image by utilising the enhanced transformer decoder blocks. The self-attention enables a focus on the most important image parts, and cross-attention helps to establish an interface between input embeddings and prompt embeddings. After the self-attention and cross-attention processes have been carried out, the generated embeddings are processed by the mask decoder. The image embeddings are then up-sampled by the model to provide a final mask with the same resolution as the original image. This up-sampled embedding and output token are used by the dynamic mask prediction head to predict the mask's probability. On the other hand, a single prompt presents many masks, such as the object's part, a smaller sub-part, and the entire object. For that, the model evaluates the generated masks according to a confidence score using overlap measures. Utilising the most appropriate and precise mask, it reduces the chance of inaccurate segmentation. The overview of the segmentation process using the SAM model is illustrated in Figure 4a.

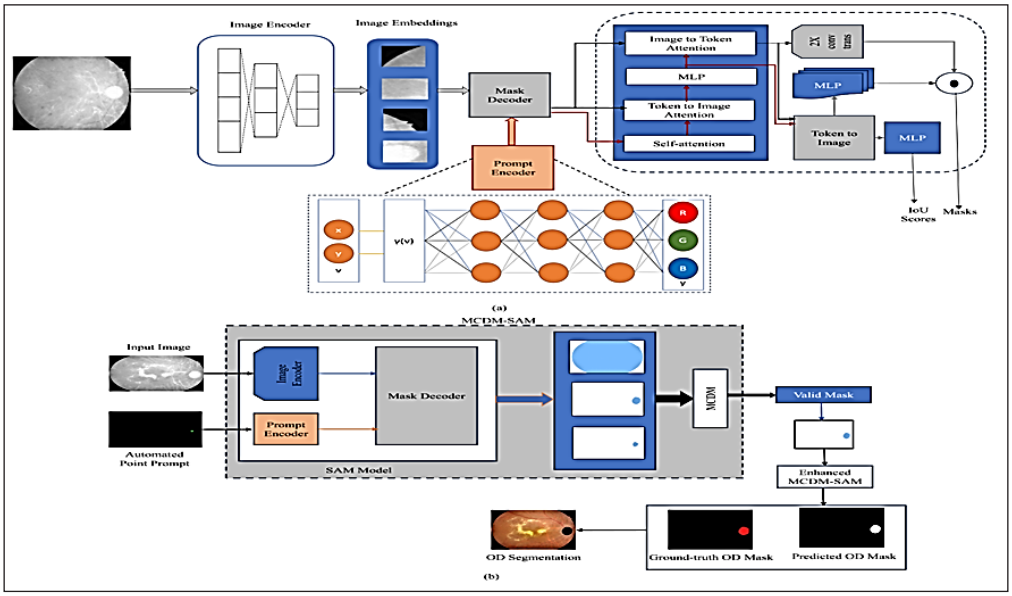


Figure 4. (a) Illustration of the OD segmentation process in the SAM architecture; (b) OD segmentation in the MCDM-SAM

Optic Disc Segmentation

After the above-mentioned retinal segmentation and OD localisation process, the image is converted into the desired colour channel, and then the converted image is fed into the proposed SAM model with the positional information. Positional information obtained from the OD localisation is assigned to the SAM model as a point prompt. Although the models used in this work are well-established, the current study focuses on their innovative usage in retinal segmentation, particularly in medical image segmentation. Figure 4b provides an overview of the suggested OD segmentation method. This proposed approach ensures precise and efficient segmentation, streamlining the process without requiring real-time human input.

The SAM model, which is initialised with the ViT-B, is given the input image. The image encoder of the SAM is specially designed with the Vision Transformer (ViT) to handle more complex and volumetric data, and it is scaled from ViT-B (Base) to ViT-H (Huge). In comparison with ViT-B, ViT-H can capture more significant and detailed features. Since it modestly improves accuracy and substantially raises processing requirements (Kirillov et al., 2023). Therefore, for computational efficiency, the ViT-B model was used as the image encoder in this study, although ViT-L and ViT-H also exhibit considerable performance. After processing the given V-channel image, the pre-trained Vision Transformer with the MAE approach He et al. (2022) creates a 64x64 embedding that preserves rich, downscaled information from the high-resolution input. SAM needs positional information in addition to the image to generate a mask on it. To analyse the positional data, SAM provides a prompt encoder.

The prompt encoder supports real-time user interaction. Instead of real-time interactions, the points that are identified automatically during the OD localisation process are transferred to the prompt encoder in the present study. These predetermined points, encoded by their positions and associated foreground labels, will serve as prompts to guide the segmentation process. The prompt encoder will translate these fixed coordinates into embeddings that precisely match the image's features accordingly. The mask decoder will predict the final segmentation mask using a transformer-based architecture that integrates the prompt token and image embeddings. The model generates multiple segmentation masks to capture different interpretations and address an issue in a single prompt. During interpretation, the model usually produces three optimal masks. The model ranks predicted masks by their Intersection over Union (IoU) with the object and evaluates the loss for each to obtain the most accurate prediction.

However, as shown in Figure 5, the successful SAM segmentation is not always granted to the maximum accuracy mask. Because the original SAM structure was designed for the segmentation of 2-D natural images. As a result, when evaluated with medical images, the model exhibits performance degradation (Gong et al., 2023). Therefore, to improve

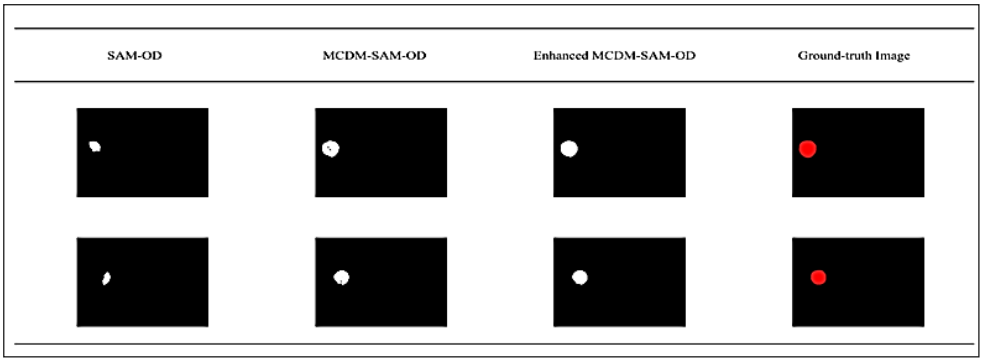


Figure 5. Results of OD segmentation: columns (1) using SAM; (2) SAM with MCDM; (3) enhanced MCDM-SAM; and (4) ground-truth image

optic disc segmentation, this study suggests using the Multiple Criteria Decision Making (MCDM) technique with the SAM (MCDM-SAM). MCDM aids in choosing the optimal solution among several choices for making informed decisions when multiple criteria are involved (Ali & Broumi, 2024; Barretta et al., 2023; Kim et al., 2022). For each possible segmentation to produce more accurate OD segmentation, the MCDM considers confidence to choose segments with a higher probability and area score to compare each segment to a threshold range. The confidence scores help evaluate the validity of each segment, and the area-based approach ensures the identified segment region is neither too large nor too small. To reflect the anatomical size range of the OD, these area scores were determined empirically through validation experiments on the training dataset. This integrated criterion generates a superior and well-conformed OD segmentation.

As seen in Figure 5(2), determining boundaries can be challenging due to the presence of blood vessels and poor contrast. Thus, this work uses ellipse fitting (Pallawala et al., 2004) along with the MCDM-SAM technique to streamline the detection process and capture the essence of the OD boundary. To find the best-fitting ellipse, the algorithm uses the least squares techniques. It determines the ellipse that best fits a collection of 2D points, such as the edges of an object. It returns a rotated rectangle that completely contains the fitted ellipse. This rectangle is useful because it shows the position and orientation of the ellipse in relation to the points and ensures that the fitting is done effectively. It helps minimise the difference between the actual points and the ellipse. Finally, the ellipse-fitting algorithm can accurately place an ellipse around the object's edges. This Enhanced MCDM-SAM ensures more representation of actual object shapes. The general steps for the enhanced MCDM-SAM segmentation algorithm are as follows:

Input: Preprocessed image (obtained using Equation 1) for segmentation, points, and label for mask prediction, and area constraints for mask selection.

Output: Optic disc segmented image.

Step-by-step Process:

Step 1: Input the V-channel image and apply the median blur to reduce noise. The median blur operation can be represented in the following equation.

$$I' = \text{medianBlur}\left(I_{v,}\ k = 15\right) \quad [2]$$

where I_v represents the v-channel input image, I′ denotes the smoothed output image, and k indicates the kernel value with size 15.

Step 2: Load the segmentation predictor model and set the input image in the predictor.

$$sam = \text{sam_model_registry}[model_type]\ \ (\text{checkpoint} = path) \quad [3]$$

$$Pred = \text{SamPredictor}(sam) \quad [4]$$

Step 3: Generate segmentation masks using points and labels of an image.

$$m, s, l = \text{Pred}(sam).\text{predict}(I_{Pt}, I_l, mm_O = True) \quad [5]$$

where, m- masked images, s-score of the masked images, l-logits values, I_{pt}- Points of an input image, I_l- Labels of an input image, mm_O -Multi-mask output mode.

Step 4: Compute the pixel area on the generated segmentation mask image with its confidence score.

$$A_i = \sum_{x,y} M_i(x, y) \quad [6]$$

Where M_i- masked images, A_i- Area of the i-th mask

Step 5:

a. Sort the masked images based on their confidence scores.

$$S = \text{argsort}(s)[::-1] \quad [7]$$

Here, S contains indices sorted by descending score.

b. Apply criteria with area constraints for selection. If the first condition is met, it selects the highest-scoring mask as the final one. If the first condition is not satisfied,

it selects the second-high score mask. Otherwise, it selects the third-smallest-area mask, or else the second-highest-score mask as the final one.

$$m_{final} = \begin{cases} m_{idx[0]}, & \text{if } A_{min} \leq A_{idx[0]} \leq A_{max[1]} \quad [8] \\ m_{idx[1]}, & \text{if } A_{idx[1]} \leq A_{max[2]} \quad [9] \\ m_{idx[2]}, & \text{otherwise (fallback)} \quad [10] \end{cases}$$

Here, A_{min} represents the minimum area threshold, $A_{max[1]}$ is the second maximum area for the high-score mask, and $A_{max[2]}$ represents the maximum area for other masks. From this, we get the final selected mask image that satisfies the conditions.

Step 6: Apply the ellipse fitting on the selected mask image. The ellipse-fitting approximates the ellipse parameters to minimise the least-squares distance between the ellipse (E) and the set of contour points (P). This fitting process can be expressed as follows:

$$E = \text{fitEllipse}(P) \quad [11]$$

As depicted in Figure 6, the following gives a comprehensive explanation of each step of the OD segmentation process:

- The original image (Figure 6a) is assigned to the image encoder to transform it into an image embedding after being transformed into the desired colour channel (Figure 6b).
- To obtain the desired region, the prompt encoder is provided with the coordinate point that was previously found (Figure 6c).
- The predicted probability of masks (Figure 6d, Figure 6e, Figure 6f) is obtained.
- Based on the multiple criteria, the enhanced optic disc segmentation result is extracted (Figure 6g, Figure 6h).
- The optic disc boundary is enhanced using the ellipse fitting (Figure 6i) on the extracted OD image.

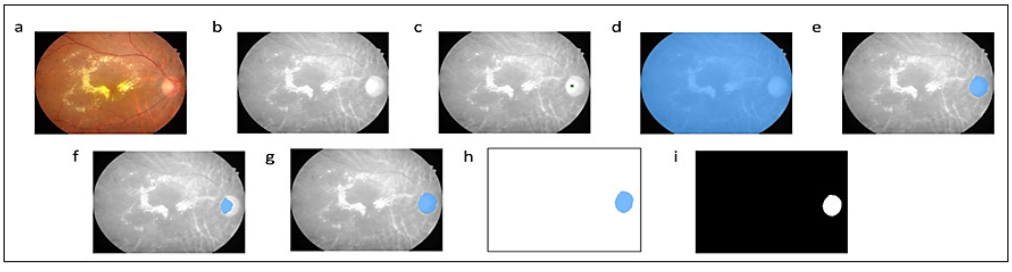


Figure 6. Results of each process of OD segmentation: (a) input image; (b) V-channel image; (c) positional information; (d-f) different mask probabilities; (g-h) final masks using MCDM-SAM; and (i) enhanced OD boundary

RESULTS AND DISCUSSION

Performance Measurement

The efficiency of the proposed OD segmentation approach is assessed using several evaluation metrics, such as precision, recall, sensitivity, F1 score, IoU (Jaccard score), and dice coefficient (DICE), relative to the ground truth. (Sarvamangala & Kulkarni, 2022). Accuracy is the number of accurate predictions among all predictions, used to assess overall correctness. Recall measures the number of actual positives among all actual positives, indicating how well the model detects real positive occurrences. Precision measures the accuracy of positive predictions by dividing the number of actual positives by the total number of predicted positives. The F1-score is perfect for unbalanced data since it balances recall and precision as their harmonic mean. This scale provides a single measurement of false negatives and positives. The Intersection over Union (IoU) is an important metric in segmentation work because it quantifies the overlap between the predicted and ground-truth areas. In the medical field, both IoU and Dice coefficient (DICE) are widely used to determine the degree of specificity of the segmentation methods (Furtado, 2020).

$$\text{Accuracy} = \frac{T_P + T_N}{T} \quad [12]$$

$$\text{Recall}(T_{PR}) = \frac{T_P}{F_N + T_P} \quad [13]$$

$$\text{Precision } (P_{Pv}) = \frac{T_P}{F_P + T_P} \quad [14]$$

$$\text{F1 score} = 2 \; X \; \frac{P_{PV} \; X \; T_{PR}}{P_{PV} + T_{PR}} \quad [15]$$

$$\text{IoU} = \frac{T_P}{T_P + F_P + F_N} \quad [16]$$

Here, T_P and T_N represent correctly classified positive and negative results, respectively. Similarly, F_P and F_N represent false-positive and false-negative events. T denotes the total number of pixel-level predictions evaluated, T_{PR} represents the true positive rate, and P_{PV} represents the positive predictive value.

Experimental Analysis using a Deep Learning Segmentation Approach

In the experiment, the effectiveness of the proposed method is compared with the current state-of-the-art (Rayed et al., 2024) deep learning-based models, such as a fully convolutional network (FCN) (Shelhamer et al., 2014), U-Net (Ronneberger et al., 2015), and DeepLab (Chen et al., 2018). The parameter settings used in this experiment are given in Table 1.

Although FCN and DeepLab achieved adequate precision and recall, their low segment quality is reflected in their F1 Scores and IoUs. The proposed enhanced MCDM-SAM method performs better than U-Net in all aspects of recall, F1-score, and IoU. This result shows that the proposed method is superior to all others in terms of overlap and segmentation accuracy. SAM and MCDM-SAM are similar in terms of precision and recall. The proposed Enhanced MCDM-SAM approach offers the highest level of accuracy and reliability for segmentation tasks, achieving the best possible recall, F1 score, and IoU. The results are presented in Table 2.

According to Table 2, the proposed approach consistently outperforms other approaches across all key measures. It achieves the best balance between precision and recall, resulting in the highest F1 score and IoU. FCN gives high accuracy, but extremely low recall due to class imbalance at the pixel level.

Table 1
Implementation setup details of baseline models

Model	Framework	Model Availability?	Pre-trained	Max iteration	Batch size
FCN	TensorFlow	Yes	ImageNet	25	1
U-Net	TensorFlow	Yes	ImageNet	25	1
DeepLab	TensorFlow	Yes	ImageNet	25	1

Table 2
Comparative analysis with the state-of-the-art deep learning-based segmentation approach

Methods	Accuracy	Precision	Recall (Sensitivity)	Specificity	F1 score/ DICE	IoU/ Jaccard
FCN	0.869	0.984	0.055	0.999	0.130	0.072
DeepLab	0.972	0.429	0.763	0.977	0.314	0.205
U-Net	0.995	0.927	0.857	0.998	0.696	0.544
SAM	0.965	0.948	0.838	0.968	0.853	0.789
MCDM-SAM	0.983	0.948	0.890	0.985	0.895	0.84
Proposed enhanced MCDM-SAM	0.981	0.942	0.916	0.982	0.907	0.862

It emphasises the importance of balanced evaluation metrics, such as DICE/ F1 and IoU. The results demonstrate that FCN is primarily accurate at predicting negatives but fails to identify most actual lesions, as evidenced by low IoU and F1 Scores, indicating subpar performance in OD pixel-level segmentation. DeepLab achieves good accuracy and recall; however, low precision results in a substantial number of false positives. A modest IoU suggests that the segmentation does not closely match the ground truth. The U-Net network delivers high accuracy, good precision, and high recall. The IoU scores indicate that the segmentation quality is acceptable, although not outstanding. The SAM consistently demonstrates high performance across all evaluation metrics. Good image segmentation quality can be inferred from high precision, a high recall, and a strong IoU. The MCDM-SAM enhances SAM by increasing its recall rate while sustaining a high level of precision. An improved IoU indicates high-quality segmentation results. But further enhancement in MCDM-SAM facilitates the best recall for identifying the most lesions. It maintains high precision, so few false positives. Moreover, the best F1 score and IoU indicate the most accurate segmentation overall. The IDRiD dataset images utilised in experimental analysis display a range of clinical conditions, such as blood vessel anomalies, comparable intensity to exudates, and other anomalies. The overlap results for various pathological images obtained using the proposed enhanced MCDM-SAM approach and alternative methods are illustrated in Figure 7. In terms of OD segmentation, the Enhanced MCDM-SAM approach outperforms the other models overall.

Experimental Analysis using Image Processing Approach

To compare the OD segmentation results with traditional image processing, k-means clustering with a template matching algorithm from another paper was reproduced. Table 3 shows the segmentation results assessed using the IoU.

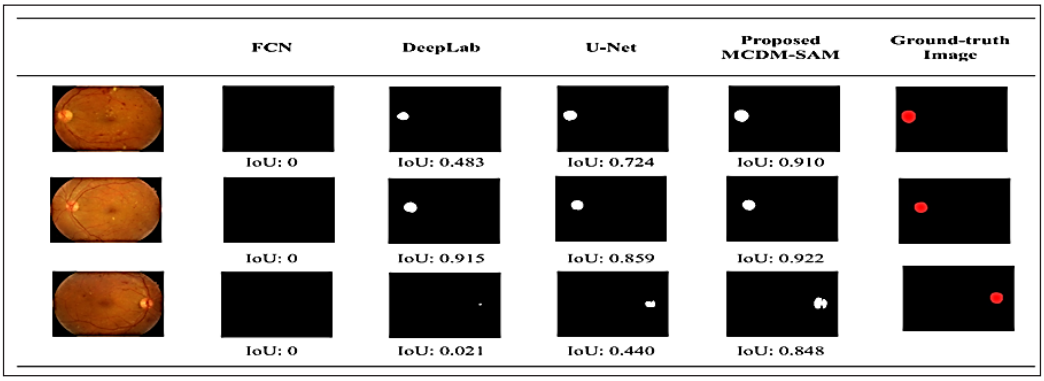


Figure 7. Comparison of overlaps in a few example cases

The table findings indicate that the K-means clustering with template matching method performs moderately across all metrics, with acceptable Recall but lower Precision and overlap metrics (IoU, Dice coefficient). This is because, as Figure 8 illustrates, template matching relies on exact matches to identify the optic disc; it has difficulty in handling variations such as varying lighting conditions or partial occlusions. Conversely, the proposed method significantly outperforms the K-means approach across all metrics. The Enhanced MCDM-SAM is more precise and efficient at segmenting regions of interest, as evidenced by its significantly higher Precision, Recall, F1 Score, IoU, and Dice coefficient. The study's results show that the proposed method performs better in OD segmentation.

To eliminate inter-study variability and assess the efficacy of the proposed OD method, controlled comparisons were conducted using FCN, U-Net, DeepLab, and template matching. This approach has been represented from the early days of FCN through popular U-Net, which is currently well-represented in clinical segmentation, to advanced semantic segmentation (DeepLab), and to foundational models (SAM).

Table 3
Evaluation result with traditional image processing methods

Metrics	Traditional Method (Basu et al., 2021)	Enhanced MCDM-SAM
Precision	0.49	0.94
Recall	0.84	0.92
F1 Score	0.62	0.91
IoU	0.47	0.86
Dice coefficient	0.62	0.91

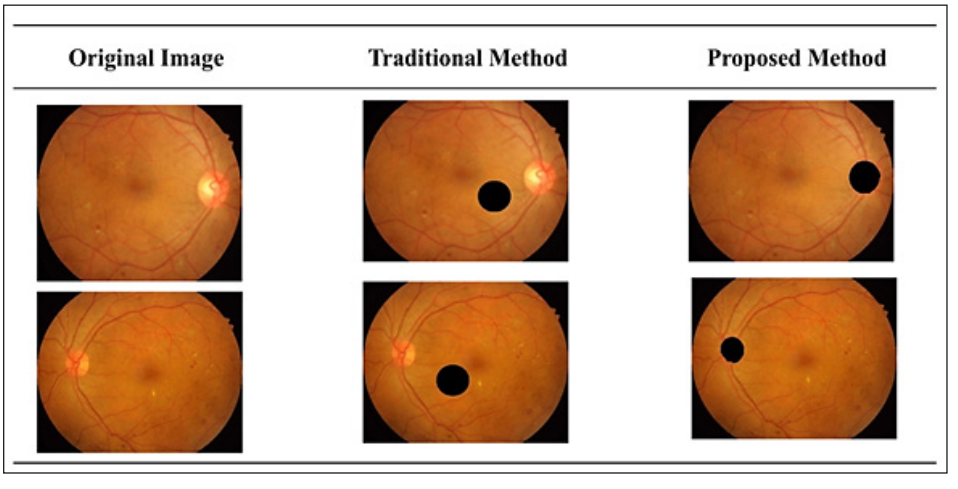


Figure 8. Comparison result of the segmentation

To ensure a fair comparison, the optimised hyperparameters for each model were obtained using the same validation set. Performance was evaluated using precision, recall, F1/DICE score, and IoU on a test dataset consisting of 27 fundus images with expert-verified OD annotations.

Analysis of Detection Results: With and Without OD Segmentation

The impact of OD segmentation on exudate detection was evaluated by utilising retinal images with and without OD segmentation. For this experiment, 54 high-resolution fundus images for training and 27 for validation, each with annotated hard exudate lesions, are used. In the preprocessing stage, the image was converted into the LAB colour space, and the initial enhancement of each image was performed using Contrast Limited Adaptive Histogram Equalisation applied to the L-channel within the LAB colour space. To enhance the model's robustness, the albumentations library was employed for data augmentation, including random scaling, flipping, rotation, and adjustments to brightness and contrast. The detection model was trained for 100 epochs using a YOLO-based framework, with an input image size of 1024×1024 pixels, a batch size of 8, and no model-specific hyperparameter tuning.

The results shown in Table 4 demonstrated improvements in hard exudate detection, specifically in precision from 0.468 to 0.476, recall from 0.314 to 0.344, and mAP50 from 0.310 to 0.332. Excluding the optic disc area appears to reduce the likelihood of false positives, mainly because its appearance and intensity are comparable to those of hard exudates. The overall fitness score increased from 0.157 to 0.165, which further supports the benefits of OD segmentation in DR screening at the lesion level.

From the above-mentioned experiments, the proposed framework was more successful at optic disc segmentation than the traditional segmentation model. In fully supervised deep learning models such as FCN, DeepLab, and U-Net, overfitting often occurs when the dataset has few training images. However, the proposed work utilises a pre-trained zero-shot SAM without task-specific retraining, and MCDM-based selection and ellipse fitting are rule-based post-processing steps. Hence, the risk of overfitting is reduced compared to fully supervised deep learning models. However, this study also has some limitations.

Table 4
Hard exudate test comparison results

Metric	Before OD Segmentation	After OD Segmentation
Precision	0.47	0.48
Recall	0.31	0.34
mAP@0.5	0.31	0.33
Images	130	130

Multi-centre dataset validation is necessary to confirm the model's generalisability, as the proposed work was conducted on a single dataset. Additionally, cross-dataset validation is required to evaluate the OD localisation approach, despite the V-channel being less sensitive to moderate illumination variations.

CONCLUSION

This study proposes an enhanced MCDM-SAM method to segment OD in fundus images while simultaneously addressing the problem of a lack of annotated training data. It uses a predetermined criterion with a wide range of possibilities to accurately determine the optic disc. Elliptical fitting has been used to smooth the OD boundary. The study's results show that the proposed method performs better than other approaches. The proposed OD segmentation approach will help prevent false positives and increase the accuracy of true hard exudate lesion detection. In this study, optic disc segmentation was performed using the proposed framework, and its impact on the detection accuracy of hard exudate lesions was investigated. No other lesions were examined. Therefore, investigation of other lesions would be helpful in grading interpretation. Future work will involve designing an experiment across all lesion types, testing it on larger, more diverse datasets, and evaluating the impact on grading accuracy for diabetic retinopathy. This work is significant for improving optic disc detection and segmentation, thereby supporting more reliable exudate detection.

ACKNOWLEDGEMENT

This work was supported by the Centre for Machine Learning and Intelligence (CMLI), an ISO-certified centre (ISO/IEC 20000-1:2018), funded by the Department of Science and Technology (DST-CURIE).

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF COMPETING INTEREST

The authors report there are no competing interests to declare.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used QuillBot and Grammarly to improve language and readability. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the publication.

DATA AVAILABILITY

The data supporting the findings of this study are openly available in the Indian Diabetic Retinopathy Image Dataset (IDRiD) at https://dx.doi.org/10.21227/H25W98.

CODE AVAILABILITY

The code supporting the findings of this study is available from the corresponding author upon reasonable request.

REFERENCES

Abdullah, A. S., Rahebi, J., Özok, Y. E., & Aljanabi, M. (2020). A new and effective method for human retina optic disc segmentation with fuzzy clustering method based on active contour model. *Medical and Biological Engineering and Computing, 58*(1), 25-37. https://doi.org/10.1007/s11517-019-02032-8

Ali, A. M., & Broumi, S. (2024). Machine learning with a multi-criteria decision-making model for thyroid disease prediction and analysis. *Multicriteria Algorithms with Applications, 2*, 80-88. https://doi.org/10.61356/j.mawa.2024.26961

Antonetti, D. A., Silva, P. S., & Stitt, A. W. (2021). Current understanding of the molecular and cellular pathology of diabetic retinopathy. *Nature Reviews Endocrinology, 17*(4), 195-206. https://doi.org/10.1038/s41574-020-00451-4

Barretta, R., Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia, 3*(1), 77-87. https://doi.org/10.3390/encyclopedia3010006

Basu, S., Mukherjee, S., Bhattacharya, A., & Sen, A. (2021). Segmentation of blood vessels, optic disc localisation, detection of exudates, and diabetic retinopathy diagnosis from digital fundus images. In *Proceedings of the International Conference on Intelligent Computing and Communication (ICICC 2020)* (pp. 173-184). Springer. https://doi.org/10.1007/978-981-16-1543-6_16

Bengani, S., Angel Arul Jothi, J., & Vadivel, S. (2020). Automatic segmentation of optic disc in retinal fundus images using semi-supervised deep learning. *Multimedia Tools and Applications, 80*(3), 3443-3468. https://doi.org/10.1007/s11042-020-09778-6

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, Y., Bai, Y., & Zhang, Y. (2024). Optic disc and cup segmentation for glaucoma detection using attention U-Net incorporating residual mechanism. *PeerJ Computer Science, 10*, Article e1941. https://doi.org/10.7717/peerj-cs.1941

Chincholi, F., & Koestler, H. (2023). Detectron2 for lesion detection in diabetic retinopathy. *Algorithms, 16*(3), Article 147. https://doi.org/10.3390/a16030147

Cleland, C. (2023). Comparing the International Clinical Diabetic Retinopathy (ICDR) severity scale. *Community Eye Health, 36*(119), 10. https://pmc.ncbi.nlm.nih.gov/articles/PMC10436766/

Dharmawan, D. A., Ng, B. P., & Rahardja, S. (2020). A new optic disc segmentation method using a modified Dolph-Chebyshev matched filter. *Biomedical Signal Processing and Control, 59*, Article 101932. https://doi.org/10.1016/j.bspc.2020.101932

Dietter, J., Haq, W., Ivanov, I. V., Norrenberg, L. A., Völker, M., Dynowski, M., Röck, D., Ziemssen, F., Leitritz, M. A., & Ueffing, M. (2019). Optic disc detection in the presence of strong technical artefacts. *Biomedical Signal Processing and Control, 53*, Article 101535. https://doi.org/10.1016/j.bspc.2019.04.012

Fu, Y., Zhang, G., Lu, X., Wu, H., & Zhang, D. (2023). RMCA U-Net: Hard exudates segmentation for retinal fundus images. *Expert Systems with Applications, 234*, Article 120987. https://doi.org/10.1016/j.eswa.2023.120987

Furtado, P. (2020). Segmentation of diabetic retinopathy lesions by deep learning: Achievements and limitations. In *Proceedings of the 7th International Conference on Bioimaging (BIOIMAGING 2020)* (pp. 95-101). https://doi.org/10.5220/0008881100002513

Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.-A., & Dou, Q. (2023). 3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable medical image segmentation. *arXiv*. https://doi.org/10.1016/j.media.2024.103324

Guan, G., Zhao, C., Li, Z., Zhang, P., Chen, Y., Ye, P., Wong, M. K., Chan, L. Y., Yan, H., Tang, C., & Zhao, Z. (2025). EmbSAM: Cell boundary localisation and Segment Anything Model for fast images of developing embryos. *Communications Biology, 9*(1), Article 8. https://doi.org/10.1038/s42003-025-09220-3

Halder, N., Roy, D., Banerjee, R., Roy, P., Sarkar, P. P., & Bandyopadhyay, S. (2020). Automatic detection and segmentation of optic disc (ADSO) of retinal fundus images based on mathematical morphology. In *Proceedings of the 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA 2020)*. https://doi.org/10.1109/NCETSTEA48365.2020.9119931

Hasan, M. K., Alam, M. A., Elahi, M. T. E., Roy, S., & Martí, R. (2021). DRNet: Segmentation and localisation of optic disc and fovea from diabetic retinopathy image. *Artificial Intelligence in Medicine, 111*, Article 102001. https://doi.org/10.1016/j.artmed.2020.102001

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16000-16009). https://doi.org/10.1109/CVPR52688.2022.01553

Joshi, S., & Karule, P. T. (2018). A review on exudates detection methods for diabetic retinopathy. *Biomedicine and Pharmacotherapy, 97*, 1454-1460. https://doi.org/10.1016/j.biopha.2017.11.009

Kamble, R., Kokare, M., Deshmukh, G., Hussin, F. A., & Mériaudeau, F. (2017). Localisation of optic disc and fovea in retinal images using intensity-based line scanning analysis. *Computers in Biology and Medicine, 87*, 382-396. https://doi.org/10.1016/j.compbiomed.2017.04.016

Kim, B. S., Shah, B., He, T., & Kim, K. I. (2022). A survey on analytical models for dynamic resource management in wireless body area networks. *Ad Hoc Networks, 135*, Article 102936. https://doi.org/10.1016/j.adhoc.2022.102936

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3992-4003). https://doi.org/10.1109/ICCV51070.2023.00371

Kumar, S., Adarsh, A., Kumar, B., & Singh, A. K. (2020). An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation. *Optics & Laser Technology, 121*, Article 105815. https://doi.org/10.1016/j.optlastec.2019.105815

Liu, S., Jiang, X., Zhang, J., & Zou, W. (2025). Hard exudates segmentation for retinal fundus images based on longitudinal multi-scale fusion network. *Medical and Biological Engineering and Computing, 63*(12), 3761-3775. https://doi.org/10.1007/s11517-025-03426-7

Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis, 89*, Article 102918. https://doi.org/10.1016/j.media.2023.102918

Misera, L., Nebelung, S., Carrero, Z. I., Bressem, K., Ligero, M., Kühn, J.-P., Hoffmann, R.-T., Truhn, D., & Kather, J. N. (2026). Reducing manual workload in CT and MRI annotation with the Segment Anything Model 2. *BMC Medical Imaging, 26*(1), Article 75. https://doi.org/10.1186/s12880-025-02075-4

Motevali, S., Khanal, A., & Estrada, R. (2021). Optic disc segmentation using disk-centred patch augmentation. *arXiv*. https://doi.org/10.48550/arXiv.2110.00512

Nagpal, D., Pattanaik, P. A., Madaan, V., & Agrawal, P. (2024). Enhancement of retinal images through modified anisotropic diffusion. *Journal of Medical Artificial Intelligence, 7*, Article 80. https://doi.org/10.21037/jmai-23-80

Pallawala, P. M. D. S., Hsu, W., Lee, M. L., & Eong, K. G. A. (2004). Automated optic disc localisation and contour detection using ellipse fitting and wavelet transform. In *Lecture Notes in Computer Science* (Vol. 3022, pp. 139-151). https://doi.org/10.1007/978-3-540-24671-8_11

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD). *IEEE Dataport*. https://dx.doi.org/10.21227/H25W98

Ramani, R. G., & Shanthamalar, J. J. (2020). Improved image processing techniques for optic disc segmentation in retinal fundus images. *Biomedical Signal Processing and Control, 58*, Article 101832. https://doi.org/10.1016/j.bspc.2019.101832

Rayed, M. E., Islam, S. M. S., Niha, S. I., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2024). Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked, 47*, Article 101504. https://doi.org/10.1016/j.imu.2024.101504

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (Vol. 9351, pp. 234-241). https://doi.org/10.1007/978-3-319-24574-4_28

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence, 15*(1), 1-22. https://doi.org/10.1007/s12065-020-00540-3

Septiarini, A., Hamdani, H., Setyaningsih, E., Junirianto, E., & Utaminingrum, F. (2023). Automatic method for optic disc segmentation using deep learning on retinal fundus images. *Healthcare Informatics Research, 29*(2), 145-152. https://doi.org/10.4258/hir.2023.29.2.145

Sevastopolsky, A. (2017). Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis, 27*(3), 618-624. https://doi.org/10.1134/S1054661817030269

Shelhamer, E., Long, J., & Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640-651. https://doi.org/10.1109/TPAMI.2016.2572683

Stojmenovic, M., & Nayak, A. (2007). Direct ellipse fitting and measuring based on shape boundaries. In *Lecture Notes in Computer Science* (Vol. 4872, pp. 221-235). https://doi.org/10.1007/978-3-540-77129-6_22

Teoh, C. S., Wong, K. H., Xiao, D., Wong, H. C., Zhao, P., Chan, H. W., Yuen, Y. S., Naing, T., Yogesan, K., & Koh, V. T. C. (2023). Variability in grading diabetic retinopathy using retinal photography and its comparison with an automated deep learning diabetic retinopathy screening software. *Healthcare, 11*(12), Article 1697. https://doi.org/10.3390/healthcare11121697

Theera-Umpon, N., Poonkasem, I., Auephanwiriyakul, S., & Patikulsila, D. (2019). Hard exudate detection in retinal fundus images using supervised learning. *Neural Computing and Applications, 32*, 179-192. https://doi.org/10.1007/s00521-019-04402-7

Usman Akram, M., Khalid, S., Tariq, A., Khan, S. A., & Azam, F. (2014). Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine, 45*(1), 161-171. https://doi.org/10.1016/j.compbiomed.2013.11.014

Valverde, C., Garcia, M., Hornero, R., & Lopez-Galvez, M. (2016). Automated detection of diabetic retinopathy in retinal images. *Indian Journal of Ophthalmology, 64*(1), 26-32. https://doi.org/10.4103/0301-4738.178140

Xue, X., Wang, L., Du, W., Fujiwara, Y., & Peng, Y. (2022). Multiple preprocessing hybrid level set model for optic disc segmentation in fundus images. *Sensors, 22*(18), Article 6899. https://doi.org/10.3390/s22186899

Yin, P., Xu, Y., Zhu, J., Liu, J., Yi, C., Huang, H., & Wu, Q. (2021). Deep level set learning for optic disc and cup segmentation. *Neurocomputing, 464*, 330-341. https://doi.org/10.1016/j.neucom.2021.08.102

Zhang, Y., Cai, X., Zhang, Y., Kang, H., Ji, X., & Yuan, X. (2021). TAU: Transferable attention U-Net for optic disc and cup segmentation. *Knowledge-Based Systems, 213*, Article 106668. https://doi.org/10.1016/j.knosys.2020.106668

Pertanika J. Sci. & Technol. 34 (2): 913 - 950 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Review Article

A Critical Review of User-centric QoS Provisioning Frameworks in 5G Networks

Wai Leong Pang[1], Wen Hao Anselm Chow[1], Hui Hwang Goh[1], Swee King Phang[1], and Kah Yoong Chan[2,3]*

[1]*School of Engineering, Faculty of Innovation and Technology, Centre for Sustainable Societies, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia*
[2]*Faculty of Artificial Intelligence and Engineering, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia*
[3]*Centre for Advanced Devices and Systems, COE for Robotics and Sensing Technologies, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia*

ABSTRACT

5G networks provide the Quality of Service (QoS) to various services, including Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC) with different requirements. However, the existing 5G network deploys the static 5G QoS Identifiers (5QI) framework, which provides a basic level of services differentiation and supports the class-based prioritisation. The static 5QI fails to support the dynamic and user-centric QoS requirements. This paper reviews user-centric QoS provisioning schemes that focus on the Software-Defined Networks (SDN), Machine Learning (ML), and network slicing schemes. A structured evaluation methodology is used to analyse the existing works in terms of architecture, mechanisms, QoS metrics, scalability, and real-world feasibility. The reviews discover that SDN improves programmability and policy-based control, ML predicts the traffic adaptively, and network slicing provides service isolation. However, user-centric schemes can increase computational load, raising controller overhead and latency. This paper identifies the research gaps in the user-centric schemes related to scheme readiness, scalability, security, and heterogeneity. The paper also proposed a direction for deploying user-centric QoS schemes in real 5G networks that support future 5G enhancement that covers the adaptive, efficient, and personalised QoS provisioning.

Keywords: 5G, quality of service, software-defined networking, user-centric

ARTICLE INFO

Article history:
Received: 26 February 2025
Accepted: 25 March 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.14

E-mail addresses:
Waileong.pang@taylors.edu.my (Wai Leong Pang)
0361738@sd.taylors.edu.my (Wen Hao Anselm Chow)
jonathan.goh@taylors.edu.my (Hui Hwang Goh)
sweeking.phang@taylors.edu.my (Swee King Phang)
kychan@mmu.edu.my (Kah Yoong Chan)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The environments, network speed, reliability, and service consistency play a crucial role in ensuring user satisfaction across different applications in existing mobile networks. The latest 5G wireless networks improve network speed and capacity compared with previous generations. Consequently, 5G can support various user-centric and mission-critical applications, including Virtual Reality (VR) applications, autonomous vehicles, smart manufacturing, and real-time applications. However, there is still a challenging task to ensure a stable QoS provisioning for various types of services.

The previous cellular networks, with static configurations and specialised hardware, fail to support the current high complexity and scalability of network conditions (Haji et al., 2021). The current mobile networks fail to provide QoS to all users, especially premium customers, when the networks are overloaded. Conventionally, the 5G network relies on fixed QoS levels for traffic management, which are pre-defined and cannot be changed quickly in the event of varying network conditions. The current design of QoS allows for up to nine levels of priority for varying telco service provided, designed to prioritise traffic across different service types, including voice, video, and best-effort data transfers. Although this approach can work well under normal conditions, it may break during congestion events, resulting in higher latency, packet loss, and general degradation of user experience. Therefore, the key technical challenge is not just to define QoS classes, but also to ensure QoS enforcement stability and efficiency under dynamic load conditions.

As summarised in Table 1, the existing QoS framework for 5G networks allocates bandwidth resources according to the predefined priorities of the services. Although the QoS framework based on priorities can provide a certain level of service differentiation, it is not fully capable of satisfying the new requirement for personalised QoS provisioning, especially for the premium segment that requires high availability, reliability, and integrity of the services provided. Regarding the QoS in the 5G network architecture, the framework specifically supports a flexible configuration of 64 QoS Identifiers known as 5QI (5G QoS Identifier) for differentiating the services according to utilisation demands (Debnath et al., 2023). The 5G QoS architecture essentially supports a completely different configuration than the existing 4G architecture. It facilitates a total of nine Quality Class Identifiers (QCIs) of resources for the services, resembling the nine levels of priorities. The levels of priorities in the 5G architecture are essentially determined by the 3rd Generation Partnership Project (3GPP) for standardising the overall 5G communication framework (Hussain et al., 2020; Shrivastava et al., 2022). As shown in Table 1, delay-sensitive services such as conversational voice and real-time gaming are assigned higher priority and stricter delay constraints to preserve interactivity, while less delay-sensitive services such as non-conversational video are allocated comparatively relaxed QoS budgets. This structure design allows for differentiation in service level establishment. However, the issue

Table 1
Standardised 5QI mappings

5QI	Service type	Resource type	Priority	Packet Delay Budget	Packet Error Rate
1	Conversational Voice (VoIP)	Guaranteed Bit Rate (GBR)	Highest	100ms	10^{-2}
2	Conversational video	GBR	High	150ms	10^{-3}
3	Real-time gaming	GBR	High	50ms	10^{-3}
4	Non-conversational video streaming	GBR	Medium	300ms	10^{-6}
5	IMS signalling	Non-GBR	Medium	100ms	10^{-6}
6	Voice/video signalling	Non-GBR	Low	100ms	10^{-6}
7	Live streaming	Non-GBR	Medium	100ms	10^{-6}
8	Low-latency data (Interactive gaming/ Buffered streaming)	GBR	High	10ms	10^{-6}
9-63	Mission-critical applications	Non-GBR/GBR	Variable	Variable	Variable

Note. Adapted from Al-Shammari et al. (2018); Hoyhtya et al. (2018); Hussain et al. (2020); and Trifan et al. (2015)

now is how to ensure optimum bandwidth utilisation with optimum QoS performance in a real-world scenario, especially in a network with unbalanced network utilisation (Khan, 2022). It is important to note that the static QoS mapping may not consider the dynamic user behaviour or traffic patterns. Especially in cases when premium users who are guaranteed the highest priority and bandwidth do not utilise their granted resources. The network bandwidth will be underutilised, while the performance of non-premium users will be low due to bandwidth constraints. This highlights the need to shift QoS provisioning from strictly service-class differentiation toward more adaptive and user-centric resource allocation strategies.

The 5G QoS Identifier (5QI) is used to provide service differentiation in the existing 5G framework. The priority and resource allocation are assigned according to the service class (eMBB, URLLC, mMTC). All the users receive the same QoS for the same service class, and this is clearly a service-centric framework. However, the QoS for the user-centric framework is delivered to the individual users according to the users' priority, subscription tier, real-time user experience, etc. The user-centric QoS can be deployed on the programmable SDN with the ML-powered prediction model to adaptively manage the limited resources.

In this review paper, various algorithm-based and Machine Learning (ML)-assisted approaches have been explored to analyse network traffic and improve bandwidth allocation under both overloaded and non-overloaded scenarios. These approaches are generally integrated with SDN, which allows for programmatic control to dynamically control traffic. For instance, the closed-cycle process for traffic management, where traffic can be monitored in real time to adapt policies to traffic conditions, has already been validated in research using machine learning to analyse traffic patterns to determine overload conditions in traffic (Hyder & Lung, 2018). As opposed to the previous reliance, the method aims to facilitate the adaptation in the allocation of prioritising the bandwidth based on the time-varying demands.

Several works further indicate that dynamic allocation schemes have been reported to enhance resource utilisation by redistributing unused high-priority resources when demand is lower and still granting service guarantees at increased demands. For example, Nikolaidis et al. (2023) discuss per-user QoS guarantees in network slicing scenarios, illustrating how more flexible provisioning of resources across users can take place. The literature indicates that the combination of SDN and ML enhances QoS flexibility and bandwidth efficiency but also introduces practical constraints, which include control-plane overhead, inference latency, and scalability challenges in dense 5G environments (Amin et al., 2021). Conveying the rapidly changing demands of various modern network services, 5G networks will have to be highly flexible, scalable, and dynamic. As compared to traditional networks, which are bound to their configuration and are only supported through hardware-based proprietary solutions such as routers and firewalls. SDN is a relatively new paradigm that has come to offer a novel approach in which the control plane is decoupled from the data plane (Etxezarreta et al., 2023). The integration of SDN reduces the Capital Expenditure (CAPEX) and Operational Expenditure (OPEX) in 5G networks. There are also opportunities for managing dynamic resources and services in response to real-time conditions.

The evolution from Long Term Evolution (LTE) to 5G improved the service quality in mobile communications. Resource allocation is a promising solution to support the surge in mobile data traffic under limited licensed spectrum. The coexistence of LTE and Wi-Fi has a significant impact on performance in terms of throughput degradation and energy efficiency loss under various modulation and coding schemes (Yaqoob et al., 2020). A robust mobile network system should protect the QoS of the real-time streaming applications (Yaqoob et al., 2019).

Beyond technical performance, the QoS stability problem has a growing global impact on how 5G services are perceived and monetised. As increased deployments of 5G technology within urban cities and industrial sites are underway, the deterioration of service within such scenarios of increased congestion stands to have an immediate impact on the end consumer with respect to the services they can access within the 5G environment,

including more robust applications such as smart manufacturing, the Internet of Vehicles, and smart cities. As such, the issue of scalable QoS within the 5G environment continues to become a determining aspect in the successful rollout of 5G technology across the globe.

In addition to QoS classification challenges, practical QoS performance is also influenced by coexistence effects and resource allocation efficiency in multi-access and multi-user environments. Furthermore, the involvement of trade-offs in resource allocation problems of Non-Orthogonal Multiple Access (NOMA) systems for 5G networks is focused on where researchers investigate how a balance between throughput maximisation and fairness can result in optimally allocating user data rates by using sophisticated algorithms like Integer Linear Programming (ILP) and Particle Swarm Optimisation (PSO) (Abuajwa et al., 2022). A fair bandwidth distribution to the users through the resource allocation techniques in poor channel conditions, regardless of the types of services. The enhanced exponential rule-based scheduling algorithms are proposed to enhance the performance of the real-time traffic for the LTE systems (Yaqoob et al., 2020). The latency and QoS guarantees of the real-time data services are improved to ensure a stable performance for these services. The effective QoS provisioning for real-time services depends on QoS classification, practical scheduling and resource allocation strategies under real network constraints.

5G and beyond systems improve the efficiency of QoS management and resource allocation amid multi-access coexistence. Bandwidth-optimisation techniques can improve reliability but introduce trade-offs. In this review, SDN, ML and adaptive QoS-driven frameworks are examined with a focus on their strengths, limitations, and outstanding research challenges. Prior studies report theoretical QoS improvements that depend on deployment feasibility and operational constraints.

The limitations observed in 5G QoS deployment are largely associated with maintaining reliable service guarantees under fluctuating traffic conditions, rather than with the absence of formal QoS definitions. Static service class prioritisation offers an initial structure, but real-world performance varies according to user behaviour, congestion, mobility, and competing resource demands among services. Hence, the adaptive QoS provisioning needs to be considered in the context of the implementation challenges, such as controller scalability, signalling overheads, and decision latency, apart from the reported improvements.

The large-scale commercial 5G network deployed the static 5QI-based management system, but the existing works have shown that the performance of the dynamic QoS provisioning schemes is better than the static 5QI-based approaches. The experimental test shows that the dynamic scheme provides a lower packet loss rate and a more stable bitrate for the video streaming in an outdoor environment compared to the static 5QI scheme (Ito et al., 2025). Another work that incorporated adaptive network management

has shown a significant improvement in throughput and packet drop rate compared to the static approach (Yin et al., 2020). The dynamic SDN resource allocation scheme also demonstrated a significant improvement in bandwidth utilisation compared to the static QoS approaches (Yang & Tsai, 2024). These show that the dynamic approaches have significant advantages over the static 5QI-based schemes in the 5G network.

Addressing QoS Challenges in 5G

The motivation for this review results from the growing challenge of providing consistent QoS service delivery within modern 5G networks. As 5G infrastructure is used to deliver mission-critical services like autonomous transportation systems and immersive virtual environments, the tolerance for service delivery instability is much lower. Yet, providing consistent QoS service delivery remains an ongoing challenge within operational 5G networks, especially against dynamic fluctuations, congestion due to mobility, and burst-based demand patterns.

One of the major drawbacks of the conventional QoS approaches is that they are based on a set of static rules of prioritisation that are not effective enough to deal with real-time dynamic changes in traffic and user behaviours. In fact, during congestion, inflexibility in priority rules can cause less-than-optimal use of available system resources and service performance, especially when there are a variety of applications competing for limited radio and core network resources. The issue is therefore not just about service classification anymore. It is also about service allocation strategies that are not only adaptive and context-aware but also provide a certain level of fairness and reliability between users. To address these limitations, prior studies have investigated adaptive and smart QoS management approaches that integrate SDN and ML. SDN has depicted its merits in terms of making flexible programmable traffic management possible for enforcing dynamic policies according to the real-time evolving states of networks. In parallel, ML techniques have been applied to enhance traffic awareness, enabling congestion prediction, traffic classification, and decision support for prioritisation strategies. Although these existing solutions demonstrated strengthened adaptive capabilities over traditional static management techniques, their own constraints on the control plane delay over SDN, the processing latency during ML-assisted decision-making processes, as well as scalability in high traffic density, are depicted.

Consequently, this review critically examines SDN, ML, and dynamic QoS-based frameworks for user-centric QoS provisioning in support of 5G networks through the critical assessment of mechanisms, strength and gaps associated with various approaches. The objective is to determine the suitability of different frameworks under service requirements and network scenarios, along with the key trade-offs and open issues to be addressed for effective QoS assurance in 5G and beyond wireless communication systems.

Existing works that simulated ML-powered SDN and network slicing in 5G have reported the effectiveness of the proposed solution. These works provide service-centric QoS provisioning to the services without considering the users' privileges and priorities. There are limited studies on the impact of the user-level behaviour on the 5G QoS provisioning. A comprehensive review is carried out in this paper to evaluate the potential of the SDN, ML and slicing-based framework from a user-centric perspective. The review reported the ability of these solutions and the challenges in supporting the user-centric driven framework for the 5G network.

Paper Outline

This review paper adopts a structured and comparative approach to analyse existing research on user-centric QoS provisioning in 5G networks. Generally, the review is focused on SDN-based frameworks, ML-assisted QoS intelligence, and dynamic QoS and resource management mechanisms, as these represent the most common and promising directions for enhancing QoS adaptability under real-world traffic variability.

The paper is organised as follows. Firstly, the background and motivation for user-centric QoS provisioning in 5G are introduced to highlight the limitations associated with the traditional QoS support under congested and heterogeneous environments. Next, the review examines the SDN-enabled QoS control techniques, emphasising how programmability and centralised orchestration can support flexible bandwidth management. Then, this is followed by an observation on solutions based on ML, and their significance associated with enhanced traffic knowledge and prediction-based QoS management is explained. Exploring dynamic QoS solutions is also one of the steps taken in this review.

Lastly, a comparative synthesis is presented to summarise significant technical trade-offs among all the studied solutions in terms of performance enhancement, scaling limits, control-plane overhead, inference delay, and deployment simplicity. The rest of this paper is concluded with an overview of existing research gaps and a discussion of future studies related to deployable and user-oriented QoS in 5G and future communication systems.

A structured screening process was applied to ensure the review is systematic instead of descriptive. The review covered the works reported since 2017 in the areas of SDN, ML-powered, network-slicing, adaptive scheduling and user-centric QoS provisioning solutions. The QoS-related solution that is deployable on 5G or beyond is reported for each study. The mode of evaluation, either through simulation or experiment, is provided. The QoS metrics and the significant performance likes delay, throughput, fairness, and packet loss are analysed. The properties of the 5G network, like model architecture (SDN, ML, network slicing), scalability, adaptability, feasibility and user-centric QoS are evaluated for all the reviewed papers.

Trends and Evolutions of User-Centric QoS Research

The development of user-centric QoS provisioning in wireless networks has witnessed colossal evolution from 2G to 5G and is now expanding into the realm of beyond 5G (B5G) technologies. The early mobile networks, such as Global System for Mobile Communications (GSM) and Wideband Code Division Multiple Access (WCDMA), were voice-oriented with less support for data. In these networks, QoS was static and primarily preconfigured, with static bandwidth reservations and minimal differentiation by user or application. The introduction of LTE and LTE-Advanced brought about QoS provisioning as a concept. As explained in (Toor et al., 2019) Improvements to the Random-Access Procedure (RAP) in LTE allowed devices to transmit connection requests using contention-based and contention-free procedures. Access Class Barring (ACB) and Extended Access Barring (EAB) were added to control network congestion for Human-Type Communications (HTC) and Machine-Type Communications (MTC). There is better control of device access and a reduction in delay.

In 5G networks, the transition is more tense; technologies such as the grant-free transmission, NOMA, and Pattern Division Multiple Access (PDMA) were employed to facilitate massive connectivity and ultra-low latency, as demonstrated in simulation evaluations described in (Toor et al., 2019). The protocols allow devices to transmit data without the need for a grant from the base station (gNodeB), which greatly reduces the overhead while making the protocols more effective for use in environments of high density. As explained in (Sufyan et al., 2023), 5G leverages an extremely programmable and flexible system design that supports context-dependent, user-preference, and type-of-service-sensitive dynamic QoS adaptation. Edge computing, massive Multiple-Input Multiple-Output (MIMO), and network slicing complement support 5G to ensure real-time, user-level quality guarantee. However, with the sudden increase in demands from Internet of Things (IoT) ecosystems and immersive applications like the Metaverse, VR and Augmented Reality (AR), even 5G is insufficient. As a result, the research strategy is progressively shifting towards B5G technology with ultra-dense networks, an AI-driven strategy, and edge-cloud coordination, having the capability to further improve the user-centric QoS as described in (Sufyan et al., 2023). Besides that, Mahmood et al. (2021) indicate that mission-critical services based on industrial applications, such as Factory 5G, necessitate deterministic QoS, with a tendency towards vertical-specific QoS design for manufacturing and automation sectors. Moreover, the evolution of mobile communication networks and how 5G technology paves the way for B5G use cases with even more stringent latency, reliability, and flexibility requirements. In general, the evolution of user-centric QoS is a shift from fixed best-effort delivery to intelligent, adaptive, and service-aware provisioning. The current trend is centred on the use of AI/ML in prediction and real-time decision-making, adaptive access methods, and architectural shifts that extend the decision

logic to the network edge. The goal is to provide reliability, low latency, and optimal user experience in highly heterogeneous environments.

Despite the reviewed evolution, it indicates that the QoS-related problems have changed rather than disappeared. Although techniques such as grant-free access and advanced multiple access techniques reduce the overhead of scheduling and connectivity, new problems arise in contention resolution, fairness, and QoS guarantee. Thus, a major research gap that remains to be addressed is the design of user-centric QoS frameworks that can meet strict latency and reliability requirements while being scalable and computationally tractable.

Software-defined Network

Literature on the QoS provisioning for 5G networks highlights the need for SDN, Network Function Virtualisation (NFV), and ML to facilitate the adaptation of resource management to satisfy the heterogeneous QoS demands of the applications. The 5G network is equipped with various applications such as URLLC, eMBB, and mMTC, which have different QoS requirements, such as latency, throughput, and reliability (Khan, 2022; Kunasegran et al., 2025a). Thus, the QoS paradigms may not work, and there is a need to implement a flexible control plane architecture to satisfy the applications' demands.

SDN revolutionises conventional networking by introducing a separation between the control and data planes of communication to allow for centralised programmatic controls for managing network traffic according to predefined policies. As shown in Figure 1, network administrators can leverage intelligent network management techniques to dynamically set up QoS policies, rather than depending upon conventional static configurations of networks (Kunasegran et al., 2025b). In a typical 5G network setup, due to dynamic changes in user traffic requirements for bandwidth over short intervals of time, it is necessary to integrate SDN-based network control functionalities. For instance, bandwidth-intensive applications such as video-on-demand services, interactive applications such as online gaming that require lower latency or applications which requires forestalling over other applications. These can take advantage of the network monitoring provided by SDN in supporting QoS depending on the network loads (Akinola et al., 2024). Figure 2 shows the traffic classifications of SDN. It highlights the importance of dynamic QoS rearrangement. SDN was applied to optimise the bandwidth allocation of the 5G network. Similarly, SDN controllers have been interfaced with optical networks to improve bandwidth allocation efficiency and adaptability of traffic for hybrid environments (Yang & Tsai, 2024).

Within the SDN architectures, communication across network layers is enabled through Northbound and Southbound Interfaces (NBI and SBI), which connect applications, controllers, and forwarding elements. The NBI in SDN enables applications and management systems to specify service-level QoS requirements,

allowing network configurations to adapt to real-time operating conditions (Alabarce et al., 2020). These requirements are then mapped by the SBI into device-level instructions for flow-level policy enforcement. Although this interface design improves programmability, it also increases implementation complexity, especially in dense networks that require frequent reconfiguration. Moreover, SDN deployment itself presents practical challenges. As observed by Jiang et al. (2023), configuring appropriate SDN functions can be cumbersome, prompting the use of ML-based performance prediction to estimate metrics such as Round-Trip Time (RTT), Switch-to-Controller (S2C) Traffic, and Controller-to-Controller (C2C) traffic before deployment.

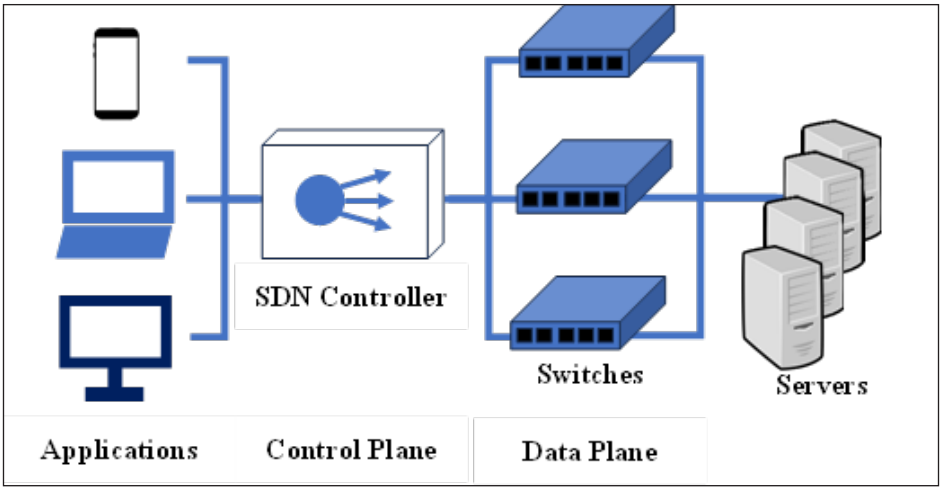


Figure 1. Network architecture of SDN

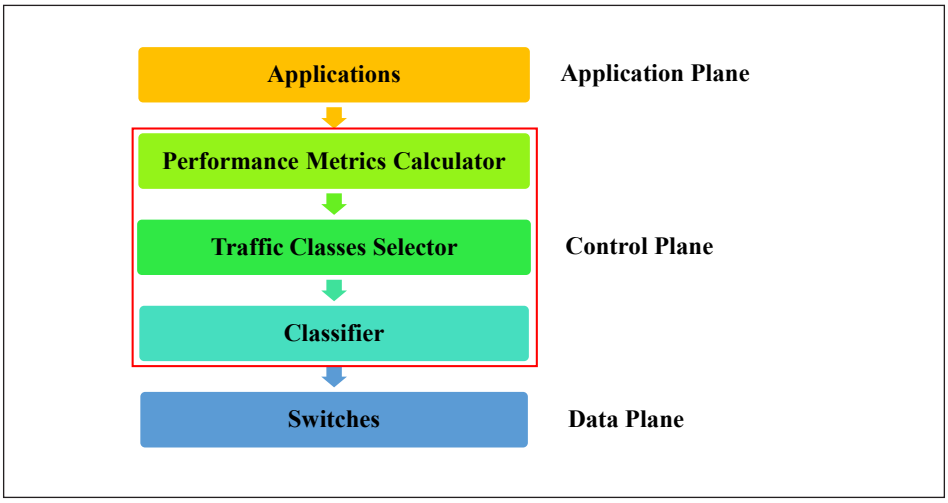


Figure 2. Traffic classifications of SDN

Their findings show that a neural network boosting regression model (NNBoost) achieves superior prediction accuracy compared with baseline ML and deep learning models such as RF, AdaBoost, XGBoost, LightGBM, DNN, and LSTM. This suggests that QoS provisioning in SDN networks needs to consider not only the optimisation during runtime, but also the assessment before deployment to avoid configuration risk, enhance scalability, and reduce trial and error in practice (Jiang et al., 2023).

Overall, SDN-based QoS frameworks provide a solid foundation for user-centric and dynamic QoS provisioning in 5G networks due to the centralised control, programmability, and policy-driven adaptability they offer. However, the reviewed analysis identified the critical trade-offs involved in the integration of SDN, which invariably include an increase in the control plane overhead, interface complexities, and scalability challenges, especially when the density of devices is high. All these challenges suggest that the QoS solutions based on SDN need to be critically evaluated based on the enhancements being offered for QoS and their operational feasibility, readiness, and stability.

SDN can support dynamic QoS provisioning and is feasible for the dense 5G deployment. Many works show the ability of the SDN framework in enhancing the network performance. However, the increased amount of traffic overloaded the QoS provisioning scheduler and potentially shifted the bottleneck from data plane to control plane (Shameli & Rajkumar, 2026). The experimental test result shows that SDN does not shift the complexity, but it provides QoS to support the controllers (Yang & Tsai, 2024). The conditional feasibility of the SDN supports the QoS provisioning in 5G, but a control mechanism should be deployed to manage the traffic from the dense network. A lightweight model is preferable for the edge deployment, and the lightweight SDN controller is suitable for handling the QoS at the edge. The solution can be strengthened by deploying an ML model to provide local prediction at the edge. The system can be managed and coordinated at a high level to maintain consistency among the vendors. This architecture avoids the bottleneck at the server and still provides QoS adaptively.

User-centric Quality of Service (QOS) Management

In the context of 5G networks, user-centric QoS management is seen to go beyond strict service class prioritisation and to better align QoS service delivery with individual user requirements and context. This direction has become even more important with the support of highly heterogeneous applications with different sensitivities to service performance, for which QoS information alone is not sufficient to ensure service stability with frequently changing network states. One of the means to achieve user-centric QoS optimisation is to introduce Radio Access Network (RAN) flexibility using Functional Splitting (FS), where different functions are split between central and edge entities, depending on latency and bandwidth limitations.

This flexibility can increase reliance on QoS for latency-sensitive applications by efficiently determining processing offload and enhancing system responsiveness to changing load conditions. At the same time, FS adds complexity to coordination because of the requirement to ensure persistent splitting strategies in mobility and traffic variations to exclude overhead costs of reconfiguration that might impede performance gains. Thus, various implementations of using FS have proven to make RAN more flexible for performance improvement (Ciceri et al., 2024).

In addition to RAN-level enhancements, the provision of user-centric QoS has also been explored for the integration of the transport network over optical access systems, such as Passive Optical Networks (PON), that provide more efficient bandwidth allocation for a hybrid 5G setting. Optical-based technologies have the potential to deliver better QoS assurance for supporting bandwidth-demanding and delay-sensitive services through the provision of additional network transport capacity, especially with the integration of edge computing technology. Other than the RAN layer, QoS support from a user perspective has also been explored with the integration of the 5G transport segments and optical access platforms, such as Passive Optical Networks, for improving bandwidth distribution and transport efficiency, especially with the potential to do so for high-throughput and delay-sensitive services with the integration of edge computing technology. On the other hand, the extent of the possible gains in QoS is highly dependent on the degree of interoperability between optical access networks and the existing network infrastructure. Legacy issues imply that the performance gains are more likely to be deployment-dependent rather than general (Dias et al., 2023). Moreover, another important set of issues in user-centric QoS management is concerned with adaptive scheduling and real-time prioritisation, specifically for dense and IoT-congested environments. The controlled scheduling is critical to help mitigate the situation, mainly by prioritising real-time traffic. Although the approaches are suitable in dynamic traffic environments, certain trade-offs are encountered with signal overhead, computational costs, and service classification. These limitations become increasingly important under high traffic density, where frequent adjustments within scheduling may affect scalability (Anjum et al., 2024; Albekairi, 2025).

The QoS and QoE are closely related but cannot be treated interchangeably. QoS evaluates the performance of the network by measuring the performance metrics of throughput, jitter, delay and packet loss rate. However, QoE focuses on evaluating the users' performance. The performance metrics for evaluating the user experience are startup delay of the application, bitrate stability, smoothness of the video streaming, and continuity of the application during handover. This review separated the QoE-based solutions that focused on 5G user experience and the QoS-based solutions that focused on protecting real-time services.

Furthermore, the concept of mobility-aware user-centric QoS management has been explored to enhance the QoS in the scenario of dynamic user mobility. Mode-switching schemes, for instance, need to be developed in such a way that they maintain the level of QoS while making decisions based on the anticipated effects of user mobility, which is essential in the scenario of applications where continuous service reliability is required despite the constant switching between modes (Baz et al., 2024). The comparative analysis reveals that the QoS management using deep learning prediction in mobility scenarios has the potential to surpass the existing modes of mobility management while making decisions regarding handover to avoid the loss of packets in real-time, despite generating overhead in the process.

The user-centric QoS approaches consistently enhance differentiation and adaptability in heterogeneous service environments through mechanisms such as RAN flexibility, hybrid architectures, and priority-driven scheduling. Nevertheless, scalability and deployment feasibility remain insufficiently addressed in many existing frameworks. Control overhead, reconfiguration volatility, and fairness trade-offs continue to limit operational adoption. Further work is therefore required to translate user-centric QoS concepts into deployable, implementation-level solutions.

The existing 5G system evaluates the QoS through the following performance metrics, like throughput, packet loss and packet delay. More efforts are needed to consider the users' fairness, mobility and network stability to support the user-centric solution. The user-level QoS provisioning metrics are incorporated in the following analyses that have high potential to improve the user experience.

Bandwidth Allocation

ML-assisted bandwidth allocation has increasingly been explored in SDN-based QoS management frameworks, where traffic flow detection supports dynamic prioritisation under varying traffic conditions (Khairi et al., 2021). It applies a hybrid Decision Tree-Support Vector Machine (DT-SVM) approach to improve bandwidth allocation, ensuring that higher-priority flows receive sufficient resources. Due to the implementation of ML into the SDN system, it lessens the need for manual configuration of network updates and network policies. This also allows better real-time analysis and real-time adjustments based on the network traffic conditions (Chia et al., 2025). Meanwhile, flexible bandwidth partitioning has also been discussed, where higher-priority bandwidth is split into smaller functional allocations to satisfy both latency and throughput requirements (Ciceri et al., 2024). Besides that, the usage of QoS management with the help of ML is evaluated (Raeisi & Sesay, 2023). This study applied a K-Means algorithm, which differentiates between the high-speed and low-speed users in the variable frame rate (VFR) scheme. Although the work targets autonomous vehicle QoS enhancement, its main contribution is mobility-aware bandwidth control by adjusting handover thresholds to prioritise high-speed users.

This supports more adaptive bandwidth allocation across channels, improving overall QoS consistency under varying mobility conditions. Besides, an efficient resource allocation scheme is demonstrated according to the service prioritisation, and this scheme is evaluated on a 100Gb/s Ethernet passive optical network (NG-EPON) upstream (Rawshan et al., 2024). Proportional Fairness allocation has also been proposed to distribute bandwidth according to users' real-time traffic rates (Tanuja et al., 2023). This gives users with high demands for data to receive a share of bandwidth proportionally and not completely sideline the users with lower priorities. However, while ML-driven prioritisation improves adaptability and Proportional Fairness improves multi-user stability, there are certain trade-offs with respect to responsiveness, fairness enforcement, and scalability for a highly trafficked network.

The existing works show that although the user priority and fairness are evaluated, these cannot guarantee the user experience. Hybrid solutions are proposed to reserve bandwidth for the premium users and provide best effort services to the basic users. However, the basic users may be starving when the network is overloaded, and most of the resources are utilised to support the premium users. Various approaches, such as SDN and ML-powered schedulers, are used to improve the premium users' experience. The user-centric QoS provisioning scheme must have a balance between the services and users to optimise the resource allocation.

Resource allocation can be improved using the Max-Min fairness algorithm and Weighted Fair Queuing with a preference for heavy users, besides allocating base bandwidth for other users as well. The adaptive bandwidth allocation requires a dynamic adjustment of resource distribution in response to network congestion or low-demand periods. In addition, dynamic bandwidth allocation frameworks in MIMO-based 5G systems are designed to optimise bandwidth allocation between antenna elements. As a result, spectral efficiency enhancement as well as simultaneous transmission to multiple users are achieved (Prodromos et al., 2024). Deep learning methods have been employed for estimating traffic volumes and allocating bandwidth dynamically for base station antennas. Algorithms such as the Hungarian algorithm and the Minimum Cost Flow algorithm were used to optimise bandwidth distribution. Though such strategies enhance dynamic bandwidth allocation as well as real-time traffic control, their viability can be limited by computation costs and real-time constraints, even in dense deployment. Meanwhile, a system that gives priority to the URLLC traffic in terms of bandwidth is proposed to achieve ultra-low latency, even under high traffic load (Wu et al., 2024). This dynamic allocation prevents under-provisioning for delay-sensitive applications. Regardless of that, the theoretical model proposed by the authors helps in determining the optimal amount of bandwidth required for each service type under varying conditions. Furthermore, efficient handling of the bandwidth is one of the key requirements to fulfil the requirements of the QoS for diverse classes of traffic.

The problem is more prominent when the topic is touched in the field of vehicular or high-mobility scenarios, where maintaining throughput while keeping packet loss under control is naturally difficult. To address this, Adaptive Network Coding (ANC) approaches have been investigated, where Hidden Markov Models are used to estimate packet loss behaviour and adjust coding redundancy in real time (Yin et al., 2020). By adjusting redundancy levels according to observed link conditions, these mechanisms help reduce avoidable retransmissions and improve effective throughput. In network slicing contexts, bandwidth control is often coupled with Virtual Network Function (VNF) migration to maintain service continuity through adaptive resource reassignment and reduced overprovisioning (Vidhya et al., 2025). ML-based control mechanisms react more effectively to traffic variation, while fairness-based schemes stabilise multi-user performance. URLLC-oriented approaches, in turn, are best suited to high-load scenarios that impose strict latency requirements. SDN-ML facilitates rapid policy updates and fairness-based scheduling to improve multi-user stability. The URLLC-focused allocation ensures low-latency performance for delay-critical services. These observations point to the necessity of combining multiple QoS mechanisms in practical deployments.

Adaptive Traffic Engineering and QoS or Quality of Experience (QoE) Management

Effective QoS management in 5G increasingly depends on adaptive traffic engineering, particularly during congestion events where traffic steering and prioritisation decisions directly influence service stability (Beshley et al., 2024). The service-oriented Software-Defined Networking approaches have been introduced to strengthen QoS and QoE. The Service-Oriented Software-Defined Networks (SOSDN) with centralised programmability and supports runtime control of traffic flows and resource policies. This allows prioritisation to be adjusted automatically according to the Service-Level Agreement (SLA) requirements. This capability preserves low-latency paths and critical bandwidth during network congestion. SOSDN employs real-time optimisation instead of static rules to refine prioritisation decisions based on ongoing network measurements. The architecture can accommodate diverse service demands, ranging from latency-sensitive applications such as video streaming to less time-critical traffic. The flow priorities are tuned dynamically according to QoS indicators i.e. delay, throughput, and packet loss. The QoE-oriented approaches have gained attention, particularly for delay-sensitive services where packet loss and jitter have a pronounced impact on perceived user experience (González et al., 2026). For example, a multipath redundancy framework has been introduced to minimise packet loss and latency through redundant paths and intelligent packet reordering, and it was validated through real-vehicle testing and Web Real-Time Communication (WebRTC) streaming experiments (Ito et al., 2025). In addition, SDN-based Radio Resource Management (RRM) has been explored

in a cell-less architecture to enable dynamic rerouting and interference mitigation (Zeyad & Al Janaby, 2025). This is particularly relevant for delay-sensitive mobile users, where mobility and interference significantly influence QoS continuity.

From a cross-study perspective, SOSDN-oriented prioritisation mechanisms are most effective for SLA-driven congestion handling, whereas multipath redundancy approaches show stronger reliability benefits for mobility-dominated streaming scenarios. SDN-based RRM models, in contrast, are better aligned with maintaining QoS continuity in cell-less architectures. Despite these strengths, practical deployment remains constrained by scalability limits and added control-plane overhead in real-time operation.

This paper separates the existing works that were carried out through simulations under controlled environments or experimental tests under the real network limitations. The simulations are carried out in ideal conditions with stable channels, simplified network topology and low signalling overhead. The performance of the proposed solutions may be different in the real 5G deployment. These review strategies to strengthen the practicality of the existing QoS solutions proposed in a real 5G with various network conditions, such as unstable handover, high signalling load and complex user mobility trends.

QoS-driven Network Slicing Management

QoS-aware network slicing has gained substantial research interest as a possible solution to tackle different service requirements in 5G networks, particularly in multi-service environments where resource contention is extremely high. On the other hand, in vehicular communication environments, flexible slice-based resource allocation models have been proposed to ensure that traffic like Vehicle-to-Infrastructure (V2I), which can be referred to as a critical service, is given high priority compared to other non-critical traffic. Less critical services, such as infotainment, are allocated proportionally fewer resources (Tam et al., 2024). Despite the research mentioned above being vehicular service-oriented, it makes clear the general idea concerning real-time QoS-aware resource allocation, where bandwidth allocation happens on a dynamic basis according to service needs and demand. In this regard, slice-aware resource allocation aligns with user-centric QoS requirements, whereby resource allocation policies can adapt continuously according to evolving network conditions and different service requirements.

In general, network slicing enables the logical partitioning of network resources into multiple slices to support the QoS requirements of various traffic. Dynamic slice management is central to sustaining differentiated services (Debbabi et al., 2021). The service-aware orchestration has combined VNF migration with slice-level resource allocation to limit service interruptions and maintain QoS continuity (Vidhya et al., 2025). Despite the isolation and differentiation that slicing provides, real-world deployments still confront coordination and monitoring overheads as well as scalability challenges for real-time control in dense 5G traffic.

QoS-driven slicing frameworks perform well in scenarios that demand strong service isolation and real-time differentiation, such as vehicular and mission-critical applications. The large-scale 5G deployments expose practical limitations related to monitoring overhead, orchestration latency, and the complexity of cross-domain coordination. These continue to affect scalability and operational feasibility.

The coexistence of the URLLC, eMBB, and mMTC is realistic, and the adaptive QoS-based network slicing solutions are proposed to support various types of traffic. Network slicing isolates each service class by allocating dedicated resources. It guarantees the latency of the URLLC, provides stable throughput to eMBB and high-density support for mMTC devices. However, the spectrum efficiency drops if isolation is deployed alone. Adaptive network slicing reallocates the underutilised resources in real time and maintains the QoS of each service. This solution offers a balance between isolation and efficiency for real 5G network deployment.

QoS Provisioning for 5g Network

QoS provisioning in 5G requires coordinated management across multiple layers, spanning radio resource scheduling, access procedures, core-network resource control, and traffic management. 5G deployments highly depend on adaptive mechanisms to preserve QoS stability, particularly when multiple services compete for limited spectrum resources, rather than relying on the static priority-based differentiation.

Although the application of the MDMA techniques improves the capacity of the access network when dealing with competing traffic, the application of the techniques is hindered by the coordination of interferences and the lack of scalability of the system. Beyond physical-layer access mechanisms, the QoS provisioning has been extensively investigated in cloud-based 5G and emerging 6G architectures, where resource allocation increasingly depends on flexible control and multi-metric optimisation strategies. In this context, the importance of QoS-aware resource management capable of supporting service heterogeneity and rising traffic demands has been highlighted in cloud-native network environments (Louvros et al., 2023). A multi-criteria decision-making approach has been proposed to allocate resources based on service priority across multiple QoS dimensions, including latency, jitter, and throughput, supporting a more balanced allocation under high-load conditions. Despite all these efforts, there is still a constraint towards the real-world implementation due to the complexity of orchestration and control overhead in real-time monitoring and frequent policy updates.

Although better performance has been shown in such techniques, challenges persist in terms of theoretical and practical differences in QoS provisioning frameworks. Therefore, even though MDMA might provide better bandwidth utilisation along with supporting high-prioritised traffic, high complexities along with computation might affect its effectiveness

in a larger environment (Rathod & Saxena, 2024). Similarly, in terms of QoS decision-making in a cloud environment, better resource balancing might be achieved, but in such a case, staging delays might also occur. The problem in RAN environments where strict latency requirements need to be met (Louvros et al., 2023). Therefore, any form of QoS provisioned mechanisms requires more than just simulated results but also a practical approach in terms of traffic realities, along with high-density users.

In this context, QoS-based approaches to resource management have reinforced the importance of developing resource management models depending on actual deployment scenarios by identifying the existence of a gap between assumptions and actual scenarios (Umar et al., 2024). Moreover, flexible scheduling approaches have been proposed to facilitate the adaptation of scheduling operations depending on the urgency of traffic and services, improving the alignment between resource allocation and QoS requirements (Aslam, 2025).

Comparatively, though QoS provisioning in MDMA improves access layer access, QoS in native clouds allows the decision control for multiple QoS metrics at the infrastructure level. Nevertheless, both approaches face challenges in dealing with the complexities and scalability issues in achieving timely feedback for QoS.

Functional Split Optimisation in Cloud-RAN

The implementation of functional split strategies demands an appropriate balance of resources and fronthaul bandwidth, especially to meet the QoS requirements while ensuring efficient management of energy consumption. In this context, dynamic and adaptive strategies have also been put forward to efficiently manage the application of resources. This includes the optimisation-based split selection using the Split Hybrid Particle Swarm Optimisation (SPLIT-HPSO) algorithm (Matoussi, 2020a). This strategy is expected to improve the efficiency of decision-making with respect to functional split computation-intensive functions while reducing the cost of deployment. In addition to this, learning-oriented split selection strategies have also been put forward with the application of deep learning models to learn appropriate split selection based on real-time operational parameters. This is expected to improve latency and computation efficiency based on training data and measurements (Ly & Yao, 2021). However, optimisation-driven and learning-based approaches both face practical limitations. The former is often constrained by computational burden, while the latter depends heavily on data quality and training scale, which may restrict real-time use in dense deployments.

In addition, Cloud-RAN introduces the challenge of selecting appropriate split points between centralised and edge units to balance latency, load distribution, and resource utilisation. Medium access strategies that account for functional split constraints and multi-service coexistence have also been discussed (Khodakhah et al., 2025). Scheduling

techniques have been proposed to adapt to fluctuating service demands while optimising control signalling overhead across distributed network nodes. This supports more effective C-RAN operation in dense urban 5G scenarios, although frequent split adaptation may increase coordination overhead across distributed nodes.

User-centric RAN Slice Allocation

User-centric RAN slice resource allocation has emerged as an important research direction for improving QoS flexibility in 5G systems. The End-to-End User Slice Allocation (E2E-USA) approach has been investigated using hybrid optimisation methods to address the constraints across radio, transport, and computing resources. This strategy iteratively determines slice placement while maintaining QoS targets under varying traffic conditions (Matoussi, 2020a). Vertical slicing has been incorporated into 5G QoS architectures, where traffic classes such as URLLC and eMBB are mapped to dedicated slices according to their performance requirements (Xie et al., 2022). Such designs enable more efficient resource utilisation by redistributing bandwidth when slices operate below capacity or experience reduced demand. Despite their emphasis on isolation and adaptive control, both approaches continue to face challenges in achieving scalable and consistently responsive operation, particularly during abrupt traffic changes and in dense deployment environments.

Studies on user-centric slice allocation have explored models that dynamically adjust slice properties according to user profiles, location, and mobility patterns. These models optimise the usage of the spectrum resources while also providing the service level assurances in the presence of varying user conditions, which underlines the importance of personalisation at the user level in the RAN to manage slices (Matoussi et al., 2023b).

Deep Learning for QoS Management

Deep learning has been drawing a lot of attention as an emerging solution for efficient management of QoS in 5G communications, owing to its capability of providing real-time adaptability and data-driven decision support. Besides that, learning-based user slice assignment has been researched using deep learning-based neural networks, looking at achieving fast user RAN prediction or slicing with efficient utilisation of network resources (Ly & Yao, 2021). Meanwhile, context-driven network selection solutions have also been researched for enhancing network access decisions, considering various real-time parameters like user requirement, received signal strength, as well as the network congestion status (Honarvar et al., 2022). These approaches highlight the advantage of predictive intelligence in adapting QoS decisions to evolving network conditions, especially in environments with fluctuating network demand and mobility-driven variation. However, there are also significant operational complexities incorporated in the QoS frameworks with the help of deep learning models. In most cases, large-scale training data is used

with these models, which are then required to be periodically retrained to be able to cope with the non-stationary network conditions. In addition to this, the inference latency and operational feasibility are also the major concerns when deep learning models are incorporated into the QoS control loops. Thus, when we talk about beyond deep learning, there are also some light learning-based strategies considered. As discussed above, these strategies have also been incorporated into QoS optimisation problems, wherein the major concern is to maintain responsiveness with a reduction in computational complexity. In this regard, link quality modelling with a Hidden Markov Model (HMM) has also been used in order to enable real-time adaptation of network coding redundancy, which enhances bandwidth efficiency with a reduction in unnecessary retransmissions under varying loss conditions (Yin, et al., 2020).

Comparatively, the deep learning-based QoS management solution has stronger adaptability based on prediction in terms of slice allocation and access, while the lightweight ML solution, such as HMM-based adaptation, provides lower-complexity adaptation mechanisms for bandwidth optimisation. However, one of the most important research gaps is to achieve scalable and deployable learning-based QoS solutions for dense 5G networks.

The ML-based QoS schemes support adaptability, but they come with fundamental performance degradation. Extra resources are required by the high-performance deep learning models (LSTM, CNN, etc.) that require higher energy consumption to support the larger amount of computation with longer computational time, which is not available at the edge nodes. It may hurt the latency-sensitive URLLC. The higher accuracy provided by the SDN scheme comes at the trade-off of the heavy computational load (Jiang et al., 2023). Lightweight models are proposed for edge nodes with faster judgment, but the accuracy is lower in active scenarios (Yin et al., 2020). A balance between the accuracy, latency and edge cost for the ML-driven provisioning schemes. The edge nodes support the lightweight models for the latency-sensitive deployment. The deep learning models are preferable for high-precision prediction applications.

The additional energy consumption by the complex ML-based QoS control, especially at the edge, remains a challenge. Deep learning models require longer processing cycles and larger memory access, which significantly increases the power consumption. The proposed solution should maintain a balance between maintaining energy efficiency and adaptive QoS provisioning. The lightweight ML models with lower energy consumption are preferable for edge deployment.

Traffic Management through Network Slicing

Traffic management using network slicing is extensively researched as an important QoS support feature in the context of 5G networks, particularly during congestion states where multiple services are competing to access shared infrastructure resources.

Vertical slicing enables the segmentation of service traffic according to the requirements of the application, thereby providing more predictable QoS performance through service isolation. Moreover, the use of slice-aware bandwidth re-allocation has been recognised as one of the key utilisation methods, whereby idle slices' underutilised resources may be used to support active slices to avoid underperformance for high-priority traffic (Xie et al., 2022). Apart from traffic segregation, another approach that has been mentioned as a secondary option that could potentially reduce fronthaul bandwidth requirements, without compromising QoS for high-priority services, is the functional split approach, particularly in computation-intensive RAN scenarios (Matoussi, 2020a). These techniques improve flexibility in slice-level traffic handling, but they introduce additional management complexity, particularly during abrupt traffic surges that require frequent reconfiguration. Slice-aware traffic control mechanisms have been investigated across both edge and core network domains. For instance, the robust distribution of traffic in Multi-access Edge Computing (MEC) based crowd sensing has been researched to minimise the deterioration of services in case of an overload situation (Xiang et al., 2025). SDN-based slicing architectures have also been investigated to enhance the centralised control of traffic in slices to improve isolation while providing effective orchestration (Ramesh et al., 2024). A cross-comparison of these methods indicates that vertical slicing, together with dynamic redistribution of unused resources, can improve utilisation efficiency and service differentiation. The function splitting mechanisms are helpful in optimising fronthaul for latency-critical services, SDN and MEC-based control that can improve slice isolation in overload situations. Nevertheless, the scalability of 5G networks is a challenge in the implementation of real-time slices due to orchestration, monitoring, and scalability issues.

Though the simulation-based investigations are beneficial from a theoretical point of view, in practical application scenarios, issues regarding scalability as well as latency are more challenging to overcome. For instance, the SDN infrastructure has demonstrated its efficiency across a range of testbed environments, but poses inherent issues regarding its control plane overhead when applied to 5G scenarios (Rathod & Saxena, 2024). ML models have demonstrated a superior capacity for increased efficiency under testbed environments but remain more challenging to retain when dealing with dynamic environments that are perpetually changing (Louvros et al., 2023). Field tests performed by (Ito et al., 2025) have effectively demonstrated the QoS improvement for video streaming by leveraging the benefits of multipath redundancy.

Table 2 presents a comparison of leading user-centric QoS application methods in terms of their design focus, limitations, and applicability. One study investigates the application of AI-based assistance for a 6th-generation network that focuses on a novel shift from a network function-based approach to a user-centric approach. Despite its exploratory value, inadequacies regarding its scalability and QoS metrics affect its application and applicability, respectively (Gkatzios et al., 2024). In another study, real-time user feedback

for IoT systems was investigated to promote a shift in its resource allocation approach, changing it to a form of a closed-loop system to allow for increased adaptability. Despite its potential and value, a limitation regarding its ability to perform real-time changes to the QoS (Chow et al., 2026).

Table 2
Comparative analysis of research on QoS enhancement and resource allocation

Title	Project statement	Objective	Limitation
A Proof-of-Concept Implementation of an AI-assisted User-Centric 6G Network (Gkatzios et al., 2024)	Explores the shift from network function-centric to user-centric designs in 6G slicing techniques. This change aims to address the illusion of user-centricity in 5G caused by shared network functions serving multiple users.	Proposes a self-organising, AI-assisted network that adapts to individual user requirements, enhancing scalability and user-centric service delivery.	Limited scalability due to reliance on narrow QoS metrics, difficulties in managing Message Queuing Telemetry Transport (MQTT) feedback frequency, and risks of network overhead.
Design-oriented user-centric QoS scheme for 5G (Chow et al., 2026)	Proposes a user-centric QoS provisioning scheme for 5G that prioritises the user requirements.	Designed an SDN-based control framework integrated with ML prediction and network slicing to guarantee the users under various loads.	Handover variability and multi-vendor orchestration challenges were not evaluated.
QoS-driven Selection of Web Services Considering Group Preference (Wang et al., 2015)	Examines the group preferences and QoS attributes influence web service selection in composite architectures.	Balances individual and group preferences in selecting web services, ensuring alignment with QoS criteria and stakeholder expectations.	Computational complexity escalates with the number of services and attributes, lacking real-world dynamic validations.
Edge User Allocation with Dynamic Quality of Service (Lai et al., 2019)	Proposes a dynamic allocation mechanism that prioritises QoS metrics for edge servers in fluctuating network conditions.	Adapt allocation dynamically to real-time changes in network conditions, improving resource utilisation and user satisfaction.	Susceptible to instability in resource-constrained or wide-scale scenarios, with insufficient focus on long-term signalling overhead.
Self-adaptive Resource Allocation for Cloud-based Software Services Based on Iterative QoS Prediction Model (Chen et al., 2020)	Suggests a predictive resource allocation framework for cloud services using iterative QoS forecasting.	Optimises resource use and ensures SLA compliance by predicting and adapting to changing workloads and conditions in real time.	Faces challenges in volatile environments due to predictive inaccuracies and significant computational overhead, limiting real-time applicability.

Additional slicing-related SDN architectures are also discussed beyond Table 2. While several works explore the use of 5G network clustering techniques in predicting users' behaviours for better QoS provisioning. These papers also have some challenges in terms of the poor quality of data and scalability when used in larger systems (Ramesh et al., 2024). Another paper has been written on quality-of-service-driven web service selection based on individual and group preferences in such a manner that this selection is often restricted due to challenges in terms of being computationally complex and not being well validated in real scenarios (Wang et al., 2015). Similarly, dynamic quality-of-service-aware users' allocation mechanisms in edge servers in scenarios where network conditions may become unstable have been explored to a greater degree (Lai et al., 2019). This review differs from prior studies by others due to its emphasis on not only algorithmic interpretations but also using human-centred approaches to this topic in a better manner. It critiques existing static frameworks and advocates for dynamic solutions using SDN and ML integration (Chen et al., 2020). Furthermore, an iterative QoS prediction model for cloud-based services enhances resource utilisation. However, it also poses challenges in terms of predictive inaccuracies and computational overhead in volatile environments (Vidhya et al., 2025).

As such, it reveals that there is an inevitable need for an adaptable solution which can satisfy the continuously expanding needs of 5G as well as future technologies, while at the same time showing evident gaps in the subject matter which remain unaddressed by researchers. SDN-based QoS management has an advantage in real-time traffic management, but there are challenges with scalability, especially for high-density urban areas. On the other hand, the ML-based methods have the advantage of high flexibility and adaptability to dynamic traffic conditions, but are challenged by high computation needs, which makes them suitable for IoT and edge computing. Network slicing, particularly for URLLC services, has the advantage of high reliability but is challenged by interference between slices.

Every approach has its own set of benefits targeting specific user scenarios. For instance, SDN-based approaches support centralised control and reconfiguration of QoS but may suffer from single points of failure and scalability problems. ML-based approaches support prediction-based flexibility but pose overhead concerns that must be optimised for effective deployment. Slice-based allocation of resources supports robust QoS guarantees but requires a tighter orchestration to prevent resource conflict. At the same time, network coding and redundancy schemes enhance reliability and performance in lossy or wireless environments, but at the cost of added complexity and bandwidth usage. As shown in Table 3, the comparative analysis demonstrates that no single approach is universally superior. Instead, future research should explore hybrid architectures, e.g., combining SDN with lightweight ML agents at the edge or integrating slicing with redundancy management to optimise user-centric QoS for diverse 5G and B5G application domains.

To further clarify the difference in techniques shown in Figure 3, the taxonomy chart categorises user-centric QoS approaches on four dimensions, i.e. architectural model, intelligence layer, goal of performance, and validation method. Architecture-based QoS solutions are proposed in different layers, such as RAN, the core network, or the edge/fog layer, often by using SDN, Network Function Virtualisation (NFV), or container services. Intellectually, the approaches range from traditional rule-based configurations (e.g. 5QI, Access Class Barring (ACB), and QCI) all the way to advanced solutions powered by ML (e.g., Hidden Markov Models (HMM), Deep Reinforcement Learning (DRL), and Federated Learning (FL)). The objectives span multiple target domains, including URLLC, eMBB, mMTC, and maintaining QoE consistency under mobility. The validation methods include theoretical modelling, NS-3 and MATLAB simulations, experimental tests, and real-world deployment experiments with different levels of realism.

Table 4 shows that the selected case studies demonstrate the usability and efficiency of user-centric QoS approaches through real-world tests and simulations. Techniques like IP-layer path redundancy, adaptive network coding based on machine learning, and dynamic user profile-based slice allocation have exhibited measurable QoS improvement in metrics like packet loss, throughput, and fairness.

Table 3
Comparative summary of user-centric QoS techniques

Technique Category	Strengths	Limitations	Use Case Focus	Validation Method	Research Gap
SDN-based QoS	Programmable control, dynamic policy updates	Centralised bottlenecks, latency in large-scale networks	eMBB, general mobile broadband	Simulation, real-world prototypes	Need for distributed SDN controllers
ML-based prediction	Real-time adaptability, behaviour-aware decisions	High computational demand at edge/cloud	High mobility, IoT, VR/AR	Simulation and testbeds	Lightweight and energy-efficient models
RAN slice allocation	Fine-grained isolation, customisable service levels	Inter-slice interference, complex orchestration	URLLC, mission-critical services	Conceptual and simulation	QoS assurance across slices
Multipath redundancy	Enhanced reliability and robustness to handovers	Buffering overhead, complexity in path selection	Mobile video streaming, autonomous systems	Real-world field testing	Smart path switching algorithms
Adaptive network coding	Improved throughput, error recovery	High redundancy cost, coding complexity	Vehicular communications, low-latency uplinks	Simulation and small-scale testing	Scalable coding under mixed traffic

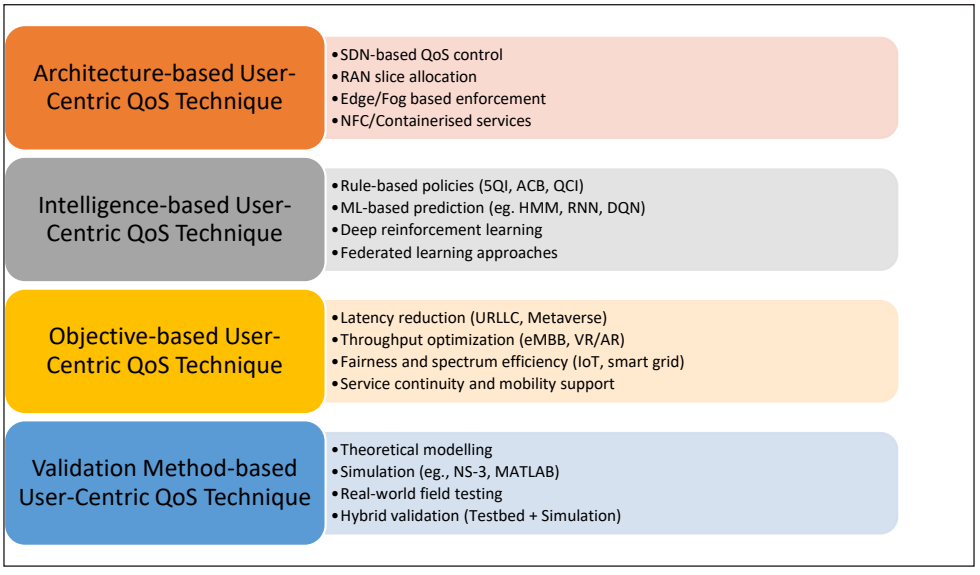


Figure 3. 5G QoS user-centric hierarchical structure

The results confirm the theoretical explanations summarised in this paper and reveal promising directions for scalable, smart QoS management in future 5G deployment. Furthermore, the experimental results from the selected case studies corroborate the applicability of user-centric QoS solutions for 5G networks. For the Multipath Redundancy for Media Streaming (Ito et al., 2025), a real-time vehicular deployment with GENEVE tunnelling achieved a packet loss ratio as low as 0.002%, stable delivery of bitrate and only a slight rise in latency (~38ms), verifying the stability of the system at high-mobility situations. The Adaptive Network Coding case (Yin et al., 2020) with HMM-based prediction on NS-2 achieved 18-25% throughput gain and up to 40% PLR reduction over static methods. Similarly, the User-Centric RAN Slice Allocation study from (Matoussi et al., 2023b) shows improved spectrum utilisation by up to 15%, then fairness index by 12%, and inter-slice contention reduction.

These results highlight that adaptive, ML-founded, and user-aware strategies provide tangible gains in essential QoS metrics, namely in dynamic and heterogeneous 5G environments. The case studies also reveal gaps and avenues for enhancement. For example, most of the implementations were tested under control or single-scenario conditions. Future work should thus target:

- Scale testing under gigantic user equipment (UE) density and mixed traffic types.
- Trade-offs in energy efficiency during ML inference at the edge.
- Long-term service continuity via handovers or network slicing reconfigurations.
- Deployment of real-time orchestration tools for runtime automation of QoS enforcement.

Table 4
Comparison of the case studies

Case Study Title	Scenario Type	Technique	Evaluation Platform	Key QoS Metrics Improved
Multipath Redundancy for Mobile Communication (Ito et al., 2025)	Real-World Field Test	IP-layer multipath & GENEVE tunnelling & packet reordering	Live vehicular test with WebRTC	↓ Packet loss (to 0.002%), ↑ Bitrate stability, ↔ Latency (minor increase)
Adaptive Network Coding for Vehicular Network (Yin et al., 2020)	Simulation & Experimental	HMM-based packet loss prediction & adaptive coding	NS-2 + real vehicular setup	↓ Packet Loss Ratio (PLR), ↑ Throughput efficiency
User-Centric RAN Slice Allocation (Matoussi et al., 2023b)	Simulation	Mobility & profile-based dynamic slice assignment	Custom 5G network simulator	↑ Spectrum utilisation, ↑ Fairness index, ↓ Inter-slice contention

Many works are carried out through simulations or small testbeds. A practical evaluation framework is used to judge the practicality of real 5G deployment for each work. The framework analyses the similarity of the network topology used in the simulation with the real 5G network. For example, the conditions such as traffic loads, users' mobility, interference and handover are modelled according to the real 5G network. The ability of the framework to support a scalable number of users is analysed. The feasibility and complexity of deploying the proposed framework into the existing 5G network are studied. The robustness of the proposed framework under the overloaded conditions is evaluated. The review evaluates the performance in the controlled testbeds and the practicality of each QoS solution in the real 5G network.

Collectively, these studies emphasise the urgency of coming up with cross-layer, modular, and policy-aware QoS solutions that can scale up readily to user behaviour, network conditions, and deployment constraints in real 5G and beyond-5G environments. Furthermore, despite the long upgrades and changes in the 5G network, there are still some outstanding research gaps in terms of the User-Centric QoS aspect. Despite the advancements, several research challenges remain unresolved in the current literature as follows:

- Lack of real-world testing for ML-driven QoS models under dynamic traffic conditions.
- Limited support for seamless interworking of QoS across heterogeneous networks (e.g., 5G, Wi-Fi, IoT).
- Minimal integration of regulatory policies (e.g. General Data Protection Regulation (GDPR), SLA enforcement) into adaptive QoS mechanisms.
- Inadequate exploration of lightweight, privacy-preserving ML models suitable for mMTC.
- Underdeveloped feedback loops between QoE metrics and network-level QoS adjustments.

The user-centric QoS schemes support the 5G network and do not fully evaluate the interoperability of heterogeneous networks with the coexistence of WiFi, fiber optic networks and 4G. It remains a challenge for the real deployment of the QoS schemes across the heterogeneous network with various performance constraints and requirements. The future lightweight user-centric QoS schemes should support cross-layer operations and guarantee consistent QoS across the heterogeneous network. Future research must also protect the privacy issues of processing and collecting the user data, especially in the user-centric framework. The mobility patterns, service usage and preference are sensitive and need protection. The proposed solutions must deploy privacy protection schemes to protect all users' information. The adaptive decisions made by the solutions protect the users' privacy.

To consolidate the comparative insights derived from the reviewed QoS frameworks and selected case studies, Table 5 summarises the most suitable user-centric QoS mechanisms across representative 5G deployment scenarios, while accounting for key practical deployment constraints. The comparative insights from the reviewed SDN, ML, and slicing-driven QoS frameworks were consolidated by mapping each major solution class to the 5G scenario where it is most technically effective. It provides a scenario-oriented interpretation of the user-centric QoS approaches in 5G. The same mechanism may yield significant gains in practical deployments for one condition (e.g., mobility-driven streaming), but it may be impractical for another condition (e.g., ultra-dense mMTC). The potential reasons that caused the performance degradation are overhead, scalability, or latency constraints. It states the advantages and limitations of suitable QoS approaches to be deployed in the mobile network.

From the comparative mapping, SDN-based mechanisms are most suitable in scenarios requiring rapid and policy-driven traffic control, such as congestion-aware traffic engineering for eMBB services. Their key advantage is centralised programmability, which allows real-time enforcement of prioritisation and routing policies across flows. However, the performance of SDN degraded when the network is overloaded.

These may be caused by the frequent flow updates, control-plane overhead, and scalability constraints due to dynamic policy reconfiguration activities. In contrast, ML-based QoS frameworks are more efficient in highly variable environments where predictive traffic awareness enables more proactive allocation decisions, such as adaptive bandwidth allocation and mobility-aware optimisation. Nevertheless, the effectiveness of ML-based models may be dependent on the quality of the training data, the latency of the models, and the associated cost of the models, which may limit the applicability of the models to meet the URLLC deadlines.

Table 5 also highlights that network slicing provides the strongest performance isolation for scenarios requiring deterministic QoS guarantees, such as URLLC and mission-critical applications. Slice-aware mechanisms improve reliability by reserving and isolating resources for critical services, but they remain constrained by orchestration complexity, monitoring overhead, and cross-domain coordination issues when deployed end-to-end across RAN and core networks. Furthermore, mobility-centric strategies, such as redundancy techniques and adaptive coding schemes, have been proven to be very effective in vehicular streaming and handover-aware scenarios, where packet loss is minimised, and throughput is maximised. Nevertheless, such techniques are accompanied by additional bandwidth, diversity, and coordination overhead, which might be more pronounced in heavily loaded networks.

There is no universally optimal user-centric QoS mechanism for all 5G scenarios. Instead, a more viable approach to QoS provisioning would be facilitated by a hybrid approach that leverages the capabilities of SDN for policy enforcement, ML for traffic intelligence, and slicing for isolation. This, in turn, points to the primary research gap within the existing literature, where most of the solutions are still validated in simulations or small-scale testbeds, with the feasibility of deployment at scale, as well as the constraints of real-time operation under mixed traffic conditions, being inadequately investigated.

A practical user-centric QoS framework that consists of SDN for fast deployment, network slicing for isolation and lightweight ML for prediction should have the following properties. A simple operation cycle for the edge deployment. The real-time monitoring of the key performance metrics. SDN and network slicing enforced the policy translated by the orchestration layer for QoS provisioning in real time. The user-level data is protected by the framework.

Challenges in 5G QoS Security

As shown in Figure 4, the deployment of user-centric QoS models in 5G networks is hindered by a set of interdependent challenges at computational, security, and regulatory levels. Computational complexity remains a key barrier, where high model latency and constrained edge-device resources weaken the real-time practicality of ML-assisted QoS enforcement.

Table 5
Scenario-based comparative summary of user-centric QoS approaches in 5G

5G Scenario / Service Context	Most Suitable QoS Approach	Why Is It the Best Fit (critical reasoning)	Main Limitation/ Trade-off
URLLC (mission-critical/low latency)	QoS-driven slicing + strict priority scheduling	Ensures isolation and latency guarantees under heavy load	Orchestration delay and control overhead
eMBB congestion (high throughput demand)	SDN-based traffic engineering + adaptive prioritisation	Supports centralised control and dynamic policy enforcement	Scalability bottleneck in dense deployments
High mobility (vehicular streaming)	Multipath redundancy (path diversity + reordering)	Reduces loss during handovers and stabilises QoE	Extra bandwidth and coordination overhead
Loss-sensitive channels (vehicular uplink)	Adaptive Network Coding (HMM/ML-based prediction)	Improves throughput by reducing retransmissions under burst loss	Coding overhead and computation cost
Dense multi-user fairness scenario	Fairness-based allocation (PF / WFQ / Max-Min)	Prevents starvation and stabilises user experience	Slower reaction to sudden traffic bursts
mMTC/massive IoT	Lightweight ML + rule-based prioritisation	Scales better with device density and avoids heavy inference cost	Limited accuracy vs deep learning
B5G/edge-cloud service placement	Edge-cloud orchestration + DRL-based optimisation	Enhances continuity across user mobility regions	Retraining needs, deployment complexity

This implies that while user-centric QoS approaches can provide performance benefits in a controlled environment, their practical applicability depends on whether such approaches can function within a tight delay constraint without overburdening the edge resources.

From a security point of view, the nature and types of threats emerging in the context of QoS management in a 5G environment appear to be getting increasingly varied and sophisticated. This includes authentication threats, data privacy issues, network slicing issues, edge computing risks, architectural risks in SDN, and those associated with mMTC. Man-in-the-middle attacks have also continued to pose a significant threat, especially in situations where traffic management and reconfiguration occur frequently, thus increasing the chances for such attacks to occur. The effectiveness and suitability of QoS-optimisation approaches should be evaluated through a security lens that evaluates their ability to preserve trust and confidence in operation.

The large-scale deployment of QoS adaptation mechanisms may also face constraints related to regulatory and policy requirements. Especially, when the adaptation of QoS is related to the enforcement of SLA and data protection between different countries.

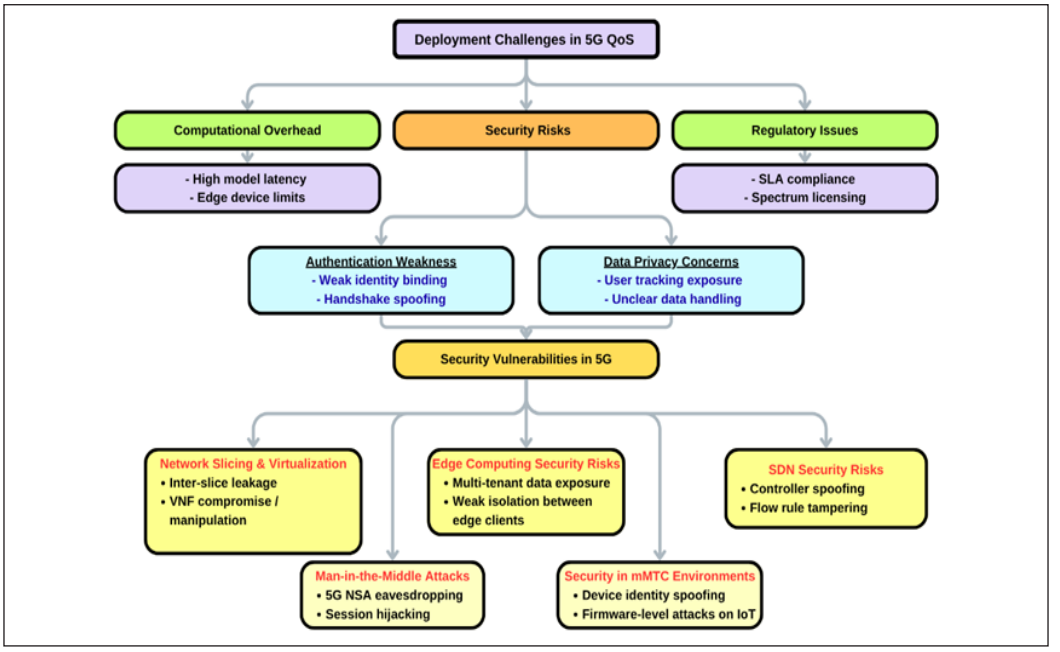


Figure 4. 5G QoS Deployment challenges, security and operational risks flowchart (Fakhouri et al., 2023)

This becomes more apparent when the system needs to cope with multiple operators and multiple vendor environments, especially when there are requirements related to the balancing of policy requirements with the need for high QoS responsiveness. Thus, since security and regulatory issues are anticipated in this manner, it is important to ensure that QoS techniques in 5G networks always remain reliable and safe.

Deployment Challenges and Mitigation Strategies

As discussed earlier, the deployment of user-centric QoS frameworks in real-world 5G environments remains beset with a variety of challenges, ranging from the ability to cope with different network environments, which require a non-trivial integration of AI SDN-QoS with conventional 5G RAN QoS, as well as a variety of scaling challenges, particularly in ultra-dense network environments as well as IoT-centric use cases. It should also be pointed out that the inherent latency overhead of real-time traffic classification and actual decision-making also contributes to the challenges in terms of responsiveness.

To reduce the challenges associated with the above issues, certain strategies have already been experimented with, as described below. One such strategy for the deployment of fog computing within the network includes the “OpticalFog Node Architecture.” In an operational surveillance deployment within a university environment, this architecture reported improvements in latency performance, energy efficiency, and bandwidth utilisation,

while applying SDN-based shortest-path routing to enhance traffic steering efficiency (Singh & Sood, 2020). The emerging B5G has the potential to support diverse mobile user requirements by combining edge-cloud orchestration with deep reinforcement learning. A Double Deep Q-Learning Recurrent Neural Network (DDQL-RNN) framework has been applied to predict user behaviour and support service placement through a water-filling allocation strategy. It considers both the latency and cost. Simulation-based evaluation suggests that this approach can help maintain service availability across the user journey (Farhoudi et al., 2023).

These architectures reviewed report notable QoS improvements. They also reveal a recurring limitation. The validations are still largely confined to simulation-based or tightly controlled test environments. The complementary strategies, such as federated or lightweight ML, intent-driven SDN, and containerised VNFs, improve the deploy ability and the implementation cost. These can help to narrow the gap between simulation studies and operational deployment for the success of deploying the user-centric QoS models in the real 5G networks.

Open Challenges and Future Research Areas

The user-centric QoS frameworks still face open challenges that affect scalability, interoperability, and practical deployment. The development of low-latency, privacy-aware ML models that can run efficiently at the network edge is highly demanded. These models enable real-time decision-making without imposing excessive computational overhead on end devices. Secondly, QoS interoperability over heterogeneous networks, including Wi-Fi 7, satellite, and legacy LTE, is still unexplored, especially for handover scenarios or multi-access convergent scenarios. Security and privacy concerns continue to exist as well, especially in AI/ML-based systems where inference models become vulnerable to adversarial attack or data compromise. On the other hand, another important challenge to be addressed is the integration of regulatory compliance, such as SLA compliance and GDPR, with dynamic QoS systems. From an operational point of view, there is no common orchestration framework that can enforce QoS policies across different network domains, vendors, and services.

The above deficiencies must be remedied by subsequent research efforts in developing as follows:

- Edge-compatible, resource-aware ML models for QoS adaptation.
- Secure, regulation-aware QoS management frameworks.
- Federated learning-based architectures for privacy-aware traffic classification.
- Cross-layer QoS methods involving application, transport, and network policies.
- Unified orchestration platforms across cloud, edge, and RAN domains.

CONCLUSION

Overall, SDN and ML-based mechanisms remain central enablers for adaptive QoS provisioning in 5G. SDN supports real-time, agile control of network resources while ML enhances the ability of the network in terms of predicting the pattern of traffic and further optimises bandwidth allocation and dynamically readjusts QoS priorities. When combined with techniques such as network slicing, these mechanisms can improve service differentiation across high and low-priority traffic classes. Despite such significant progress, the current implementation of most of these works is usually not satisfactory to meet diversified user demands, especially under high-traffic scenarios. This review identifies a persistent research gap in achieving seamless and scalable integration between SDN control logic and ML-driven decision modules for real-time QoS adaptation. Although SDN and ML demonstrate strong potential individually, their integrated deployment has not yet been fully exploited for practical QoS optimisation in 5G networks. The approaches discussed emphasise the need to ensure that premium users get the bandwidth guaranteed to them, while at the same time using the free resources intelligently to satisfy other non-premium users. The reallocation of resources has the effect of not just improving network performance but also of providing a better user experience by balancing efficiency with satisfaction. In contrast, several areas will need further research in the future. First, the scalability of the solutions of SDN-ML must be ensured for large and densely populated networks. The existing systems cannot bear such a high demand and variability typical of 5G environments. Moreover, the improvement of the accuracy of ML models used in traffic prediction and energy-efficient models will be of prime importance to enhance the performance of the 5G network while being more sustainable. In addition, the issue of interoperability with other complementary access technologies, such as Wi-Fi and IoT, should be further investigated to enable the development of hybrid QoS models that are capable of seamless operation across different heterogeneous environments. Finally, it is recommended that federated learning be further investigated as a technique that can be used to enhance the decision-making of adaptive QoS models while maintaining the privacy of users and minimising the dependency of these models on centralised data. In order to make these models more practical, it is necessary that extensive experiments are conducted while keeping the 5G network evolving so that the full potential of 5G is achieved by integrating SDN and ML.

In summary, this critical review establishes a framework of reference for future research on scalable and user-centric QoS management by critically assessing the positive aspects as well as the limitations of SDN, ML and slicing-based approaches. Further research is needed to enhance the feasibility, scalability, and real-time efficiency of these approaches, especially in view of the increasing demands of network applications for higher reliability, lower latency and more efficient resource use.

These areas will be critical for achieving more adaptive, scalable, and energy-efficient QoS management in future 5G and beyond 5G networks.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Higher Education (MOHE), Malaysia, through the Fundamental Research Grant Scheme (FRGS/1/2024/ICT06/TAYLOR/02/1).

REFERENCES

Abuajwa, O., Roslee, M., Yusoff, Z. B., Chuan, L. L., & Leong, P. W. (2022). Resource allocation for throughput versus fairness trade-offs under user data rate fairness in NOMA systems in 5G networks. *Applied Sciences, 12*(7), 3226. https://doi.org/10.3390/app12073226

Akinola, O. I., Olaniyi, O. O., Ogungbemi, O. S., Oladoyinbo, O. B., & Olisa, A. O. (2024). Resilience and recovery mechanisms for software-defined networking (SDN) and cloud networks. *Journal of Engineering Research and Reports, 26*(8), 112-134. https://doi.org/10.9734/jerr/2024/v26i81234

Alabarce, M. G., Bravalheri, A., & Marino, P. P. (2020). INSPIRING-SNI: Investigating SDN programmability improving optical south-and north-bound interfaces. In *Proceedings of the 2020 22nd International Conference on Transparent Optical Networks (ICTON)* (pp. 1-4). https://doi.org/10.1109/ICTON51198.2020.9203409

Albekairi, M. (2025). Controlled service scheduling scheme for user-centric software-defined network-based Internet of Things. *IEEE Access, 13*, 19198-19218. https://doi.org/10.1109/ACCESS.2025.3533310

Al-Shammari, B. K. J., Al-Aboody, N., & Al-Raweshidy, H. S. (2018). IoT traffic management and integration in the QoS-supported network. *IEEE Internet of Things Journal, 5*(1), 352-370. https://doi.org/10.1109/JIOT.2017.2785219

Amin, R., Rojas, E., Aqdus, A., Ramzan, S., Casillas-Perez, D., & Arco, J. M. (2021). A survey on machine learning techniques for routing optimisation in SDN. *IEEE Access, 9*, 104582-104611.https://doi.org/10.1109/ACCESS.2021.3099092

Anjum, M., Min, H., & Ahmed, Z. (2024). User-centric Internet of Things and controlled service scheduling scheme for a software-defined network. *Applied Sciences, 14*(11), 4951. https://doi.org/10.3390/app14114951

Aslam, U. (2025). Designing flexible scheduling algorithm for 5G. *International Journal of Advanced Engineering Management and Science, 11*(1), 90-108. https://doi.org/10.22161/ijaems.111.7

Baz, A., Logeshwaran, J., Natarajan, Y., & Patel, S. K. (2024). Enhancing mobility management in 5G networks using deep residual LSTM model. *Applied Soft Computing, 165*, 112103.https://doi.org/10.1016/j.asoc.2024.112103

Beshley, M., Kryvinska, N., Beshley, H., Panchenko, O., & Medvetskyi, M. (2024). Traffic engineering and QoS/QoE supporting techniques for emerging service-oriented software-defined network. *Journal of Communications and Networks, 26*(1), 99-114. https://doi.org/10.23919/JCN.2023.000065

Chen, X., Wang, H., Ma, Y., Zheng, X., & Guo, L. (2020). Self-adaptive resource allocation for cloud-based software services based on iterative QoS prediction model. *Future Generation Computer Systems, 105*, 287-296.https://doi.org/10.1016/j.future.2019.12.005

Chia, R., Pang, W. L., Phang, S. K., Goh, H. H., & Chan, K. Y. (2025). Machine learning-driven analysis of user bandwidth allocation and performance in 5G network. *IEEE Access, 13*, 173081-173095. https://doi.org/10.1109/ACCESS.2025.3615398

Chow, W. H., Pang, W. L., Goh, H. H., Tee, W. H., Md Rezali, F. A., Chan, K. Y., & Chung, G. C. (2026). Design of user-centric QoS provisioning scheme in 5G network. *Journal of Engineering Science and Technology*, 28-50.

Ciceri, O. J., Astudillo, C. A., Zhu, Z., & da Fonseca, N. L. S. (2024). Bandwidth allocation for multiple functional splitting options over TWDM-EPON networks with multi-ONU customers. In *ICC 2024-IEEE International Conference on Communications* (pp. 468-473). https://doi.org/10.1109/ICC51166.2024.10622322

Debbabi, F., Jmal, R., & Chaari Fourati, L. (2021). 5G network slicing: Fundamental concepts, architectures, algorithmics, projects, practices, and open issues. *Concurrency and Computation: Practice and Experience, 33*(20). https://doi.org/10.1002/cpe.6352

Debnath, R., Akinci, M. S., Ajith, D., & Steinhorst, S. (2023). 5GTQ: QoS-aware 5G-TSN simulation framework. In *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)* (pp. 1-7). https://doi.org/10.1109/VTC2023-Fall60731.2023.10333533

Dias, I., Ruan, L., Ranaweera, C., & Wong, E. (2023). From 5G to beyond: Passive optical network and multi-access edge computing integration for latency-sensitive applications. *Optical Fiber Technology, 75*, 103191. https://doi.org/10.1016/j.yofte.2022.103191

Etxezarreta, X., Garitano, I., Iturbe, M., & Zurutuza, U. (2023). Software-defined networking approaches for intrusion response in industrial control systems: A survey. *International Journal of Critical Infrastructure Protection, 42*, 100615. https://doi.org/10.1016/j.ijcip.2023.100615

Fakhouri, H. N., Alawadi, S., Awaysheh, F. M., Hani, I. B., Alkhalaileh, M., & Hamad, F. (2023). A comprehensive study on the role of machine learning in 5G security: Challenges, technologies, and solutions. *Electronics, 12*(22), 4604. https://doi.org/10.3390/electronics12224604

Farhoudi, M., Shokrnezhad, M., & Taleb, T. (2023). QoS-aware service prediction and orchestration in cloud-network integrated beyond 5G. In *GLOBECOM 2022 -2022 IEEE Global Communications Conference*. https://doi.org/10.1109/GLOBECOM54140.2023.10436905

Gkatzios, N., Koumaras, H., Fragkos, D., & Koumaras, V. (2024). A proof-of-concept implementation of an AI-assisted user-centric 6G network. In *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)* (pp. 907-912). https://doi.org/10.1109/EuCNC/6GSummit60053.2024.10597020

González, C. C., Pupo, E. F., Floris, A., Porcu, S., Murroni, M., & Atzori, L. (2026). OTT–MNO collaboration for a network-layer ML-based QoE prediction for video streaming over 5G O-RAN. *Computer Networks*, 112152. https://doi.org/10.1016/j.comnet.2026.112152

Haji, S. H., Zeebaree, S. R. M., Saeed, R. H., Ameen, S. Y., Shukur, H. M., Omar, N., Sadeeq, M. A. M., Ageed, Z. S., Ibrahim, I. M., & Yasin, H. M. (2021). Comparison of software defined networking with traditional networking. *Asian Journal of Research in Computer Science*, 1-18. https://doi.org/10.9734/ajrcos/2021/v9i230216

Honarvar, R., Zolghadrasli, A., & Monemi, M. (2022). Context-oriented performance evaluation of network selection algorithms in 5G heterogeneous networks. *Journal of Network and Computer Applications, 202*, 103358. https://doi.org/10.1016/j.jnca.2022.103358

Hoyhtya, M., Lahetkangas, K., Suomalainen, J., Hoppari, M., Kujanpaa, K., Trung Ngo, K., Kippola, T., Heikkila, M., Posti, H., Maki, J., Savunen, T., Hulkkonen, A., & Kokkinen, H. (2018). Critical communications over mobile operators' networks: 5G use cases enabled by licensed spectrum sharing, network slicing and QoS control. *IEEE Access, 6*, 73572-73582. https://doi.org/10.1109/ACCESS.2018.2883787

Hussain, S. M. S., Aftab, M. A., & Ustun, T. S. (2020). Performance analysis of IEC 61850 messages in LTE communication for reactive power management in microgrids. *Energies, 13*(22), 6011.https://doi.org/10.3390/en13226011

Hyder, H. K., & Lung, C.-H. (2018). Closed-loop DDoS mitigation system in software defined networks. In *2018 IEEE Conference on Dependable and Secure Computing (DSC)* (pp. 1-6). https://doi.org/10.1109/DESEC.2018.8625125

Ito, K., Nakazato, J., Fontugne, R., Tsukada, M., & Hiroshi, E. (2025). A multipath redundancy communication framework for enhancing 5G mobile communication quality. *Computer Communications*, 108157. https://doi.org/10.1016/j.comcom.2025.108157

Jiang, W., Han, H., He, M., & Gu, W. (2023). ML-based pre-deployment SDN performance prediction with neural network boosting regression. *Expert Systems with Applications, 241*, 122774. https://doi.org/10.1016/j.eswa.2023.122774

Khairi, M. H. H., Ariffin, S. H. S., Latiff, N. M. A., Yusof, K. M., Hassan, M. K., Al-Dhief, F. T., Hamdan, M., Khan, S., & Hamzah, M. (2021). Detection and classification of conflict flows in SDN using machine learning algorithms. *IEEE Access, 9*, 76024-76037.https://doi.org/10.1109/ACCESS.2021.3081629

Khan, N. A. (2022). 5G network: Techniques to increase quality of service and quality of experience. *International Journal of Computer Networks and Applications, 9*(4), 476. https://doi.org/10.22247/ijcna/2022/214508

Khodakhah, F., Mahmood, A., Österberg, P., & Gidlund, M. (2025). Adaptive user pairing with non-orthogonal medium access choices for balanced coexistence of mission-critical and eMBB services in cellular IoT. *IEEE Open Journal of the Communications Society, 6*, 5414-5433. https://doi.org/10.1109/OJCOMS.2025.3578727

Kunasegran, M. P., Pang, W. L., & Phang, S. K. (2025a). Optimising resource allocation in 5G networks: Balancing URLLC and eMBB traffic under gNB congestion. In *Proceedings of the 2025 Multimedia University Engineering Conference (MECON 2025)* (pp. 1-6). IEEE. https://doi.org/10.1109/MECON67253.2025.11277096

Kunasegran, M. P., Pang, W. L., & Phang, S. K. (2025b). SLA-aware DRL for joint slice admission control and load balancing in multi-service 5G RAN. In *Proceedings of the 9th International Conference on Recent*

Advances and Innovations in Engineering (ICRAIE 2025) (pp. 207-212). IEEE. https://doi.org/10.1109/ICRAIE65839.2025.11239148

Lai, P., He, Q., Cui, G., Xia, X., Abdelrazek, M., Chen, F., Hosking, J., Grundy, J., & Yang, Y. (2019). Edge user allocation with dynamic quality of service. In *Proceedings of the conference* (pp. 86-101). https://doi.org/10.1007/978-3-030-33702-5_8

Louvros, S., Paraskevas, M., & Chrysikos, T. (2023). QoS-aware resource management in 5G and 6G cloud-based architectures with priorities. *Information, 14*(3), 175. https://doi.org/10.3390/info14030175

Ly, A., & Yao, Y. D. (2021). A review of deep learning in 5G research: Channel coding, massive MIMO, multiple access, resource allocation, and network security. *IEEE Open Journal of the Communications Society, 2*, 396-408. https://doi.org/10.1109/OJCOMS.2021.3058353

Mahmood, A., Abedin, S. F., Sauter, T., Gidlund, M., & Landernäs, K. (2021). Factory 5G: A review of industrial-centric features and deployment options. *IEEE Industrial Electronics Magazine, 16*(2), 24-34. https://doi.org/10.36227/techrxiv.17089265.v1

Matoussi, S., Fajjari, I., Aitsaadi, N., & Langar, R. (2020a). User slicing scheme with functional split selection in 5G cloud-RAN. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-8). https://doi.org/10.1109/WCNC45663.2020.9120828

Matoussi, S., Fajjari, I., Aitsaadi, N., & Langar, R. (2023b). User-centric slice allocation scheme in 5G networks and beyond. *IEEE Transactions on Network and Service Management, 20*(4), 4268-4282.https://doi.org/10.1109/TNSM.2023.3284206

Nikolaidis, P., Zoulkarni, A., & Baras, J. (2023). Bandwidth provisioning for network slices with per user QoS guarantees. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium* (pp. 1-9). https://doi.org/10.1109/NOMS56928.2023.10154366

Prodromos, N., Diasakos, D., Kokkinos, V., Gkamas, A., Bouras, C., & Pouyioutas, P. (2024). Dynamic bandwidth allocation in MIMO 5G networks. In *2024 International Wireless Communications and Mobile Computing (IWCMC)* (pp. 97-102). https://doi.org/10.1109/IWCMC61514.2024.10592577

Raeisi, M., & Sesay, A. B. (2023). Handover reduction in 5G high-speed network using ML-assisted user-centric channel allocation. *IEEE Access, 11*, 84113-84133. https://doi.org/10.1109/ACCESS.2023.3297982

Ramesh, P., Mohan, B., Viswanath, L., & Stephen, B. J. (2024). Software defined network architecture based network slicing in fifth generation networks. *Informacije MIDEM- Journal of Microelectronics, Electronic Components and Materials, 51*(3). https://doi.org/10.33180/InfMIDEM2024.205

Rathod, I., & Saxena, S. (2024). Design, simulation and analysis of multi-dimensional multiple access (MDMA) schemes using MATLAB for quality of service (QoS) enhancement. *Journal of Intelligent Systems and Internet of Things, 11*(2), 111-128. https://doi.org/10.54216/JISIoT.110210

Rawshan, F., Hossen, M., & Islam, Md. R. (2024). Dynamic wavelength and bandwidth allocation using service class prioritisation for upstream in 100 Gb/s NG-EPON. *Results in Engineering, 24*, 103151.https://doi.org/10.1016/j.rineng.2024.103151

Shameli, R., & Rajkumar, S. (2026). Design of an AI-driven secure 5G-SDN framework with federated reinforcement learning for anomaly detection, mitigation, and attack forensics. *Frontiers in Artificial Intelligence, 9*, 1701944. https://doi.org/10.3389/frai.2026.1701944

Shrivastava, V. K., Baek, S., & Baek, Y. (2022). 5G evolution for multicast and broadcast services in 3GPP Release 17. *IEEE Communications Standards Magazine, 6*(3), 70-76. https://doi.org/10.1109/MCOMSTD.0001.2100068

Singh, K. D., & Sood, S. K. (2020). QoS-aware optical FoG-assisted cyber-physical system in the 5G-ready heterogeneous network. *Wireless Personal Communications, 116*(4), 3331-3350.https://doi.org/10.1007/s11277-020-07855-5

Sufyan, A., Khan, K. B., Khashan, O. A., Mir, T., & Mir, U. (2023). From 5G to beyond 5G: A comprehensive survey of wireless network evolution, challenges, and promising technologies. *Electronics, 12*(10), 2200. https://doi.org/10.3390/electronics12102200

Tam, P., Ros, S., Song, I., & Kim, S. (2024). QoS-driven slicing management for vehicular communications. *Electronics, 13*(2), 314. https://doi.org/10.3390/electronics13020314

Tanuja, K. S., Shanmukaswamy, C. V., Gurushankar, H. B., & Dinesh, H. A. (2023). Dynamic bandwidth allocation scheme for enhanced performance in 5G point-to-point networks. *ICTACT Journal on Communication Technology, 14*(2), 2945-2951. https://doi.org/10.21917/ijct.2023.0438

Toor, W. T., Basit, A., Maroof, N., Khan, S. A., & Saadi, M. (2019). Evolution of random access process: From legacy networks to 5G and beyond. *Transactions on Emerging Telecommunications Technologies, 33*(6). https://doi.org/10.1002/ett.3776

Trifan, R.-F., Lerbour, R., & Le Helloco, Y. (2015). Mirroring LTE scheduler performance with an adaptive simulation model. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)* (pp. 1-5).https://doi.org/10.1109/VTCSpring.2015.7145931

Umar, M. M., Mohammed, A., & Abdulazeez, A. (2024). Review of QoS-aware resource allocation schemes for 5G networks. *Dutse Journal of Pure and Applied Sciences, 10*(3c), 296-303.https://doi.org/10.4314/dujopas.v10i3c.28

Vidhya, P., K, S., R, S., & S, G. (2025). Dynamic network slicing-based resource management and service-aware virtual network function (VNF) migration in 5G networks. *Computer Networks*, 111064.https://doi.org/10.1016/j.comnet.2025.111064

Wang, H.-C., Chiu, W.-P., & Wu, S.-C. (2015). QoS-driven selection of web service considering group preference. *Computer Networks, 93*, 111-124. https://doi.org/10.1016/j.comnet.2015.10.014

Wu, C., Lu, H., Chen, Y., & Qin, L. (2024). Cross-layer optimisation for statistical QoS provision in C-RAN with finite-length coding. *IEEE Transactions on Communications, 72*(6), 3393-3407.https://doi.org/10.1109/TCOMM.2024.3370817

Xiang, Z., Ying, F., Yan, H., Zheng, Z., Zhang, Y., & Xu, Y. (2025). QoS-effective and resilient service deployment and traffic management in MEC-based crowdsensing. *Symmetry, 17*(5), 718. https://doi.org/10.3390/sym17050718

Xie, M., Gonzalez, A. J., Gronsund, P., Lonsethagen, H., Waldemar, P., Tranoris, C., Denazis, S., & Elmokashfi, A. (2022). Practically deploying multiple vertical services into 5G networks with network slicing. *IEEE Network, 36*(1), 32-39. https://doi.org/10.1109/MNET.001.2100361

Yang, D., & Tsai, W.-T. (2024). SDN-based congestion control and bandwidth allocation scheme in 5G networks. *Sensors, 24*(3), 749. https://doi.org/10.3390/s24030749

Yaqoob, J. I., Pang, W. L., Wong, S. K., & Chan, K. Y. (2020). Enhanced exponential rule scheduling algorithm for real-time traffic in LTE network. *International Journal of Electrical and Computer Engineering (IJECE), 10*(2), 1993-2002. https://doi.org/10.11591/ijece.v10i2.pp1993-2002

Yaqoob, Y. J. I. A., Pang, W. L., Wong, S. K., & Chan, K. Y. (2019). Performance evaluation of video streaming on LTE with coexistence of Wi-Fi signal. *Bulletin of Electrical Engineering and Informatics, 8*(3), 890-897. https://doi.org/10.11591/eei.v8i3.1580

Yin, C., Dong, P., Du, X., Zheng, T., Zhang, H., & Guizani, M. (2020). An adaptive network coding scheme for multipath transmission in cellular-based vehicular networks. *Sensors, 20*(20), 5902. https://doi.org/10.3390/s20205902

Zeyad, I., & Al Janaby, A. O. (2025). Interference mitigation for dynamic user connectivity using SDN and radio resource management in cell-less networks. *Journal of Engineering Science and Technology, 20*(2), 520-535.

Pertanika J. Sci. & Technol. 34 (2): 951 - 973 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Improved Speech Enhancement using Parallel MVDR Beamforming and Coherent-to-Diffuse Power Ratio Post-filtering

Sureshkumar Natarajan[1], Mohd Khair Hassan[2], Raja Kamil[2], Faisul Arif Ahmad[1], Syaril Azrad[3], June Francis Macleans[4], Hussna Elnoor M. Abdalla[1], Nurbek Saparkhojayev[5], and Syed Abdul Rahman Al-Haddad[1]*

[1]Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Department of Electrical and Electronic Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[3]Department of Aerospace Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[4]Independent Researcher, 50250 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia
[5]Chairman of the Board-Rector of the NCJSC, Rudny Industrial University, 111500 Rudny City, Kazakhstan

ABSTRACT

Speech communication involves the exchange of information between two or more individuals; however, background noise and reverberation often degrade speech clarity and intelligibility. The impact of these degradations depends on factors such as the number, intensity, and spatial characteristics of noise sources, as well as reflections from surrounding surfaces. These effects pose significant challenges for applications including teleconferencing, hearing-aid devices, and human-machine voice interfaces. This study aims to address the combined effects of background noise and reverberation by proposing a speech enhancement framework that integrates Minimum Variance Distortionless Response (MVDR) beamforming with a Coherent-to-Diffuse Power Ratio (CDR) based post-filter. The methodology relies on parallel processing, where MVDR beamforming performs spatial noise suppression, while CDR values are estimated from microphone-domain signals and used to compute post-filter gains for suppressing residual diffuse noise and reverberation in the beamformer output. The proposed algorithm was evaluated over an input signal-to-noise ratio range from 0 dB to 40 dB and compared against the Integrated Sidelobe Cancellation Linear Prediction (ISCLP) baseline using four objective metrics: Perceptual Evaluation of Speech Quality (PESQ), Extended Short-Time Objective Intelligibility (ESTOI),

ARTICLE INFO

Article history:
Received: 04 August 2025
Accepted: 04 February 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.15

E-mail addresses:
suresh45619@gmail.com (Sureshkumar Natarajan)
khair@upm.edu.my (Mohd Khair Hassan)
kamil@upm.edu.my (Raja Kamil)
faisul@upm.edu.my (Faisul Arif Ahmad)
syaril@upm.edu.my (Syaril Azrad)
macleans30june@gmail.com (June Francis Macleans)
hussnaa@hotmail.com (Hussna Elnoor M. Abdalla)
nursp81@gmail.com (Nurbek Saparkhojayev)
sar@upm.edu.my (Syed Abdul Rahman Al-Haddad)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Cepstral Distance (CD), and Weighted Spectral Slope Distance (WSS). The results demonstrate consistent improvements in PESQ and reductions in CD and WSS across all tested conditions, while gains in ESTOI remain modest. These findings indicate improved spectral fidelity and speech naturalness, highlighting the practical relevance of the proposed framework for real-time, low-complexity speech enhancement in noise and reverberation-prone communication systems.

Keywords: Acoustics, CDR, dereverberation, MVDR beamformer, noise suppression, speech enhancement

INTRODUCTION

Various speech-based application devices, such as voice-controlled robotics and day-to-day use of hands-free mobiles, process the signal captured by the microphone. Furthermore, the speech recognition and speaker recognition systems also depend on the quality of the microphone recorded signal, where the speech signal is often contaminated with both background noise and reverberant signal, which cannot be directly used for speech application devices. Hence, the requirement of a speech enhancement algorithm is crucial and essential, which filters the unwanted component and thus aims to enhance the quality and clarity of the speech. Over the years, there has been tremendous advancement in the development of various speech enhancement algorithms, right from the traditional speech enhancement, such as spectral subtraction and Wiener filter, to modern deep learning-based speech enhancement algorithms, which are driven by data (O'Shaughnessy, 2024). Despite these advancements, the same clarity of signal uttered by the source speaker will not reach the receiving end of the speaker due to the surrounding background noise and reverberation, which distorts the source signal (Beutelmann & Brand, 2006). To tackle these issues and to strengthen the desired speaker utterance, multi-microphone techniques using spatial filtering effects were extensively investigated (Delcroix et al., 2015; Habets & Benesty, 2013). Another technique based on a recursive Kalman filtering framework was proposed by Cohen et al. (2021), specifically designed to deal with both reverberation effects as well as interfering sources. The two most fundamental ideas used in modern multi-microphone speech enhancement systems are the spatial filtering concept and deconvolution.

The spatial filtering was usually realised using multiple microphones, which is called the beamforming technique, which is basically used to enhance the desired speech arriving from a particular direction and suppress the interfering sources arriving from other unknown directions. Beamforming, in general, is more successful in removing directional sources arriving from outside the region of interest as compared to reverberated signals. Among various types of beamforming techniques, the MVDR beamformer is known to be a widely used technique as it offers better improvement in speech intelligibility and clarity in a complex noise environment (Cauchi et al., 2015). Some other techniques which integrate a linear prediction framework with a beamforming method have also shown improvement in obtaining signal clarity within reverberant conditions (Yang et al., 2022).

There have been some efforts taken to integrate noise reduction with dereverberation, which has led to the advancement of hybrid techniques. For example, combining pre-whitening and Generalised Sidelobe Canceller (GSC) beamformer helps to enhance the desired speech by removing reverberation (Dietzen et al., 2015). Similarly, fusion based technique was employed to deal with both noise suppression and reverberation, which uses Multichannel Linear Prediction (MCLP) and GSC within a Kalman filtering framework (Dietzen et al., 2018). Another technique that uses a cascade approach that employs a Multiple-Input Multiple-Output (MIMO) based Weighted Prediction Error (WPE) algorithm is used as a preprocessing stage, followed by Minimum Power Distortion Less Response (MPDR) beamforming to simultaneously tackle both reverberation and background noise (Delcroix et al., 2015).

The recent studies have indicated that further enhancement of the signal can be obtained as compared to the cascaded approach. A study has shown Weighted Power Minimisation Distortion Less Response (WPD) beamformer, which jointly optimises the convolutional WPE filter and MPDR beamformer, yielded in superior performance (Nakatani & Kinoshita, 2019a). Extension to this model was carried out by including the application of sparse priors to further refine the WPD beamformer's performance (Gode et al., 2021). To obtain effective enhancement of the signal, an adaptive version has been introduced which uses RLS-WPD and the recursive WPD approach to simultaneously suppress both background noise and reverberation (Nakatani & Kinoshita, 2019b). One more interesting approach was introduced in recent times called the Weighted Binaural Linearly Constrained Minimum Power (wBLCMP) beamformer, which has shown the ability to jointly handle background noise, dereverberation and other interfering sources while maintaining binaural cues. This enhancement was achieved by extending the WPD formulation by combining the optimisation of the convolutional WPE filter with the Linearly Constrained Minimum Power (LCMP) beamformer (Gode & Doclo, 2023).

Although the proposed work is based on traditional non-deep learning (DL) based methods, in recent years, there has been a lot of work-based DL-based algorithms that have shown tremendous advancement in enhancing the desired speaker speech. For instance, an attention-based MVDR beamforming architecture (ABIC-MVDR) was introduced by Bai et al. (2025) that integrates spatial covariance matrix estimation with In-Place Gated Convolutional Recurrent Networks and frequency-independent LSTM modules. They employed an end-to-end design that yields a significant improvement in challenging conditions, including the movement of the speaker and varying acoustic environments. In another study, the MVDR beamforming technique was employed within a U-Net architecture to obtain multichannel speech in severe noise and a reverberant environment (Lee et al., 2024). A deep beamforming framework proposed to jointly enhance speech signals and localise target speakers by learning spatial filtering directly from multichannel data, which uses a convolutional recurrent network for training (CRN) (Chang et al., 2024).

A lightweight two-stage CRN was proposed for the hearing aid application that uses a microphone array for enhancing the speech signal (Xi et al., 2024). A WPD beamforming approach was proposed to jointly perform dereverberation and denoising, where the unified filter is estimated using a deep neural network (Fujita et al., 2024). A detailed systematic review that uses 187 studies highlighted the impact of various deep learning-based techniques, such as CNN, LSTM, and many more, for developing a speech enhancement algorithm to improve the quality of the speech signal, and the study also discussed speech enhancements based on beamforming techniques (Natarajan et al., 2025). In another study, which specifically discussed the impact of the beamforming technique related to enhancement in the study, this review mainly discussed the comparison of various techniques with the beamforming approach, which has shown an effective method in enhancing the clarity of speech in a far-field microphone captured signal (Natarajan et al., 2023).

Furthermore, Huang et al. (2025) have conducted a comprehensive review based on microphone array processing. The author has shown tracing evolution from the traditional beamforming approach to the latest advancement of the spatial filter using the DL approach. Although a notable performance gain through the use neural-based beamforming technique can be found to be better as compared to a traditional statistical-based approach. However, DL methods often rely on a huge amount of data, which eventually results in increasing the computational complexity, and they are less interpretable. Due to this fact, it is more often preferred to use a non-DL method as it is easier to carry out interpretation and analysis, which helps to address the issues. Due to these practical considerations, a traditional-based approach was employed in this work.

The CDR is one of the well-known techniques capable of suppressing the reverberant components (Schwarz & Kellermann, 2015). Using the idea of the CDR technique, an algorithm was developed recently that uses CDR as a preprocessing and the CDR output is fed to a recursive least squares (RLS) adaptive filter to remove noise without depending on an explicit noise reference signal (Natarajan et al., 2024). However, it is a known fact that CDR alone cannot be effective when recording is carried out in an environment where multiple speakers talk simultaneously. To handle multiple speakers talking scenarios, the MVDR beamforming method serves as one of the effective techniques in enhancing the desired speaker's speech by isolating the interfering speaker signal and helps to mitigate some reverberant components.

The idea is to take advantage of using both techniques by proposing a hybrid approach that will help to suppress the remaining residual reverberant component that exists at the output of the MVDR beamformer. In this work, a noise-corrupted signal will be processed by both the MVDR beamformer as well as CDR estimation block in parallel, and later the CDR output will be used as SNR value which in turn calculates the gain values of the

CDR is later used as CDR gain as a post filter to obtain the final output where the MVDR beamformer output is multiplied with the post-filtered gain value. In this proposed work, the aim is to suppress the babble noise and reverberation effectively as compared to the baseline ISCLP algorithm, using the same experimental conditions (Dietzen et al., 2020).

RELATED WORK

Speech enhancement algorithms play a crucial role in enhancing the quality and clarity of speech by reducing background noise in a challenging acoustic setting (Shankar et al., 2021). Single-channel techniques are the simplest and easiest to use in the system, as they depend solely on the spectral and temporal characteristics of a corrupted signal to estimate clean speech. However, their performance was limited because they do not employ spatial information, and their performance will be affected severely, specifically in a complex acoustic setting where the impact of residual noise and speech distortion is more significant (Shankar et al., 2021). To tackle these issues, the Modified Noise Reduction Method (MNRM) was proposed by Kumar and Chitra (2022), which was built upon the traditional Wiener filtering framework. This method succeeds in enhancing the speech by suppressing the noise where Wiener gain was calculated before and after SNR-based filtering, and it incorporates a decision-directed strategy that uses a twin-stage suppression mechanism approach. This approach has shown more resilience in practical applications. However, the performance of the MNRM method still depends on estimating the value of power spectral densities (PSDs) of clean speech and noise, which is one of the biggest challenges to accurately determine the PSD values under real-time and non-stationary scenarios. In a related work, the Wiener gain was determined by considering both noise and reverberation, and the gain was obtained using estimates of both SNR and CDR (Xiang et al., 2025).

In contrast, the use of microphone array processing was preferred over single-channel as it uses multiple microphones to capture spatial diversity and it offers better noise removal capability to yield higher performance over single-channel-based techniques, specifically when sources are arriving from different directions, and also it helps in handling diffuse noise sources or non-stationary signals (Song et al., 2021). Multi-channel processing allows for the spatial filtering, which ensures the isolation of the desired speaker's speech from non-desired speech. However, it is difficult to deal with when the sources are moving simultaneously because the impact of overlapping spectral content and frequent directional changes in real-time will be greater, which will reduce the performance of spatial filtering. Traditional algorithms such as multichannel Wiener and Kalman filters have been adopted for use in the multichannel domain, but they cannot produce distorted output speech signals when the estimation of noise power is inaccurate (Dietzen et al., 2020). Recently, Thimmraja et al. (2022) extended the spectral subtraction (SS) algorithm approach by combining phase information between the noisy speech and the noise model. This modified approach has

shown some improvement in cross-term estimation in the gain function, which in turn helps to align these modified to perform more closely to that of minimum mean square error (MMSE)-based filtering, and the results of this method have shown better musical noise suppression and better performance despite adverse noise conditions, specifically at low SNRs.

Research using an adaptive beamforming approach has shown significant advancement due to the requirement of robust performance in reverberant or uncertain acoustic settings. Zhao et al. (2020) focused on adaptive strategies to address the challenges, such as steering vector mismatches, statistical estimation errors, and late reverberation. To tackle the steering vector model uncertainties, the norm-constrained optimisation technique was introduced first, which results in improving the robustness against directional mismatches and multipath effects. Spatial smoothing was employed as the second step to stabilise the estimation of the noise covariance matrix, which results in reducing the diffuse interference. This method has shown notable performance improvement in terms of PESQ and STOI metrics as compared to the standard beamforming approach (Zhao et al., 2020).

In the last few years, there have been significant advancements in DL-based multichannel speech enhancement approaches as they learn the complex pattern of spatial details and help to determine the spectral relationship from the data. Some of the neural network-based approaches, such as the Gated Convolutional Recurrent Network (GCRN) proposed by Tan and Wang (2019), and the Deep Path Convolutional Recurrent Network (DPCRN) introduced by Le et al. (2021), which focused mainly on spectral mapping by considering each microphone channel as an independent input signal. Although these approaches have proved to be effective, especially for spectral modelling, they often neglect the important spatial dependencies between the microphone channels. To tackle these issues, Liu and Zhang (2021) developed the In-Place Gated Convolutional Recurrent Neural Network (IGCRN) that helps to retain the spatial information at every frequency bin without performing either down-sampling or up-sampling. Later, Tan et al. (2022) introduced a neural spectro-spatial filtering approach using a multi-microphone system that adopts a learning approach to learn both spectral and spatial filters within a convolutional network, which has shown improvement in the performance gain of the speech enhancement task.

To attain better speech enhancement performance, more advanced DL architectures such as FasNet (Luo et al., 2019), EabNet (Li et al., 2022), and MIMO-Unet (Ren et al., 2021) are explicitly integrated with traditional beamforming methods. This integration ensures the efficient use of spatial diversity in noisy acoustic settings. Despite such progress, the existing DL-based systems still process multichannel input by concatenating Short-Time Fourier Transform (STFT) features from each microphone channel. Furthermore, these approaches often fail to consider two important spatial clues, namely inter-channel phase information and spatial coherence, which are essential for obtaining accurate

source localisation, and they help to isolate the target speech from the background noise in a reverberant environment. Moreover, extracting the spatial features using a DL-based model is computationally expensive, and it may introduce latency, resulting in difficulty in employing it in a real-time system. Another issue with the DL model is related to data, because the performance of the DL model mostly depends on a huge number of labelled datasets required for effective training of the model. More importantly, the dependency of labelled data used for training may limit the performance of the model when a model is fed with new unseen data.

Non-neural network or analytical approaches have received more attention due to the fact they do not need large datasets for operation, and it is less computational cost and more importantly it is possible to perform the interpretation and signal analysis at any stage using an analytical approach whereas in case of DL model, interpretation of the signal is difficult as the model itself acts like a black-box. One such notable speech enhancement filtering technique is the ISCLP algorithm (Dietzen et al., 2020) that integrates MCLP with GSC in a Kalman filter framework. This design allows for performing spatial and temporal filtering operations simultaneously, and this approach has resulted in significant performance as compared to traditional sequential processing pipelines, specifically in suppressing the diffuse reverberation and interfering sources in distant talking acoustic settings. However, its practical implementation may face some constraints and difficulties because the accuracy of this method heavily depends on the precise estimation of Relative Early Transfer Functions (RETFs). If the RETFs are not estimated accurately, then in such scenarios, self-cancellation artefact issues can occur, which result in disrupting the filter updates, which can be seen during pauses in speech. Furthermore, the MCLP filters depend on active speech segments to learn active filter coefficients, making it difficult to jointly optimise with other models in real-time. Another drawback of the ISCLP algorithm is computationally complex, which puts a limitation on real-time implementation.

To overcome these issues, the proposed method uses the MVDR beamformer and CDR to implement a hybrid approach. The role of the MVDR beamformer in this work is to preserve the desired speech by suppressing the directional interferences. However, the performance of the MVDR beamformer tends to degrade in a scenario if the level of reverberation is high, where the spatial selectivity reduces due to early reflection effects. To compensate for the issue of reverberation, the CDR estimation technique is incorporated in this work, which is well-known for dereverberation. However, the CDR technique cannot handle any background noise or interfering sources. So, the proposed method considers the advantage of both methods, and in this research work, the MVDR beamformer and the CDR technique were used to implement a hybrid approach where both the MVDR beamformer and the CDR process the noise-corrupted signal. Since the output of the MVDR beamformer does contain some residual noise, which will be further reduced by

employing CDR as a post-filter. The integration of the CDR technique with the MVDR beamformer was to improve the desired speech clarity and suppress the reverberation and background noise in acoustic settings where a single dominant speech source is present.

To evaluate the effectiveness of the proposed method and to compare its performance with the baseline ISCLP algorithm, we used the same ISCLP experimental settings. The experimental setup mirrors the ISCLP evaluation environment, featuring a single target speaker within 2 meters of a microphone array, moderate reverberation (RT60 = 0.6 seconds), and multi-talker babble interference. The objective is to assess whether the MVDR-CDR-SS framework can achieve comparable or improved performance relative to ISCLP, thereby establishing its viability as a lightweight and competitive alternative for real-time, far-field speech enhancement.

SIGNAL MODELLING

In this section, the speech enhancement system is discussed. Consider the microphone recorded signals consist of mixed components of desired speech and noise/interfering sources in a reverberant environment. The desired speech gets convolved with the room impulse response (RIR), and these signals can be mathematically expressed in the STFT domain at frame m and frequency f, and the i^{th} microphone signal can be written using Equation 1 and Equation 2:

$$X_i(m,f) = o_i(m,f)S_i\,(m,f) + N_i(m,f) \quad [1]$$

$$X(m,f) = [X_0(m,f), X_1(m,f), \ldots\ldots X_{N-1}(m,f)]^T \quad [2]$$

where, $X_i(m,f)$ represents the direct signal component $S_i(m,f)$, where $o_i(m,f)$ represents the steering vector of the sound coming from the target speaker and $N_i(m,f)$ representing noise and late reverberation signals.

Minimum Variance Distortionless Response Beamformer

To process a mixed signal containing the target speaker's speech and background noise, the MVDR beamformer is used. The mixed-noise signal is first transformed into shorter segments using STFT for the Fourier transformation. The MVDR filter weights are then computed to minimise the power of the noise component while maintaining the signal power of the target speaker. This Equation 3 can be expressed as:

$$W_{MVDR}(m,f) = argmin\, w^H(m,f)\, N_i(m,f) \quad [3]$$

Based on the received multichannel covariance matrix of the noise component $N_i(m, f)$ From the microphone array, we calculate the steering vector for both the target speaker and the noise signal using the eigenvector method suggested in (Sarradj, 2010). To estimate the covariance matrix of the target signal, we need to compute the covariance matrix of the combined signal and the noise signal. Once we have the target signal's covariance matrix, we can compute the steering vector using the formulas and concepts derived. The eigenvalue decomposition of the microphone array's covariance matrix for the target speaker can be expressed using Equation 4:

$$G = \mathbb{V}\Lambda\mathbb{V}^H \quad [4]$$

where G, is the covariance matrix for the target speaker in the microphone array, and it can be written in Equation 5. Λ is a diagonal matrix that consists of the positive eigenvalues of G can be mathematically written using Equation 6, while $\mathbb{V}$ is the eigenvector corresponding to these eigenvalues. We assume that the smallest eigenvalue associated with the noise component of the signal is n^2, following the method proposed in (Su & Morf, 1983). Therefore, the eigenvalue can be split into two parts in Equation 5, one for the signal and the other for the noise. The steering vector is estimated using eigenvalue decomposition of the spatial covariance matrix, following the approach proposed by Su and Morf (1983). This method assumes sufficient separation between the signal and noise subspaces. In low-SNR or highly diffuse reverberant environments, partial overlap between these subspaces may occur, potentially affecting estimation accuracy. While this limitation is inherent to subspace-based methods, the impact is mitigated in the proposed framework through subsequent diffuseness-based post-filtering.

$$G = \mathbb{V}\begin{bmatrix}\Lambda_s & 0\\ 0 & 0\end{bmatrix}\mathbb{V}^H + n^2\mathbb{V}\mathbb{V}^H = \mathbb{V}_s\Lambda_s\mathbb{V}^H{}_s + n^2\mathrm{I} \quad [5]$$

$$\Lambda = \begin{bmatrix}\Lambda_s + n^2 I & 0\\ 0 & n^2 I\end{bmatrix} \quad [6]$$

Each part corresponds to either the signal or the noise, where the eigenvectors of $\mathbb{V}_s$ belong to the signal subspace $\mathbb{V}$, while the remaining ones belong to the noise subspace. In real experiments, the eigenvalues in the noise subspace have minimal impact on the eigenvalues in the signal subspace. We calculate the smallest eigenvalue for the noise signal in the matrix and identify the highest eigenvalues as the target source values in the matrix. Next, we assign high values to frequency bin values associated with the target speaker and low values to those associated with noise. The weights are first derived from the mixed input signal, and each frequency bin is then assigned its own corresponding weight. These

weights are applied back to the mixed signal so that every frequency bin is scaled according to its computed value. In this way, the enhancement is distributed more precisely across the frequency range, ensuring balanced improvement throughout the spectrum. Finally, we obtain a time-domain signal by converting the beamforming signal to an inverse short-time Fourier transform, following the weight calculation with the signal.

Coherent-to-Diffuse Power Ratio

To enhance the results of the MVDR beamformer, we employ a postfilter based on the spectral subtraction technique to remove diffuse noise and reverberation. We start by determining a short-time SNR value, which is used in the spectral subtraction method to calculate the filter gain. Furthermore, the estimation of the CDR only necessitates the computation of the SNR, which in turn determines the expression of the gain $G(m,f)$ which can be written using Equation 7:

$$G(m,f) = max\left\{1 - \mu \frac{1}{1+SNR(m,f)}, G_{min}\right\} \quad [7]$$

The gain floor has a predetermined value of 0.1 and an overestimation factor μ is set to 1.3 (Schwarz & Kellermann, 2015). The SNR value is derived from the CDR, and the gain floor G_{min} has a predetermined value of 0.1 while μ is set to 1.3. A mathematical definition for the CDR between two omnidirectional microphones is suggested in (Schwarz & Kellermann, 2015) that can be written using Equation 8:

$$CDR(m,f) = \frac{\Gamma_n(m,f) - \Gamma_x(m,f)}{\Gamma_x(m,f) - \Gamma_s(m,f)} \quad [8]$$

To estimate the correlation between two omnidirectional microphones, we compute the coherence function of the signal recorded by the microphone $\Gamma_x\ (m,f)$ and the diffuse noise $\Gamma_n(m,f)$. We also use short-time spatial coherence to estimate the coherence function of the diffuse field. The spatial coherence function of the diffuse field can be represented using Equation 9:-

$$\Gamma_{diffuse}\ (f) = \Gamma_n(m,f) = \frac{sin\left(2\pi f \frac{d_{1,2}}{c}\right)}{\left(2\pi f \frac{d_{1,2}}{c}\right)} \quad [9]$$

The coherence function of the desired signal component is defined in Equation 10:-

$$\Gamma_s(f) = e^{jkdsin\ (\theta)} = e^{j\,2\pi f\,\Delta t} \quad [10]$$

where, Δt represents the time difference of arrival (TDOA). Using an automatic cross-power spectrum estimator that can be estimated from the microphone signal through recursive averaging to compute the post-filter gain function, and it can be determined using Equation 11:

$$\widehat{\Phi}_{x_p x_q}(m,f) = \lambda \widehat{\Phi}_{x_1 x_2}(m-1,f) + (1-\lambda) X_1(m,f) X^*{}_2(m,f) \quad [11]$$

The microphone signals are processed with a smoothing factor λ, which is a constant value between 0 and 1. We used λ = 0.68 in the experiment. The temporal smoothing factor λ was fixed to 0.68 for all experiments, consistent with prior work on CDR-based dereverberation (Schwarz & Kellermann, 2015). Larger values of λ improve estimation stability but reduce responsiveness to rapid acoustic changes, whereas smaller values enhance temporal tracking at the risk of increased estimation variance and artefacts. In this study, λ was kept constant to ensure stable behaviour and fair comparison with the baseline. The coherence function can be determined by using Equation 12:

$$\hat{\Gamma}_x(m,f) = \frac{\widehat{\Phi}_{x_1 x_2}(m,f)}{\sqrt{\widehat{\Phi}_{x_1 x_1}(m,f)\widehat{\Phi}_{x_2 x_2}(m,f)}} \quad [12]$$

A direction-of-arrival (DOA) independent CDR estimator was employed to reduce computational complexity and improve robustness under diffuse reverberation. While this approach is effective in late-reverberant conditions, it may limit dereverberation accuracy when early reflections dominate. Incorporating DOA-aware diffuseness estimation represents a promising direction for future improvement. In this work, we use the CDR estimator represented in Equation 13 due to its independence from the DOA (Schwarz & Kellermann, 2015).

$$\widehat{CDR}_{DOAindep} = max\left(0, \frac{\Gamma_n Re\{\widehat{\Gamma_x}\} - |\widehat{\Gamma_x}|^2 - \sqrt{\Gamma_n{}^2 Re\{\widehat{\Gamma_x}\}^2 - \Gamma_n{}^2 |\widehat{\Gamma_x}|^2 + \Gamma_n{}^2 - 2\Gamma_n Re\{\widehat{\Gamma_x}\} + |\widehat{\Gamma_x}|^2}}{|\widehat{\Gamma_x}|^2 - 1}\right) \quad [13]$$

We use the real part (Re{.}), magnitude (|.|), and phase (arg{.}) operations along with the maximum operation to avoid negative values. The CDR estimator was initially designed only for microphones (Schwarz & Kellermann, 2015). However, to ensure optimal performance, we need to estimate the CDR for more than just microphones. To achieve this, we estimate the CDR for each pair of microphones, determining the diffuse value from each estimate. Then, we take the arithmetic mean of all the diffuse values of the microphone pairs to obtain the average diffuse estimate.

Instead of the average CDR for a specific pair of microphones, we will directly calculate the average diffusion estimate. This is because CDR can take any value from 0 to infinity, while diffusion values lie within a specific interval. To obtain the average diffuseness value (Schwarz & Kellermann, 2015), we apply the following Equation 14:

$$D(m,f) = (1 + \widehat{CDR}\ (m,f))^{-1} \quad [14]$$

where $\widehat{CDR}\ (m,f)$ represents the estimated CDR value. After obtaining the diffuseness value, the CDR value for each microphone can be computed using Equation 15:

$$\widehat{CDR}_{In}(m,f) = \frac{1-D(m,f)}{D(m,f)} \quad [15]$$

where $\widehat{CDR}_{In}(m,f)$ is the CDR value obtained from the microphone input, not the output of the beamformer. To account for this, a correction factor $A_\Gamma(m,f)$ is applied to $\widehat{CDR}_{In}(m,f)$. The CDR estimate from the beamformer output, $\widehat{CDR}_B(m,f)$ can be defined using Equation 16:

$$\widehat{CDR}_B(m,f) = \frac{\widehat{CDR}_{In}(m,f)}{A_\Gamma(m,f)} \quad [16]$$

The correction factor $A_\Gamma(m,f)$ can be calculated using Equation 17,

$$A_\Gamma(m,f) = w^H(m,f)\, J_{diff}\ (f)\, w(m,f) \quad [17]$$

where $w^H(m,f)$ represents the beamformer weight and $J_{diff}\ (f)$ is the $N \times N$ spatial coherence matrix of a diffuse noise field with the $(1,2)th$ element given by Equation 9. Figure 1 shows the flow process starting from the data acquired by microphones, which consists of a mixed version of target speaker speech, interfering speaker speech and diffuse noise. The corrupted signal is converted into the STFT domain using an STFT block and then applied to coherence estimation to calculate the CDR estimation value without using DOA. Also, the STFT signal is applied to the MVDR beamformer and the post filter gain, which is calculated using the spectral subtraction technique, which uses the SNR value that is obtained from the CDR estimator. Finally, the time domain output is determined by applying the inverse STFT to the spectral enhancement output.

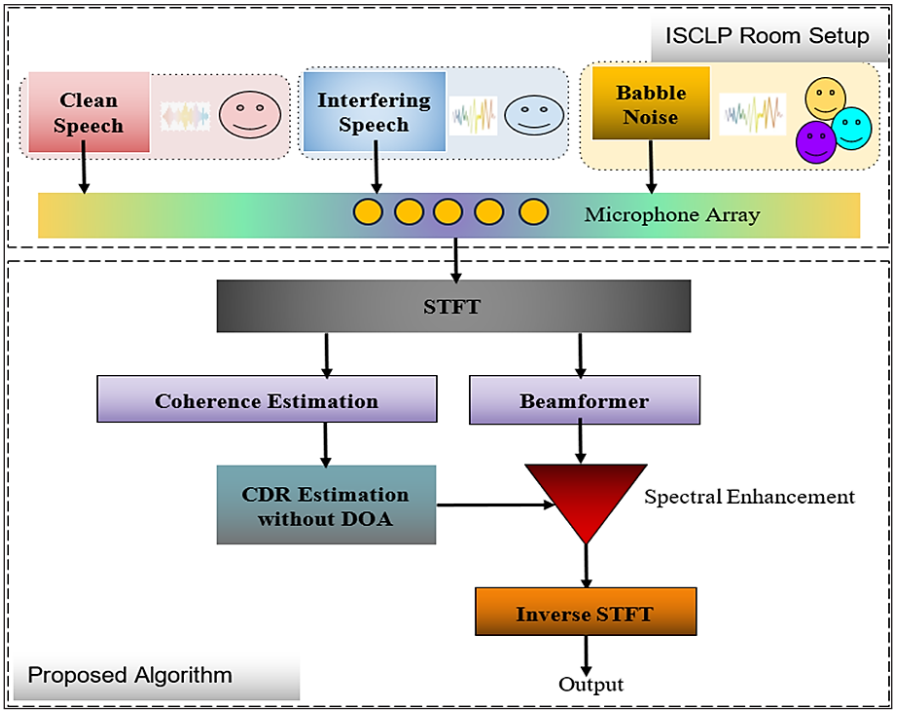


Figure 1. Diagram of proposed algorithm integrates CDR method with MVDR beamformer

Combination of MVDR Beamformer and CDR as Post-filter

The proposed framework adopts a parallel processing structure in which MVDR beamforming and CDR-based post-filtering operate independently. This design decouples spatial noise suppression from diffuseness estimation, as the CDR is computed directly from microphone-domain signals rather than from the beamformer output. Such an approach avoids bias and spatial colouration introduced by beamforming, limits error propagation between processing stages, and preserves the spatial coherence characteristics of the acoustic scene, leading to improved robustness in reverberant and diffuse noise conditions.

To obtain the final output $Z(m,f)$ of the speech enhancement approach, we multiply the weights of the MVDR beamformer $Y_B(m,f)$ signal by the post-filter gain $G(m,f)$, which is calculated using the CDR output of the microphone and output $Z(m,f)$ can be defined using Equation 18:

$$Z(m,f) = G(m,f)\, Y_B(m,f) \quad [18]$$

Finally, the signal is converted back to a time-domain waveform through the utilisation of the inverse short-time Fourier transform.

Experiment Setup and Data Collection

To allow a fair comparison of the results, we decided to use the same data generated in (Dietzen et al., 2020), where two methods were proposed, namely ISCLP SNR and ISCLP smooth SNR. Therefore, we considered the results from (Dietzen et al., 2020) as benchmark results to compare the results with our proposed algorithm. The experiment considers a target speaker, an interfering speaker, and noise in a reverberant environment. We denote the target reverberant speech, the interfering reverberant speech, and the babble noise component as $x1(t), x2(t)$, and v(t), respectively. To generate $x1(t)$ and $x2(t)$, using RIRs with a linear microphone array of M = 5 microphones, each 8 cm apart. We measured RIRs in a room with a reverberation time of 0.61s, with the source 2m away from the array. The target source and the interfering speaker were positioned at 2m from the microphone, with the target speaker placed at an angle of 0 degrees. Recorded babble noise at random locations 2m from the microphone. Both female and male speech source signals are considered in the experiment, and the babble noise is generated by a specific process. The simulations had a sampling frequency of fs = 16kHz, and square Hann windows with STFT = 512 samples and 50% overlap were used in STFT analysis and synthesis. All experiments were conducted offline on a MacBook Pro (14-inch, 2021) equipped with an Apple M1 Pro processor, 16 GB RAM, and a 10-core CPU (8 performance and 2 efficiency cores). The implementation was carried out in Python (version 3.9.12) using the *Pyroomacoustics* toolbox for acoustic simulation and signal processing.

Evaluation Metrics

To determine the effectiveness and to compare the results of the proposed method with the baseline work results, we consider four evaluation metrics, namely: PESQ (Rix et al., 2001), ESTOI, which calculates speech intelligibility without assuming mutual independence between frequency bands, which sets it apart from STOI's correlation analysis (Kobayashi & Kondo, 2019). Beyond these, other important evaluation measures include WSS and CD (Loizou, 2007). The PESQ metric is used to assess the perceived quality of the enhanced speech, with scores ranging from 1, representing very poor quality, to 4.5, representing excellent clarity. ESTOI produces values between 0 and 1, where scores closer to 1 reflect higher speech intelligibility and clarity. The CD was selected as an additional quality measure because it is computationally efficient and effectively captures spectral distortions in the enhanced signal. A higher CD value indicates greater deviation from the clean reference, while a lower value signifies minimal distortion and better signal fidelity.

RESULTS AND DISCUSSION

The proposed method results were compared with the baseline method results of the ISCLP SNR method, the ISCLP Smooth SNR method, which is illustrated using two evaluation

metrics, such as PESQ and ESTOI scores, as shown in Table 1. The performance of the proposed approach is evaluated against both baseline methods across various input SNR levels, ranging from 0 dB to 40 dB. As shown in Table 1, the proposed algorithm demonstrates slightly better performance than the baselines in terms of PESQ, indicating improved overall speech quality. Additionally, the ESTOI results of the proposed algorithm are slightly better than the baseline results at input SNR = 0dB, 5dB, and 40dB, with slightly lower results for other SNR values compared to the baseline result.

Table 2 presents the results of the ISCLP SNR method, the ISCLP Smooth SNR method, and the proposed method using CD and WSS scores. The performance of the proposed algorithm is compared to the baseline results of the ISCLP SNR method and the ISCLP smooth SNR method. The proposed algorithm was tested for different input SNR levels ranging from -20dB to 40dB. It can be observed from Table 2 that the performance of the proposed algorithm achieves the lowest value of CD for all input SNR levels ranging from 0dB to 40dB. The CD results indicate less distortion in the output compared to the two baseline methods. In addition, the WSS also attains the lowest value compared to the baseline results, indicating that the proposed method outperforms the baseline methods. From Table 1 and Table 2, it is clear that the proposed method performs better than the two baseline methods, especially for the metrics PESQ, CD, and WSS, across all input SNR levtheb), (c), and (d) display the time domain waveforms of the target speech, interfering speech, babble noise, and mixed speech, respectively, affected by an input SNR of 0dB and in a reverberant environment with rt60 = 0.61. The time domain waveforms of the output of the ISCLP SNR method, output of the ISCLP smooth SNR method, and output of the proposed method are shown in Figures 2e, 2f, and 2g, respectively. It is evident from Figure 2g compared to Figure 2e and 2f that the amplified speech signal is clearly seen at the output of the proposed method when compared to the results of the two baseline methods.

Table 1
Presents a comparison of the results obtained from the ISCLP SNR method, ISCLP smooth SNR method, and the proposed method using PESQ and ESTOI scores

SNR Level (dB)	ISCLP SNR PESQ	ISCLP Smooth PESQ	Proposed Method PESQ	ISCLP SNR ESTOI	ISCLP Smooth SNR ESTOI	Proposed Method ESTOI
0	1.0918	1.088	1.1192	0.4356	0.4389	0.5008
5	1.1574	1.1514	1.1785	0.5539	0.5535	0.5628
10	1.2166	1.2072	1.2199	0.6159	0.6199	0.6075
15	1.2444	1.2422	1.2601	0.648	0.6519	0.6334
20	1.2561	1.2562	1.2875	0.6614	0.6647	0.6484
25	1.2653	1.2632	1.2993	0.6666	0.6705	0.6574
30	1.2722	1.2726	1.3204	0.6686	0.6724	0.6643
35	1.2765	1.2754	1.3299	0.6674	0.6719	0.6689
40	1.2757	1.2762	1.3364	0.6663	0.6705	0.6719

Table 2
Presents a comparison of the results obtained from the ISCLP SNR method, ISCLP smooth SNR method, and the proposed method using CD and WSS scores

SNR Level (dB)	ISCLP SNR CD	ISCLP smooth SNR CD	Proposed method CD	ISCLP SNR WSS	ISCLP smooth SNR WSS	Proposed method WSS
0	6.3878	6.2441	5.7246	91.5309	85.5473	59.8161
5	5.6725	5.5536	5.2302	72.4086	67.8177	56.4334
10	5.2501	5.1336	4.9275	64.9385	60.8119	54.0634
15	5.0212	4.9056	4.7879	61.0293	56.8611	52.6873
20	4.8965	4.7925	4.7254	59.0557	55.2304	51.82
25	4.8181	4.7317	4.7008	57.6582	54.1183	51.0073
30	4.7847	4.6989	4.6887	57.0284	53.7325	50.6551
35	4.7674	4.6836	4.6742	56.7206	53.5247	50.2594
40	4.7571	4.6696	4.6549	56.6754	53.4617	50.1757

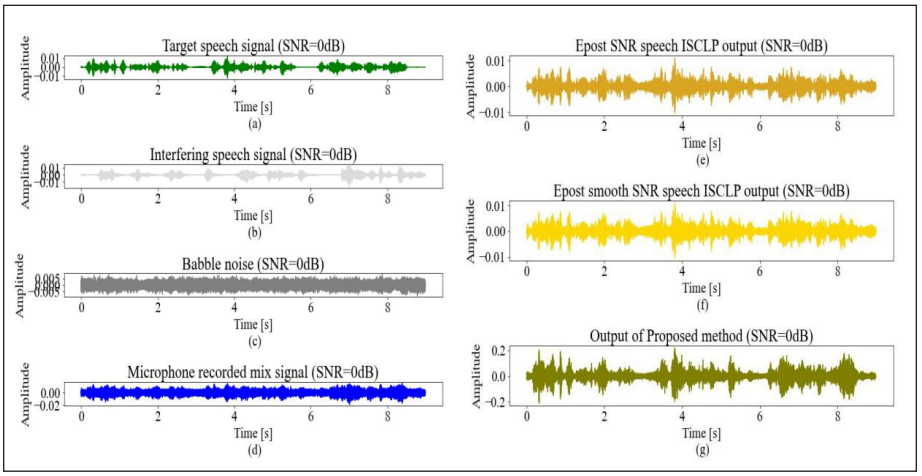


Figure 2. Different time-domain input and output waveforms processed at input SNR = 0 dB. (a) Target speech signal; (b) Interfering speech signal; (c) Babble noise; (d) Microphone-recorded mixed signal; (e) Epost SNR speech ISCLP output; (f) Epost smoothed SNR speech ISCLP output; (g) Proposed method output

Figure 3a, 3b, 3c, and 3d display the time domain waveforms of the target speech, interfering speech, babble noise, and mixed speech, respectively, affected by an input SNR of 40dB and in a reverberant environment with rt60 = 0.61. The time domain waveforms of the output of the ISCLP SNR method, the output of the ISCLP smooth SNR method, and the output of the proposed method are shown in Figures 3e, 3f, and 3g, respectively. Similarly, it is evident from Figure 3g compared to Figure 3e and 3f that the amplified speech signal is clearly observed at the output of the proposed method when compared to the results of the two baseline methods.

The proposed method consistently achieves higher PESQ scores than the ISCLP baseline across all tested SNR conditions, as shown in Table 1, while improvements in ESTOI remain modest. In contrast, Table 2 demonstrates a clear and consistent reduction in CD and WSS values. These reductions indicate improved spectral fidelity and reduced speech distortion, suggesting that the proposed framework enhances speech naturalness even when intelligibility gains are limited. In practical communication scenarios such as teleconferencing and hearing assistance, such improvements are associated with reduced listening effort and improved long-term listening comfort.

The proposed framework combines spatial filtering and diffuseness-based suppression to enhance speech in reverberant environments. Although the CDR estimator operates directly on microphone-domain signals while enhancement is applied to the MVDR beamformer output, the system exhibits stable behaviour under the evaluated conditions. The MVDR stage attenuates spatially off-target and incoherent components of the babble noise field, which reduces the sensitivity of the overall system to representation mismatch between microphone-domain CDR estimation and beamformer-domain enhancement. Performance under strong directional interference was not explicitly evaluated and represents an important direction for future investigation.

While MVDR beamforming contributes to spatial noise reduction, residual reverberation and diffuse noise components remain, particularly in environments with moderate to high reverberation. The consistent improvements observed in PESQ, CD and WSS across all tested SNR levels can therefore be primarily attributed to the CDR-based post-filter, which explicitly models and suppresses diffuse energy. By complementing MVDR beamforming, the post-filter improves spectral fidelity beyond what can be achieved through spatial selectivity alone.

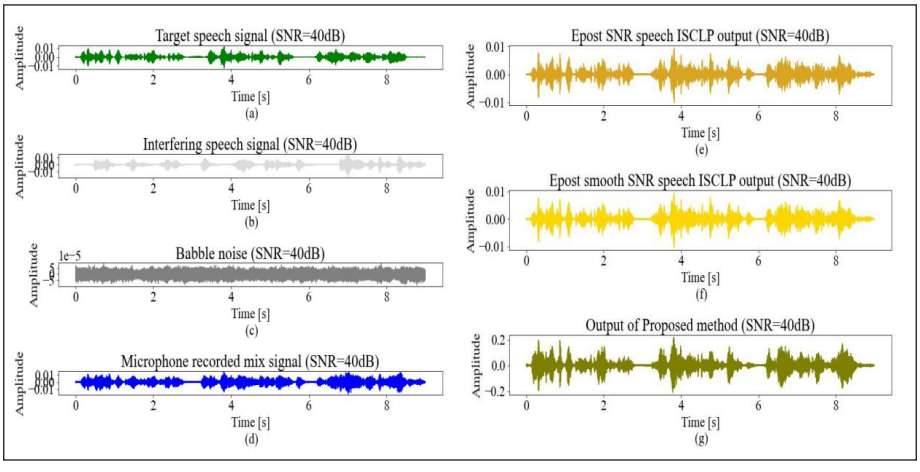


Figure 3. Different time-domain input and output waveforms processed at input SNR = 40 dB. (a) Target speech signal; (b) Interfering speech signal; (c) Babble noise; (d) Microphone-recorded mixed signal; (e) Epost SNR speech ISCLP output; (f) Epost smoothed SNR speech ISCLP output; (g) Proposed method output

Beyond objective metric comparison, the proposed framework differs qualitatively from ISCLP in its architectural behaviour during speech pauses and low-energy segments. The CDR estimation is derived from temporally smoothed cross-power spectral statistics of the microphone signals and does not rely on explicit speech activity detection. Consequently, the diffuseness-based post-filter remains active during silent or low-energy intervals and primarily reflects the characteristics of background noise and reverberation. Similarly, the MVDR beamformer continues to minimise spatially correlated noise power irrespective of speech presence. While pause-specific behaviour was not explicitly analysed, the system design supports continued suppression of diffuse and reverberant components during such segments.

Formal listening tests were not conducted as part of this study. However, informal inspection of the enhanced signals did not reveal severe musical noise artefacts. Nevertheless, subtle speech colouration effects may still be present and are not fully captured by objective metrics such as PESQ and ESTOI. A comprehensive perceptual evaluation is therefore identified as an important direction for future work.

From a computational perspective, the proposed framework relies on closed-form MVDR beamforming and coherence-based CDR estimation and does not involve iterative state-space tracking or adaptive filtering stages such as those employed in the ISCLP baseline, which incorporates Kalman filtering for interference power estimation. Consequently, the proposed method primarily consists of matrix operations and recursive spectral averaging, resulting in reduced computational complexity. Although all experiments were conducted offline and explicit runtime measurements were not reported, the algorithmic structure suggests suitability for real-time implementation on moderate embedded hardware.

Overall, the proposed framework effectively integrates spatial filtering and diffuseness-based suppression to improve speech quality under reverberant conditions while maintaining stable behaviour and low computational complexity.

CONCLUSION

This study presents a new framework that integrates the CDR estimator with the MVDR beamformer to address both reverberation and suppression of babble noise in complex acoustic settings. The CDR estimator in the proposed system supports spectral enhancement through SNR estimation and acts as a post-filter to suppress residual reverberation. The proposed approach improves robustness against babble noise and late reflections. The experimental results have confirmed that the proposed method has shown slightly better evaluation metric scores of PESQ, CD, and WSS as compared to two ISCLP-based baseline systems, especially the perceptual quality, and spectral accuracy improvement was observed to be better in the proposed method over the baseline method. The experiment has been tested considering moderate reverberant conditions with RT60 = 0.61 s, with one source

and babble noise. The experiment was tested by recording the voice of the speaker at 2m from the microphone. The proposed method assumes the presence of a single dominant target speaker. In scenarios involving multiple speakers with comparable power and partial spatial overlap, steering vector estimation and coherence-based separation may become unreliable, potentially leading to target suppression or interference leakage. Addressing multi-speaker conditions remains an important extension of this work. In addition, future studies will investigate the robustness of the proposed framework under increased source-to-microphone distances beyond 2 m, diverse noise conditions, and higher reverberation levels. The correction factor used to adapt the microphone-domain CDR estimate to the beamformer output was evaluated under fixed array geometry and stationary acoustic conditions. While stable performance was observed in this setting, rapidly changing acoustic scenes or microphone array perturbations were not investigated. Future work will focus on adaptive correction mechanisms to improve robustness under dynamic conditions. The experimental evaluation was limited to babble noise, which predominantly exhibits diffuse characteristics. Performance under transient, impulsive, or strongly directional interference sources was not evaluated and therefore remains an open research question to be explored in future studies. The proposed framework assumes a fixed linear array geometry consisting of five microphones with an inter-element spacing of 8 cm. Changes in array spacing, the number of microphones, or physical array deformation can affect spatial covariance estimation and steering vector computation, which in turn may influence the performance of MVDR beamforming and the subsequent CDR-based post-filter. The current study does not explicitly evaluate sensitivity to such array variations, and investigating robustness under different array configurations remains an important direction for future work.

The proposed framework offers several practical advantages over learning-based approaches due to its fully analytical formulation. The method provides interpretability and predictable behaviour, does not require training data, and exhibits low computational complexity. These properties make it particularly suitable for real-time and resource-constrained applications such as hearing aids, embedded communication devices, and safety-critical speech enhancement systems, where robustness, transparency, and reliable behaviour across acoustic conditions are essential.

ACKNOWLEDGEMENT

This research was funded by the Research Management Centre, Universiti Putra Malaysia, Grant Putra 9695500.

REFERENCES

Bai, J., Li, H., Zhang, X., & Chen, F. (2025). Attention-based beamformer for multi-channel speech enhancement. *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1-5. https://doi.org/10.1109/ICASSP49660.2025.10890720

Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 120(1), 331-342. https://doi.org/10.1121/1.2202888

Cauchi, B., Kodrasi, I., Rehr, R., Gerlach, S., Jukic, A., Gerkmann, T., Doclo, S., & Goetze, S. (2015). Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech. *EURASIP Journal on Advances in Signal Processing*, 2015, Article 61, 1-12. https://doi.org/10.1186/s13634-015-0242-x

Chang, H., Hsu, Y., & Bai, M. R. (2024). Deep beamforming for speech enhancement and speaker localisation with an array response-aware loss function. *Frontiers in Signal Processing*, 4, 1-8. https://doi.org/10.3389/frsip.2024.1413983

Cohen, N., Hazan, G., Schwartz, B., & Gannot, S. (2021). An online algorithm for echo cancellation, dereverberation and noise reduction based on a Kalman-EM method. *EURASIP Journal on Audio, Speech, and Music Processing*, 2021, Article 33, 1-17. https://doi.org/10.1186/s13636-021-00219-2

Delcroix, M., Yoshioka, T., Ogawa, A., Kubo, Y., Fujimoto, M., Ito, N., Kinoshita, K., Espi, M., Araki, S., Hori, T., & Nakatani, T. (2015). Strategies for distant speech recognition in reverberant environments. *EURASIP Journal on Advances in Signal Processing*, 2015, Article 1, 1-15. https://doi.org/10.1186/s13634-015-0245-7

Dietzen, T., Huleihel, N., Doclo, S., Moonen, M., Waterschoot, T. V., Spriet, A., & Tirry, W. (2015). Speech dereverberation by data-dependent beamforming with signal pre-whitening. In *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO)*, Nice, France. https://doi.org/10.1109/EUSIPCO.2015.7362827

Dietzen, T., Doclo, S., Moonen, M., & Waterschoot, T. V. (2018). Joint multimicrophone speech dereverberation and noise reduction using integrated sidelobe cancellation and linear prediction. In *Proceedings of the 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 221-225). https://doi.org/10.1109/IWAENC.2018.8521250

Dietzen, T., Doclo, S., Moonen, M., & Waterschoot, T. V. (2020). Integrated sidelobe cancellation and linear prediction Kalman filter for joint multimicrophone speech dereverberation, interfering speech cancellation, and noise reduction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 740-754. https://doi.org/10.1109/TASLP.2020.2966869

Fujita, Y., Nugraha, A. A., Di Carlo, D., Bando, Y., Fontaine, M., & Yoshii, K. (2024). Run-time adaptation of neural beamforming for robust speech dereverberation and denoising. In *Proceedings of the 2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. https://doi.org/10.1109/APSIPAASC63619.2025.10849318

Gode, H., Tammen, M., & Doclo, S. (2021). Joint multi-channel dereverberation and noise reduction using a unified convolutional beamformer with sparse priors. In *Proceedings of the ITG Conference on Speech Communication* (pp. 144-148). https://doi.org/10.48550/arXiv.2106.01902

Gode, H., & Doclo, S. (2023). Adaptive dereverberation, noise and interferer reduction using sparse weighted linearly constrained minimum power beamforming. *arXiv*. https://doi.org/10.23919/EUSIPCO55093.2022.9909809

Habets, E. A. P., & Benesty, J. (2013). A two-stage beamforming approach for noise reduction and dereverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5), 945-958. https://doi.org/10.1109/TASL.2013.2239292

Huang, G., Jensen, J. R., Chen, J., Benesty, J., Christensen, M. G., Sugiyama, A., Elko, G., & Gaensler, T. (2025). Advances in microphone array processing and multichannel speech enhancement. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 1-5). https://doi.org/10.1109/ICASSP49660.2025.10888510

Kobayashi, Y., & Kondo, K. (2019). Japanese speech intelligibility estimation and prediction using objective intelligibility indices under noisy and reverberant conditions. *Applied Acoustics*, 156, 327-335. https://doi.org/10.1016/j.apacoust.2019.07.034

Kumar, C. R., & Chitra, M. P. (2022). Implementation of modified Wiener filtering in frequency domain in speech enhancement. *International Journal of Advanced Computer Science and Applications*, 13(2), 434-439.https://doi.org/10.14569/IJACSA.2022.0130251

Le, X., Chen, H., Chen, K., & Lu, J. (2021). DPCRN: Dual-path convolution recurrent network for single channel speech enhancement. In *Proceedings of Interspeech* (pp. 2811-2815). https://doi.org/10.21437/Interspeech.2021-296

Lee, C. H., Patel, K., Yang, C., Shen, Y., & Jin, H. (2024). An MVDR-embedded U-Net beamformer for effective and robust multichannel speech enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8541-8545). https://doi.org/10.1109/ICASSP48485.2024.10448366

Li, A., Liu, W., Zheng, C., & Li, X. (2022). Embedding and beamforming: All-neural causal beamformer for multichannel speech enhancement. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 6487-6491). https://doi.org/10.1109/ICASSP43922.2022.9746432

Liu, J., & Zhang, X. (2021). Inplace gated convolutional recurrent neural network for dual-channel speech enhancement. *arXiv*. https://doi.org/10.21437/Interspeech.2021-899

Loizou, P. C. (2007). *Speech enhancement: Theory and practice*. CRC Press.

Luo, Y., Han, C., Mesgarani, N., Ceolini, S. C., & Liu, E. (2019). FaSNet: Low-latency adaptive beamforming for multi-microphone audio processing. In *Proceedings of the 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 260-267). https://doi.org/10.1109/ASRU46091.2019.9003849

Nakatani, T., & Kinoshita, K. (2019a). A unified convolutional beamformer for simultaneous denoising and dereverberation. *IEEE Signal Processing Letters*, 26(6), 903-907. https://doi.org/10.21437/Interspeech.2019-1286

Nakatani, T., & Kinoshita, K. (2019b). Simultaneous denoising and dereverberation for low-latency applications using frame-by-frame online unified convolutional beamformer. In *Proceedings of Interspeech*, Graz, Austria (pp. 111-115). https://doi.org/10.21437/Interspeech.2019-1286

Natarajan, S., Al-Haddad, S. A. R., Hassan, M. K., Kamil, R., Azrad, S., Ahmad, F. A., & Macleans, J. F. (2023). A comprehensive review of beamforming-based speech enhancement techniques, IoT, and smart

city applications. In *Proceedings of the 2023 IEEE 2nd Industrial Electronics Society Annual On-Line Conference (ONCON)* (pp. 1-6). https://doi.org/10.1109/ONCON60463.2023.10431158

Natarajan, S., Al-Haddad, S. A. R., Hassan, M. K., Kamil, R., Azrad, S., Ahmad, F. A., & Macleans, J. F. (2024). Revolutionising speech clarity: Unveiling a novel approach with hybrid coherent-to-diffuse power ratio and recursive least square algorithm for reverberation removal. In *Proceedings of the 2024 8th International Conference on Digital Signal Processing (ICDSP '24)* (pp. 140-145). https://doi.org/10.1145/3653876.3653887

Natarajan, S., Al-Haddad, S. A. R., Ahmad, F. A., Kamil, R., Hassan, M. K., Azrad, S., Macleans, J. F., Abdulhussain, S. H., Mahmmod, B. M., Saparkhojayev, N., & Dauitbayeva, A. (2025). Deep neural networks for speech enhancement and speech recognition: A systematic review. *Ain Shams Engineering Journal*, 16(7), Article 103405, 1-35. https://doi.org/10.1016/j.asej.2025.103405

O'Shaughnessy, D. (2024). Speech enhancement-A review of modern methods. *IEEE Transactions on Human-Machine Systems*, 54(1), 110-120. https://doi.org/10.1109/THMS.2023.3339663

Ren, X., Zhang, X., Chen, L., Zheng, X., Zhang, C., Guo, L., & Yu, B. (2021). A causal U-Net based neural beamforming network for real-time multi-channel speech enhancement. In *Proceedings of Interspeech* (pp. 1832-1836). https://doi.org/10.21437/Interspeech.2021-1457

Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (2001). Perceptual evaluation of speech quality (PESQ): A new method for speech quality assessment of telephone networks and codecs. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Vol. 7, pp. 749-752). https://doi.org/10.1109/ICASSP.2001.941023

Sarradj, E. (2010). A fast signal subspace approach for the determination of absolute levels from phased microphone array measurements. *Journal of Sound and Vibration*, 329(9), 1553-1569. https://doi.org/10.1016/j.jsv.2009.11.009

Schwarz, A., & Kellermann, W. (2015). Coherent-to-diffuse power ratio estimation for dereverberation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(6), 1006-1018. https://doi.org/10.1109/TASLP.2015.2418571

Shankar, N., Bhat, G. S., Panahi, I. M. S., Tittle, S., & Thibodeau, L. M. (2021). Smartphone-based single-channel speech enhancement application for hearing aids. *The Journal of the Acoustical Society of America*, 150(3), 1663-1673.https://doi.org/10.1121/10.0006045

Song, S., Cheng, L., Luan, S., Yao, D., Li, J., & Yan, Y. (2021). An integrated multi-channel approach for joint noise reduction and dereverberation. *Applied Acoustics*, 171, Article 107526. https://doi.org/10.1016/j.apacoust.2020.107526

Su, G., & Morf, M. (1983). The signal subspace approaches multiple wide-band emitter locations. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(6), 1502-1522. https://doi.org/10.1109/TASSP.1983.1164233

Tan, K., & Wang, D. L. (2019). Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 380-390. https://doi.org/10.1109/TASLP.2019.2955276

Tan, K., Wang, Z. Q., & Wang, D. L. (2022). Neural spectrospatial filtering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, 605-621. https://doi.org/10.1109/TASLP.2022.3145319

Thimmraja, Y. G., Nagaraja, B. G., & Jayanna, H. S. (2022). A spatial procedure to spectral subtraction for speech enhancement. *Multimedia Tools and Applications*, 81, 23633-23647. https://doi.org/10.1007/s11042-022-12152-3

Xi, J., Xu, Z., Zhang, W., Zhao, L., & Xie, Y. (2024). Speech enhancement algorithm based on microphone array and lightweight CRN for hearing aid. *Electronics*, 13(22), Article 4394. https://doi.org/10.3390/electronics13224394

Xiang, Q., Chen, J., Benesty, J., Lei, T., & Pan, C. (2025). Design of the Wiener gain in noisy and reverberant environments. *Applied Acoustics*, 231, Article 110491. https://doi.org/10.1016/j.apacoust.2024.110491

Yang, W., Huang, G., Brendel, A., Chen, J., Benesty, J., Kellermann, W., & Cohen, I. (2022). A bilinear framework for adaptive speech dereverberation combining beamforming and linear prediction. In *Proceedings of the 2022 International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 1-5). https://doi.org/10.1109/IWAENC53105.2022.9914728

Zhao, Y., Jensen, J. R., Jensen, T. L., Chen, J., & Christensen, M. G. (2020). Experimental study of robust acoustic beamforming for speech acquisition in reverberant and noisy environments. *Applied Acoustics*, 170, Article 107531, 1-13. https://doi.org/10.1016/j.apacoust.2020.107531

Pertanika J. Sci. & Technol. 34 (2): 975 - 1004 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

A Genetic Algorithm-optimised Hybrid Framework Integrating Statistical Forecasting Models and Extreme Learning Machine for Dengue Case Forecasting in the United States of America

Dipankar Das[1]*, Arijit Chakraborty[2], Sajal Mitra[3], and Soumyabrata Nag[4]

[1]*Department of Computer Application and Science, Institute of Engineering & Management, University of Engineering and Management Kolkata, 700091 West Bengal, India*
[2]*Department of BCA, The Heritage Academy, 700107 West Bengal, India*
[3]*Department of Computer Science and Business Systems, Heritage Institute of Technology, 700107 West Bengal, India*
[4]*Department of Microbiology, All India Institute of Medical Sciences, Nagpur, 441108 Maharashtra, India*

ABSTRACT

Dengue, a vector-borne disease, has become a global calamity. A robust early warning mechanism could assist authorities in mitigating this hazard. The objective of the present study is to propose a dengue forecasting model employing a high-throughput dataset for eleven months of dengue progression in the United States of America (USA). The methodology integrates the efficacy of well-renowned statistical methods, i.e., Auto ARIMA, Auto ETS, and machine-intelligent methods, namely Extreme Learning Machine (ELM). The final architecture of the proposed approach is realised by using an evolutionary algorithm, i.e., Genetic Algorithm (GA), which demonstrates improvement of model performance and enhances out-of-sample predictive accuracy. Our proposed model obtained a Mean Error (ME) of 16.66, a Root Mean Square Error (RMSE) of 66.84, a Mean Absolute Error (MAE) of 55.22, a Mean Percentage Error (MPE) of 0.30, and a Mean Absolute Percentage Error (MAPE) of 0.79. Our approach convincingly identifies linearity, seasonality, and nonlinearity imprints of the dengue progression in the USA. The model outperforms nineteen other techniques, including seven traditional, three Artificial Intelligence (AI) -based, a generative AI, and eight hybrid methods. The significance of the findings lies in a rigorous validation technique, specifically non-parametric tests, underscoring the practicability of the proposed model in dealing with noisy or incomplete data environments inherent in coping with health-related time-series challenges.

ARTICLE INFO

Article history:
Received: 09 June 2025
Accepted: 11 March 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.16

E-mail addresses:
dipankr.das@gmail.com (Dipankar Das)
arijitchakraborty.besu@gmail.com (Arijit Chakraborty)
sajal.mitra@heritageit.edu (Sajal Mitra)
soumyabratanag@gmail.com (Soumyabrata Nag)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Potential applications of the research could be the development of a robust early warning mechanism that empowers public health efforts with enhanced epidemiological surveillance.

Keywords: Dengue, extreme learning machine, genetic algorithm, hybrid model, model optimisation, time-series forecasting

INTRODUCTION

Dengue is a vector-borne viral disease transmitted by Aedes mosquitoes. It is common in tropical and subtropical regions of Asia, Africa, and the Americas. Dengue often exhibits spatial heterogeneity and variations in distribution. The clinical symptoms of dengue range from ordinary nausea, vomiting, body aches, and fever to critical conditions such as dengue shock syndrome. It becomes a significant challenge in developing anti-dengue immunisation programmes, as highlighted by Mobin (2024), due to the presence of multiple serotypes in dengue and tropical and subtropical climatic conditions. Patil and Pandya (2021) established a positive link between weather fluctuations and dengue progression in their work. The epidemiological forecasting models developed by Mahanty et al. (2024) and Tomov et al. (2023) portrayed disease outbreaks over time.

The forecasting models are primarily classified as statistical, Artificial Intelligence (AI)-based, and hybrid. Their performance is often limited by non-linearity and interpretability issues. Often, relying on a single approach may restrict our insight into the disease outbreaks. It results in sketchy support for healthcare authorities in combating dengue. This work proposes a novel approach. It integrates Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing State Space Model (ETS), and a Genetic Algorithm (GA) optimised Extreme Learning Machine (ELM) method to portray the progression of dengue in the USA. This novel approach is expected to furnish more precise and timely dengue outbreak forecasts, allowing for more efficient public health measures and, eventually, lessening the disease's impact on afflicted communities.

LITERATURE REVIEW

In America, Sanchez-Gendrizet et al. (2022) applied the Long Short-Term Memory (LSTM) method to predict dengue outbreaks. They employed weekly data from 2016 to 2019. The proposed model was also used to predict the dengue prevalence rates and mosquito egg density in Natal, Brazil. The model achieved a high correlation. The values were between 0.87 and 0.92. The model was used to predict dengue infections and outbreaks 4 to 6 weeks in advance.

Saleh and Baiwei (2021) also used LSTM networks and Support Vector Regression (SVR) to forecast dengue infections, using satellite-based weather data for the years 1990-2008 in San Juan, Puerto Rico. In this regard, the LSTM performed better than the SVR.

It (the LSTM) achieved R^2 of 0.75, Mean Absolute Error (MAE) 8.76, Mean Squared Error (MSE) 245.61 and Root Mean Squared Error (RMSE) 15.67.

Hasan et al. (2024) proposed time series models pertaining to several contexts of healthcare planning and resource allocations. They also used external factors and interactions for dengue prevalence in Bangladesh. The ARIMA (1,1,1) had an MAE of 1266.08. Seasonal ARIMA (SARIMA) had MAE, MSE and RMSE values of 1246.442, 531.837 and 5175.86, respectively, for predicting the dengue prevalence in the forecasted year 2024.

Mohammed and Rahman (2023) employed time series analysis to forecast the monthly prevalence of dengue disease in Bangladesh. They specifically applied ARIMA and SARIMA models. The SARIMA (1,0,0) (1,1,1) model was the best-fit model with an RMSE of 9073.657, which was used to forecast the rate of dengue infection period from January 2022 to December 2022.

Sutriyawan et al. (2024) performed an ARIMA model study to predict the dengue incidence by analysing the data from 2014 to 2023, in Bandung, Indonesia. The best-fitted model is ARIMA (3,0,3), having Mean Absolute Percentage Error (MAPE) 33.34, RMSE 10.0493 and MAE 73.42905. Their proposed model predicted a peak for dengue cases in September 2024, estimated at 320 cases.

Rangarajan et al. (2019) forecasted dengue disease in Singapore, Taiwan, Thailand, Brazil, and Mexico using a sparse representation from electronic health records, Google trends, and time series data. Their proposed method, Autoregressive Likelihood Ratio (i.e., ARLR), achieved an average RMSE of 0.514, MAE of 0.531, and MAPE of 0.550. It forecasted dengue incidence in the USA four weeks ahead, which accomplished a 26% average reduction in RMSE, 21% in MAE, and 6% in MAPE for existing forecasting techniques.

Lu et al. (2024) proposed a Susceptible-Infected mosquito vectors and Susceptible-Infected-Recovered human hosts model (SI-SIR). According to expected climate conditions and mosquito bites in Selangor, Malaysia, they concluded that their model outperformed Multiple Linear Regression (MLR) and was better than LSTM at predicting climate. Forecasts from the model extended to 60 weeks. Before the Movement Control Order (MCO) in Malaysia, it had an MAPE of 13.97 and an MAE of 204.36.

Torres et al. (2025) employed an LSTM network-based model on Monte-Carlo dropout to classify and distinguish between healthy and infected Aedes Aegypti mosquitoes, the dengue virus's primary vectors. The proposed model, a 2-layer Bi-LSTM with 256 neurons, detected approximately 5% of uncertainty data from wing beat signals and achieved 94.87% accuracy, 94.87% precision, 94.87% recall, and 94.87% F1-score, which outperformed the K-Nearest Neighbours (KNN), Support Vector Machine (SVM), 2-layer LSTM with 64 neurons, and 2-layer LSTM with 256 neurons models.

Using weather-related datasets, Baker et al. (2021) employed thirteen machine-learning approaches to forecast weekly dengue incidences in San Juan, Puerto Rico, and in the city of Iquitos, Peru. The authors considered the dengue progression data of San Juan from 1999 to 2008 and the dengue progression data from 2000 to 2010 for Iquitos, which indicated that the Poisson Regression Model (PRM), Negative Binomial Regression Model (NBRM), and Random Forest (RF) models achieved the lowest MAE values of 25.6, 25.8, and 26.6, respectively.

Mendoza Chirinos (2024) developed a spatiotemporal neural network model that combines LSTM and Graph Convolution Network (GCN) layers to predict dengue outbreaks. The author used climatic and socioeconomic data assembled from meteorological records. The author's model achieved impressive results with an MSE of 11.0948 for the combination of cases, temperature, and precipitation features set.

The Artificial Neural Networks (ANN), Convolution Neural Networks (CNN), and LSTM models evaluated the effectiveness for diagnosing dengue cases in the study (Dhaked et al., 2025) using monthly surveillance and meteorological data from 2015 to 2019 in Jaipur, Rajasthan, India. Their proposed one-dimensional CNN (1-DCNN) model demonstrated significantly better accuracy than the ANN and LSTM models, achieving an MAE of 31.49, MSE of 3187.43, and RMSE of 56.45 in predicting dengue cases.

In summary, the investigation of dengue propagation and its effects primarily utilised time series and machine learning models, such as ARIMA, LSTM, SVM, ANN, CNN, and KNN. The performance of these models was evaluated using various statistical measures, including RMSE, MSE, MAPE, and MAE.

Forecasting dengue cases is not limited to possible earlier identification. It is also related to various aspects of potential solutions associated with dengue diseases, related to the protection of human health and the safety of social life. This leads to finding a suitable and relevant solution for identifying and managing a possible multi-objective outlook. Thus, the role of computational optimisation techniques is paramount. It presents substantial options for enhancing the forecasting approach for dengue cases and for building a decisive forecasting mechanism. Table 1 presents a statistical summary of these studies.

Motivation and Research Gap

Dengue fever is another major issue of global health that has been made worse by climate change. The use of AI models has been found to be effective in the prediction of dengue outbreaks. The creation of an AI-based prediction system would enable the health officials to receive crucial warnings and be better positioned to control an outbreak and ultimately save lives.

Wang et al. (2023) emphasised the fact that real-life time-series data can be mixed with both linear and non-linear components. They are also non- stationary because of the trends and seasonality. Pereira da Veiga et al. (2024) emphasised that the ARIMA is an

Table 1
Observations of the past investigations for dengue analysis

Author	Dengue disease investigation methods		Statistical measures used				
	ML-based	Time series	RMSE	MSE	MAE	MAPE	R^2
Sánchez-Gendriz et al. (2022)	✔	✔	-	-	-	-	✔
Saleh and Baiwei (2021)	✔	✔	✔	✔	✔	-	✔
Hasan et al. (2024)	✔	✔	✔	✔	✔	-	-
Mohammed and Rahman (2023)	-	✔	✔	-	-	-	-
Mendoza Chirinos (2024)	✔	✔	-	✔	-	-	-
Sutriyawan et al. (2024)	-	✔	✔	-	✔	✔	-
Dhaked et al. (2025)	✔	✔	✔	✔	✔	-	-
Rangarajan et al. (2019)	✔	✔	✔	-	✔	✔	-
Lu et al. (2024)	✔	✔	-	-	✔	✔	-
Baker et al. (2021)	✔	✔	-	-	✔	-	-
Flores et al. (2026)	✔	-	✔	-	-	✔	-
Bunprom et al. (2026)	-	✔	-	-	-	✔	-

effective model that can capture linear elements. Equally, the ETS can overcome trends and seasonality. In their work, Beck et al. (2025) also indicated that the ETS can face difficulties in addressing complex non-linearity.

Artificial Intelligence models can spot complex nonlinear trends. Wu et al. (2025) demonstrated that an ELM, which is an artificial intelligence model, can be applied successfully to non-linear trends of time-series data. However, overall, the models that are based on AI tend to be uninterpretable, and they need a significant amount of hyperparameter tuning to maximise their performance. Those problems indicate shortcomings of the existing models of dengue forecasting using the data of time-series dengue cases.

We found out that hybrid models are improved over individual models in most instances. They integrate numerous strengths and boost forecasting accuracy and strength. Considering the benefits of the ARIMA, ETS and ELM forecasting methods, we suggest a hybrid forecasting model that combines Auto ARIMA, Auto ETS, and ELM to predict dengue.

This combined strategy is expected to improve the out-of-sample predictive performance, thus improving the accuracy and reliability of dengue predictions. Moreover, a GA is used to optimise the ELM model. Though the application of GA is common in multi-objective optimisation, the ability of the technique to narrow down the number of hidden neurons in ELM to predict dengue has not been explored. ELM architecture is composed of one hidden layer with a very large number of neurons,

the number of which is computed to give optimal model complexity and performance. GA systematically recognises the most appropriate number of concealed neurons in ELM by exploring a large solution space in a more systematic manner to reduce forecasting errors, which promotes a decrease in the costs of computation and the general complexity of the model. However, the evolution of the proposed model based on the hybridisation of different forecasting methods with the help of GA should deal with the following research issues:

- How effectively can the GA-tuned hybrid model, integrating Auto ARIMA, Auto ETS, and ELM, improve dengue case predictions in the USA compared to standalone models regarding accuracy and reliability?
- Can GAs optimise the hidden neurons in the GA-tuned ARIMA-ETS-ELM hybrid model for better dengue forecasting?
- What are the contributions of Auto ARIMA, Auto ETS, and ELM in enhancing the accuracy of the GA-tuned ARIMA-ETS-ELM model for dengue forecasting?

Accordingly, the objectives of the present study are outlined as follows:

- Analysing dengue case trends in the USA by exploring historical data.
- Creating a hybrid forecasting model that combines two automated statistical methods, Auto ARIMA and Auto ETS, with an AI approach, specifically ELM, to enhance out-of-sample predictive performance.
- Utilising GA to optimise the number of hidden neurons in ELM ensures an optimal balance between model complexity and forecasting accuracy.
- Comparing the forecasting accuracy of the proposed hybrid model with advanced automatic time-series forecasting models using benchmark metrics.
- Evaluating the reliability of the GA-tuned ARIMA-ETS-ELM model through statistical tests on its forecasts for effective dengue surveillance and early warning systems in public health.

METHODOLOGY

Our product is a new hybrid system that includes the ARIMA, ETS, and ELM models, which are predicted by the GA optimisation to predict cumulative dengue cases in the United States. To make use of a general-purpose GA to maximise the number of hidden nodes, which is a real-valued parameter. To systematically optimise the fitness of the candidate solutions, the GA utilised the conventional evolutionary operators, namely selection, crossover and mutation. The ARIMA-ETS-ELM model used to tune GA is in two layers. The former layer is the automated statistical modelling and involves two methods: ARIMA and ETS. The second layer involves the use of an AI-based forecasting method with an ELM. Here, a GA is used to optimise the ELM regarding the number of hidden nodes, thus making it easier to generate correct forecasts. Figure 1 displays a detailed description of the methodology applied in this research.

The out-of-sample forecasting accuracy of the suggested GA-optimised hybrid approach, which blends ARIMA, ETS, and ELM, is assessed. The following metrics are used in this study: Mean Error (ME), MAE, Mean Percentage Error (MPE), MAPE, and RMSE. ME is the average signed difference between forecasted and observed values, while RMSE emphasises more significant errors, making it sensitive to outliers. It helps assess the stability of a model's predictions. MAPE treats all errors uniformly. It is less sensitive to outliers compared to RMSE. The interpretation of MAPE is proper, where a model with MAPE $\leq$ 10 per cent is considered highly accurate (Yadav & Nath, 2019).

We compared its performance against nineteen automated techniques. Three component models ARIMA (Sucipto et al., 2025), ETS (Li, 2024), and ELM (Rahman et al., 2024); four traditional statistical methods Single Exponential Smoothing (SES) (Saragih et al., 2024), Holt's Linear Trend (HLT) (Ariyanto & Nugraha, 2024), Holt-Winters' (HW) (Santosa et al., 2024), Seasonal-Trend Decomposition using LOESS (STL), which separates a time series into seasonal, trend, and residual components (Yunisa & Siregar, 2023); three AI-based approaches – Neural Network Autoregressive (NNAR) (Ahmar & Boj, 2021), Multilayer Perceptron (MLP) (Chakraborty et al., 2024), and SVR (Parreño & Anter, 2024); one generative AI technique namely, Moirai (Woo et al., 2024); and eight hybrid models namely, GA-tuned ARIMA-ELM, GA-tuned ETS-ELM, ARIMA-ANN using equal weight, ETS-ANN using equal weight, and ARIMA-ETS-ANN using equal weight (Perone, 2021), ARIMA-ANN using CV-error (Talkhi et al., 2021), STL-ETS and STL-ARIMA (Hightower et al., 2024) are used. In these eight ANN-based hybrids, both GA-tuned techniques are applied for the ablation study to identify their contributions to the hybrid model.

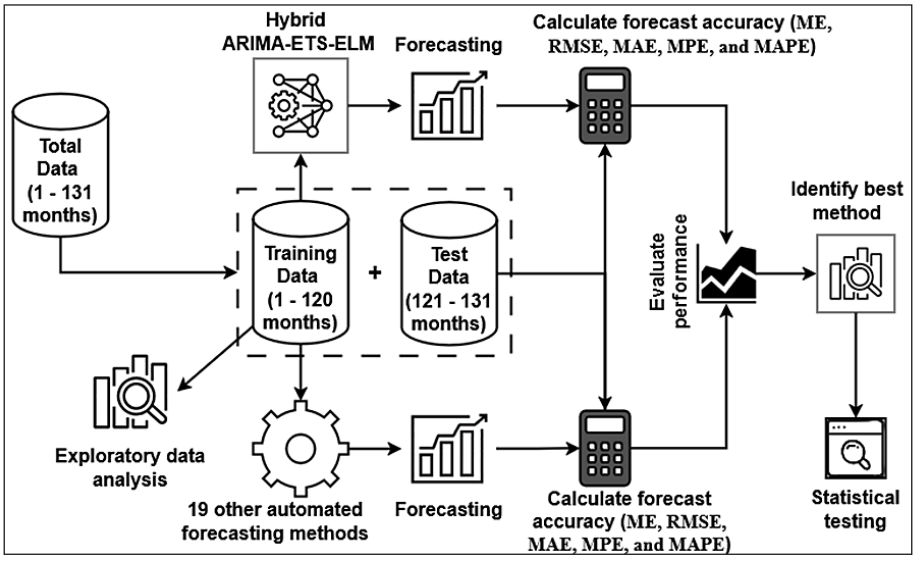


Figure 1. Overview of the research methodology

To assess the robustness of the forecasting performance of the identified best method between these twenty approaches, i.e., the GA-tuned ARIMA-ETS-ELM approach and the other nineteen techniques, we conducted statistical significance testing of the best method using the Mann-Whitney U test (Xiong et al., 2022) and the Wilcoxon Signed Rank test (Ahmadoni, 2025).

Dataset Construction

We obtained historical and high-throughput data regarding dengue cases from a publicly accessible open data source in the United States. The dataset comprises monthly dengue case statistics over 131 months, covering January 2014 to November 2024 (World Health Organisation, 2025). We utilised the monthly data to calculate the cumulative dengue cases in the USA. This cumulative case data is then split into a training set for model development and a test set for evaluating model performance. Comprehensive exploratory data analysis is conducted on the training set data to identify the trends, i.e., fitting trend lines (Das & Chakrabarti, 2021), seasonality using the Wilcoxon Signed-Rank test on the Buys-Ballot table's row variances (Nwogu et al., 2016), non-linearity employing Brock-Dechert-Scheinkman, i.e., BDS test, which is used detect nonlinear patterns in time series data (Jirasakuldech et al., 2023), and stationarity utilizing Augmented Dickey-Fuller, i.e., ADF test, which checks whether a time series is stationary (Ivanovski & Ivanovska, 2024) of the cumulative dengue cases.

The original dataset contained monthly dengue cases for each month. This monthly data was converted into cumulative dengue cases to capture the progressive trend of disease spread. Here, each observation represents the total number of cases up to that month (i.e., cumulative sum up to that month). This transformation smooths short-term fluctuations and highlights the long-term growth pattern of dengue incidence.

Cumulative dengue cases were modelled to highlight long-term outbreak growth, smoothing reporting irregularities and enabling hybrid model evaluation under mixed linear–nonlinear non-stationary conditions. It supports medium-term planning despite the monotonic nature of the cumulative series. This transformation can inflate the denominator, potentially reducing percentage-based errors like MAPE; therefore, model evaluation also relied on scale-dependent metrics (RMSE, MAE, ME) to ensure comparative consistency across all nineteen models.

Data Description

Figure 2 illustrates the cumulative data on dengue cases in the USA over time. The 131-month dataset is partitioned into a 120-month training set (1-120 months) and an

11-month test set (121-131 months). The 11-month test set, representing genuinely unseen future observations, serves as a real-time forward holdout that preserves temporal ordering, captures post-trend stabilisation, and reflects realistic one-year public-health forecasting horizons. Future work may incorporate rolling-origin evaluation for additional robustness. Table 2 provides the summary statistics and test results for the training set data of the cumulative dengue cases in the USA.

Table 2 demonstrates that the training dataset of cumulative dengue cases in the USA consists of 120 months of data without any missing values. The mean of 3435 cumulative cases has a median of 3393 and a standard deviation of 1823, indicating considerable variability. The minimum and maximum recorded cumulative cases are 54 and 6278, respectively.

The very high R^2 value of 0.976 and adjusted R^2 of 0.976 in the linear trend analysis and a statistically significant F-test with p-value < .001 confirm a strong linear trend in the cumulative cases. In contrast, the seasonality in the cumulative cases is not statistically significant, as indicated by the p-value of 0.7695 of the Wilcoxon Signed-Rank test. The BDS-test with p-value < 0.001 for both m=2 and m=3 affirms the presence of significant non-linearity in the cumulative dengue cases data. The cumulative dengue cases data is non-stationary, as suggested by the p-value of 0.2965 of the ADF test.

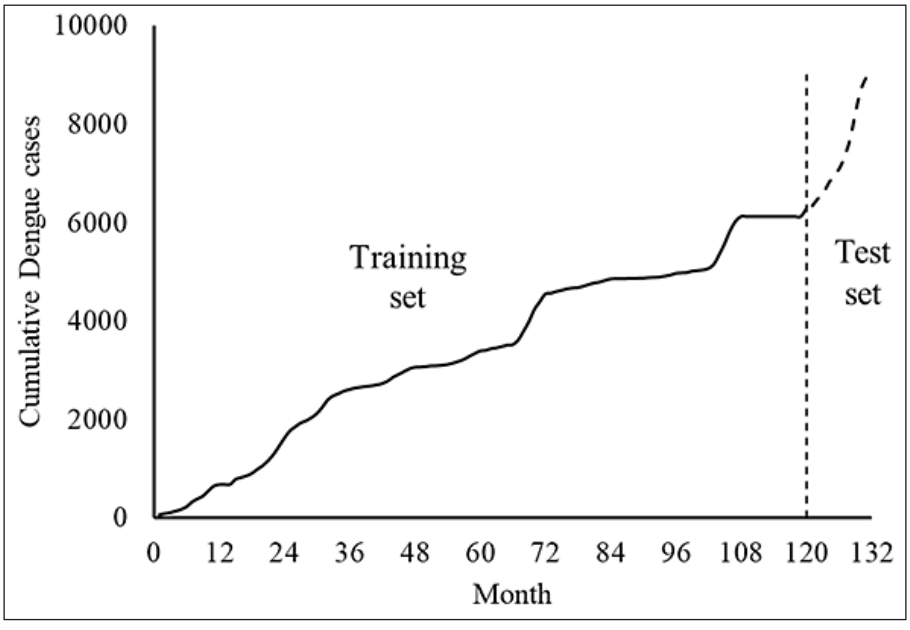


Figure 2. Cumulative dengue cases in the USA

Table 2
Summary statistics and test results for cumulative dengue cases in the USA

Statistic	Value
N (length of training set)	120
Missing	0
Mean	3435
Median	3393
Standard deviation	1823
Minimum	54
Maximum	6278
R^2 of linear trend line	.976
Adjusted R^2 of the linear trend line	.976
Significance of the F-test of the linear trend line	< .001
Significance of the Wilcoxon Signed-Rank test	.7695
Significance of the BDS test	<.001 (or embedding dimension of m=2), <.001 (or embedding dimension of m=3)
Significance of the ADF test	.2965

Proposed Model Development

Table 2 shows that the cumulative dengue case training data displays a strong linear trend. However, it also shows non-linearity and non-stationarity. These characteristics indicate that standalone models such as ARIMA and ETS may not fully capture the complex patterns within the data. Conversely, an ELM can capture non-linearity. Nevertheless, given the strong linear trend present in the data, ELM may struggle to model this effectively without incorporating trend modelling. A hybrid approach that combines ARIMA, ETS, and ELM may be more effective. ARIMA and ETS handle the trend component effectively, and ELM captures the non-linear dependencies. Figure 3 illustrates the integration of statistical (ARIMA and ETS) and AI (ELM) techniques in the proposed GA-tuned ARIMA-ETS-ELM approach forecasting framework.

The proposed hybrid GA-tuned ARIMA-ETS-ELM used the lagged values of the dengue time series and the fitted values obtained from ARIMA and ETS models as input to the ELM. During training, the fitted values from ARIMA and ETS on the training dataset served as external regressors. It enables the ELM to learn linear through ARIMA and trend, level, and seasonality through ETS components, along with the nonlinear components through ELM. During forecasting, the ARIMA and ETS models first generate forecasts for the test period. These forecasted values are then supplied as external inputs to the ELM to produce the final hybrid predictions. This ensures consistency between the training and forecasting phases. Leakage is prevented as follows: during training,

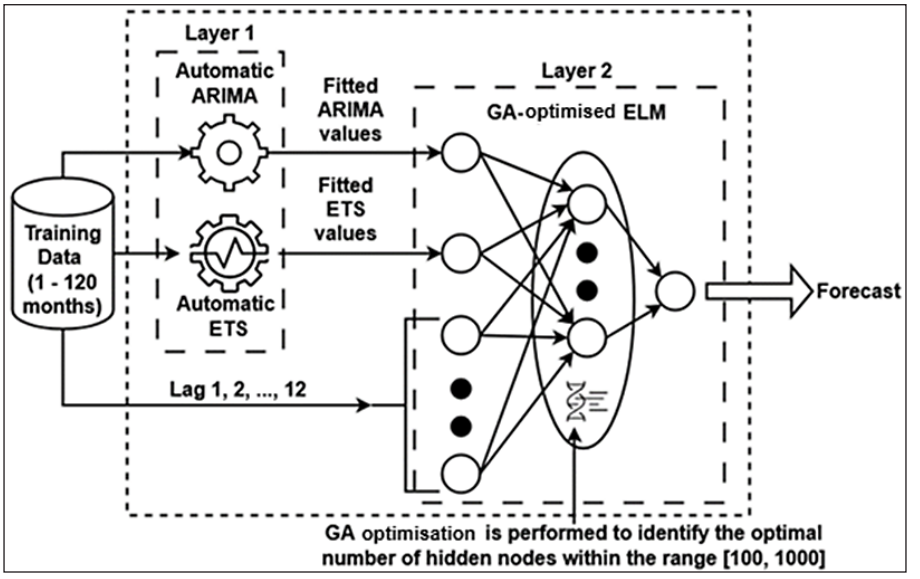


Figure 3. Proposed GA-tuned ARIMA-ETS-ELM hybrid approach

ELM uses fitted values generated strictly from the training set; during forecasting, ARIMA and ETS first generate forecasts for the test period, and these forecasted, not observed values, are then fed into ELM. Thus, no actual future observed test values enter the model during prediction, ensuring temporal integrity.

Mathematical Foundations of the Proposed Methodology

Considering the training data of the time series as $\{s_t\}_{t=1}^{N}$ (where *N* denotes the length of the training set). The methodology consists of two layers, as outlined below.

Working of Layer 1: The time series $\{s_t\}_{t=1}^{N}$ is initially modelled using automatic ARIMA and ETS techniques. The fitted values obtained from ARIMA and ETS are:

$$\hat{s}_t^{ARIMA} = \alpha_0 + \sum_{I=1}^{P} \alpha_i s_{t-i} + \sum_{j=1}^{q} \gamma_j e_{t-j} \quad [1]$$

$$\hat{s}_t^{ETS} = f(L_t, B_t, S_t, \epsilon_t) \quad [2]$$

In Equation 1, $\hat{s}_t^{ARIMA}$ represents the ARIMA fitted value at time *t*, α_0 is the intercept, α_i are autoregressive coefficients, γ_j are moving average coefficients, s_{t-i} are either lagged values or differenced values (if differencing is applied), and e_{t-j} denotes past error terms.

The order of the AR and MA components is *p* and *q*, respectively. In Equation 2, the ETS fitted values (at time *t*), $\hat{s}_t^{ETS}$, are derived from a combination of levels (L_t), trend (B_t), seasonality (S_t) and error (ϵ_t) components. ϵ_t affects the components, and not directly the fitted values. The function $f(\cdot)$ captures the relationship among these components. $f(\cdot)$ depends on the ETS model structure.

Working of Layer 2: In this layer, the ELM is employed to refine forecasting. The input to the ELM consists of the fitted values from ARIMA and ETS, along with the last 12 lagged values of the time series. The 12-lag structure reflects annual seasonality in monthly data, common epidemiological lags in vector-borne diseases, and standard practice in monthly time-series modelling. Future sensitivity analysis with alternative lags (6, 9, 18 months) could further evaluate robustness. Defining the lagged vector (v_t) as given in Equation 3:

$$v_t = [s_{t-1}, s_{t-2}, \dots, s_{t-12}] \quad [3]$$

The combined input vector (u_t) to ELM at time *t* as given in Equation 4:-

$$u_t = [v_t, \hat{s}_t^{ARIMA}, \hat{s}_t^{ETS}] \quad [4]$$

The ELM processes this input using a single hidden layer with *L* hidden nodes, where *L* is determined using a GA. The GA optimises *L* within the range [100, 1000], ensuring optimal performance. Each hidden node *j* (where $j = 1, 2, \cdots, L$) applies a transformation as follows:

$$h_j(u_t) = g\left(w_j^T u_t + b_j\right) \quad [5]$$

In Equation 5, w_j is the weight vector, w_j^T denotes its transpose, b_j is the bias, and $g(\cdot)$ is the activation function.

The final output of the ELM model is given by:

$$\hat{s}_t^{ELM} = \sum_{j=1}^{L} \beta_j h_j(u_t) \quad [6]$$

In Equation 6, β_j are the output weights connecting the hidden layer to the forecasted value and $h_j(u_t)$ is the activation (output). These weights are computed using the Moore-Penrose pseudo-inverse method:

$$\beta = H^{\dagger}T \quad [7]$$

In Equation 7, H is the hidden layer output matrix, as given in Equation 8, $H^{\dagger}$ is its Moore-Penrose pseudo-inverse, and β the output weight vector is defined in Equation 9, and T is the training target matrix, as given in Equation 10:

$$H = \begin{bmatrix} g(w_1^T u_1 + b_1) & \cdots & g(w_L^T u_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1^T u_N + b_1) & \cdots & g(w_L^T u_N + b_N) \end{bmatrix}_{N \times L} \quad [8]$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{bmatrix}_{L \times 1} \quad [9]$$

$$T = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}_{N \times 1} \quad [10]$$

To enhance robustness, the proposed approach trains 20 different networks and aggregates their forecasts using the median operator:

$$\hat{s}_t^{Hybrid} = median\left(\hat{s}_t^{ELM(1)}, \hat{s}_t^{ELM(2)}, \ldots, \hat{s}_t^{ELM(20)}\right) \quad [11]$$

In Equation 11, $\hat{s}_t^{ELM(k)}$ represents the forecast from the *k*-th ELM network.

The complete GA-tuned ARIMA-ETS-ELM approach can be mathematically summarised as:

$$\hat{s}_t^{Hybrid} = median\left(\sum_{j=1}^{L_k} \beta_j^{ELM(k)} g\left(w_j^{ELM(k)T}[v_t, \hat{s}_t^{ARIMA}, \hat{s}_t^{ETS}] + b_j^{ELM(k)}\right)\right)_{k=1}^{20} \quad [12]$$

In Equation 12, L_k is the number of hidden nodes for the k-th ELM network, selected via GA within the range [100, 1000]. Each weight, bias, and activation function parameter is unique to the corresponding ELM network.

Algorithm for Identifying the Optimal Number of Hidden Nodes of the GA-tuned ARIMA-ETS-ELM

The algorithm for determining the optimal no. of hidden nodes within the proposed GA-tuned ARIMA-ETS-ELM model is outlined in Figure 4. The range of hidden neurons [100, 1000] was chosen based on a heuristic approach. During the preliminary experiment, we observed that a sufficient search space for the GA was required to identify an optimal network size. A smaller range resulted in underfitting, and a broader range was able to minimise the RMSE and effectively regularise model complexity. The GA search range of 100-1000 hidden nodes was chosen because preliminary experiments showed underfitting below 100 neurons and RMSE stabilised above 100. A wide search space helps GA avoid local minima. The final optimal value (725) was determined via fitness minimisation rather than arbitrarily. The GA operated with a population size of 20 over 50 generations. A linear arithmetic crossover strategy was employed. The crossover probability and mutation probability were set to 0.8 and 0.1, respectively. In ELM, input weights and biases are randomly assigned, output weights are analytically computed, and the activation function is typically fixed, making hidden-node size the principal structural hyperparameter affecting the bias-variance trade-off.

Input: Training data, ARIMA and ETS models (to generate fitted values), Population size (for the GA), Maximum generations (upper limit for GA iterations), Search space (range of hidden neurons for ELM), Stopping criterion (number of generations without improvement to terminate GA)
Output: Optimised number of hidden neurons (hd*) minimising RMSE

1. Preprocessing
1.1. **Train** ARIMA model on training data and obtain fitted values.
1.2. **Train** ETS model on training data and obtain fitted values.

2. Define Fitness Function
2.1 **Train** ELM For a given number of hidden neurons (hd), with:
- **Inputs**: Lag 1, 2, ..., 12, Fitted ARIMA values, Fitted ETS values
2.2 **Compute** the RMSE of the model.
2.3 **Return** the negative RMSE as the fitness value (minimisation objective).

3. Initialize Genetic Algorithm
3.1 **Set** population size = 20 and generation limit = 50.
3.2 **Generate** an initial population of hidden neuron values (hd) randomly within [100, 1000].
3.3 **Apply** elitism: retain the top 5% best solutions in each generation.
3.4 **Set** crossover rate = 0.8 and mutation rate = 0.1.
3.5 **Set** stopping criterion to terminate if no improvement in 10 generations.

4. Evolution Process
***Repeat** until termination criteria (max generations or early stopping) are met:*
4.1 Evaluate the fitness function for each individual in the population.
4.2 Select parent solutions based on fitness scores.
4.3 Apply crossover (0.8 probability) and mutation (0.1 probability) to generate offspring.
4.4 Update population with new offspring while retaining the elite solutions.
4.5 Track the best solution found so far.

5. Termination
5.1 If no improvement in fitness for 10 generations, stop.
5.2 Return the best number of hidden neurons (hd*) minimising RMSE.

Figure 4. Hidden node optimisation algorithm

Therefore, the present work only optimised the hidden-node count, because it directly controls model capacity and complexity in ELM.

Experimental Setup

Table 3 summarises the programming languages and packages/ libraries used in this study to ensure methodological transparency and support replication efforts.

The R scripts used in this study are available at the following GitHub repository: https://github.com/drddas-kolkata/GA-Optimized-ARIMA-ETS-ELM. The codes include data loading, GA-based hyperparameter tuning of the proposed ARIMA-ETS-ELM, building the final optimised hybrid model, generating the 11-month-ahead forecasts using the GA-Optimised-ARIMA-ETS-ELM, and assessing forecast accuracy of the model.

RESULTS AND DISCUSSION

Identification of Optimised GA-tuned ARIMA-ETS-ELM Model

We used a GA to find the optimal number of hidden neurons in the ELM for the GA-tuned ARIMA-ETS-ELM model, aiming to minimise the RMSE. Figures 5 and 6 present the relationship between the number of hidden neurons and the corresponding RMSE and the best RMSE values (i.e., fitness values) in different generations during the GA-based ELM tuning, respectively.

Table 3
Summary of packages and programming languages used in the experimental setup

Language/ Package (Version)	Purpose/ Description	Author
R (4.4.2)	Traditional, AI-based, and hybrid model development and statistical tests	(R Core Team, 2024)
Python (3.11.12)	Applying the Moirai model	(Python Software Foundation, 2025)
stats (R, v4.4.2)	To perform the Wilcoxon Signed-Rank and Mann-Whitney U Test	(R Core Team, 2024)
tseries (R, v0.10.58)	To perform the BDS and ADF tests	(Trapletti & Hornik, 2024)
forecast (R, v8.24.0)	Forecasting using ARIMA, ETS, NNAR, SES, HLT, and HW	(Hyndman et al., 2024; Hyndman & Khandakar, 2008)
GA (R, v3.2.4)	Implementing GA	(Scrucca, 2013; Scrucca, 2017)
nnfor (R, v0.9.9)	Forecasting using MLP and ELM model development	(Kourentzes, 2023)
TSSVM (R, v0.1.0)	Forecasting using SVR	(Ray et al., 2022;)
forecastHybrid (R, v5.0.19)	Forecasting using ARIMA-ANN, ETS-ANN, ARIMA-ETS-ANN, STL-ETS, and STL-ARIMA	*(Shaub & Ellis, 2020)*
uni2ts (Python)	Forecasting using Moirai	(Aksu et al., 2024; Liu et al., 2024; Woo et al., 2024)

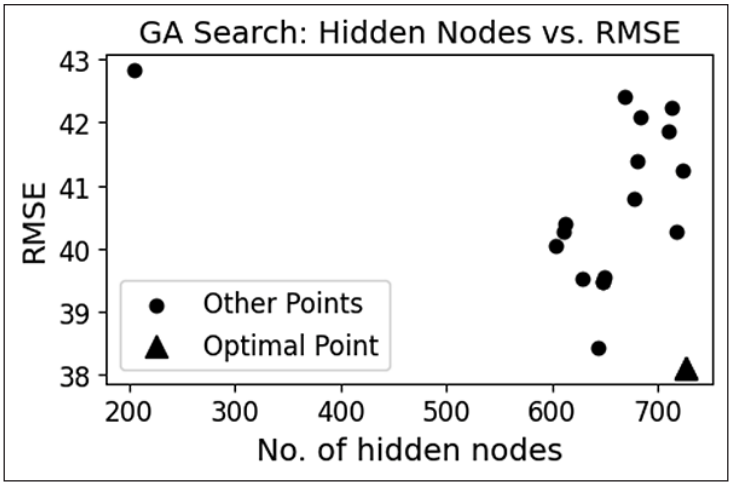


Figure 5. Identification of optimal hidden neurons in ELM via RMSE minimisation

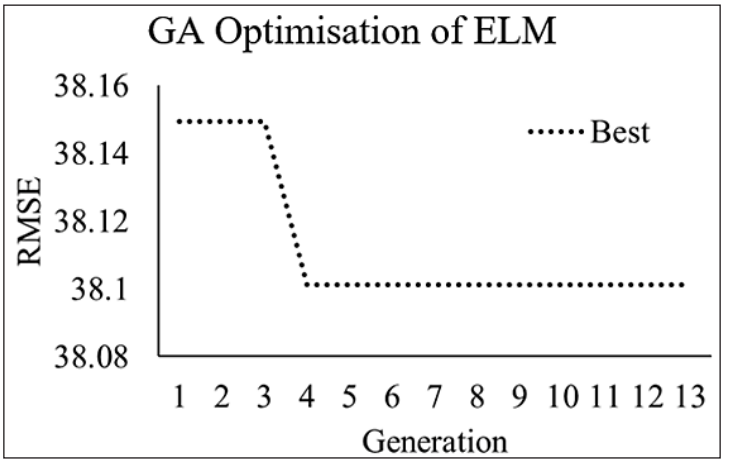


Figure 6. Evolution of the best RMSE across generations during GA optimisation of ELM.

In Figure 5, the triangle-shaped marker designates the optimal quantity of hidden neurons, which is 725, corresponding to the lowest RMSE value of 38.1. The final ELM architecture in the proposed GA-tuned ARIMA-ETS-ELM hybrid model is delineated as follows: 14 (input nodes: fitted ARIMA values, fitted ETS values, and lags 1 to 12 of the observed data) / 725 (hidden nodes: optimal number identified by the GA) / 1 (output node).

The GA-tuned ARIMA-ETS-ELM is a non-likelihood-based model. For this model, a pseudo-BIC was computed using the residual MSE and the number of hidden neurons to approximate model complexity. The resulting value (3770.09) is not directly comparable to the likelihood-based BIC values of ARIMA (1197.74) and ETS (1468.86) models.

The pseudo-BIC of the hybrid model provides only a rough indication of its relative complexity. This pseudo-BIC is used as an approximate complexity measure for the non-likelihood ELM framework. It is not directly comparable to likelihood-based BIC and serves only for internal structural interpretation and is not employed for cross-model ranking or performance comparison.

Overfitting with 725 hidden nodes was controlled using the Moore–Penrose pseudo-inverse solution, GA fitness based on RMSE minimisation, and median aggregation of 20 independently initialised ELM networks. The final evaluation was conducted strictly on unseen test data. We observed consistent improvements across 19 models, further supporting generalisation.

Forecasting Performance of the Proposed Approach

The line plot presented in Figure 7a shows that the observed values and forecasts are very similar. Meanwhile, the density plot in Figure 7b indicates that the distributions of the observed and forecasted values are closely aligned. The relationship indicates that the modes are close to one another; hence, no considerable amount of bias is present. The boxplot in Figure 7c shows that the values observed and predicted have close central tendencies, which indicates that the model can track the overall trend. The model further fails to produce excessive errors. The medians are very close, meaning that the forecast does not significantly overestimate or underestimate the mid-values. Figure 7d reveals the violin plot demonstrates that the overall shape of the observed distributions and forecasted distributions are almost equal and this is the way of knowing that the model is capturing the data. It is worth noting that the average values of the two violins are positioned at a similar point. This observation means that the average observed values are close to the average of the foregoing values. This therefore implies that little to no systematic bias in the forecasting of this model.

Table 4 presents the performance metrics of the GA-tuned ARIMA-ETS-ELM model, including ME, RMSE, MAE, MPE, and MAPE. It shows these metrics for a forecast horizon of 11 months, with individual values labelled as ME (11-month), RMSE (11-month), MAE (11-month), MPE (11-month), and MAPE (11-month). Additionally, Table 4 includes the average values for each of these metrics over the 11 months, denoted as average ME, average RMSE, average MAE, average MPE, and average MAPE.

The ME of 11 months was -52.73, and the average ME is 16.66, indicating that the model does not show a consistent over- or under-forecasting bias. The RMSE penalises significant errors. The RMSE -11 month and the average RMSE of the GA-tuned ARIMA-ETS-ELM were 141.79 and 66.84, respectively. It suggests a reasonable overall forecasting performance. The MAE of 11 months was 100.20, and the average MAE was 55.22. It indicates relatively low absolute error magnitudes.

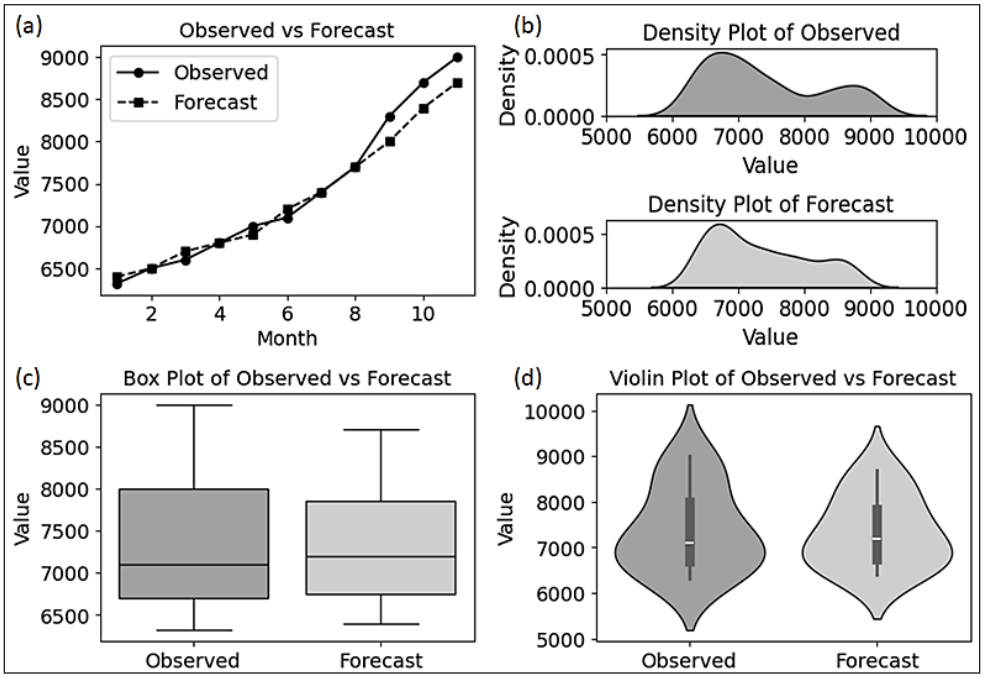


Figure 7. Forecast-performance of GA-tuned ARIMA-ETS-ELM: (a) Observed vs Forecast plot; (b) Density plots of observed and forecast values; (c) Box plot comparison; and (d) Violin plot comparison.

The MPE and MAPE are two percentage-based error measures. Low 11 months MPE of -0.58 and average MPE of 0.30 suggested minimal overall bias in forecast direction. All the MAPE values are well below the threshold of 10 for excellent performance. The MAPE of 11 months was 1.27, and the average MAPE was 0.79. It displays the GA-tuned ARIMA-ETS-ELM's strong forecasting ability.

Assessment of GA-tuned ARIMA-ETS-ELM Performance over Component Models

We performed a comparative analysis between the GA-tuned ARIMA-ETS-ELM and its components – Auto-ARIMA, Auto-ETS, and Auto-ELM, using out-of-sample forecasts over an 11-month horizon to evaluate the performance of the GA-tuned ARIMA-ETS-ELM relative to its individual components. The average ME of the GA-tuned ARIMA-ETS-ELM was 16.66. It was closest to zero among all. It exhibited that the GA-tuned ARIMA-ETS-ELM had the least forecast bias. The average RMSE (66.84) of the hybrid model was the lowest. It indicated that the hybrid model was more stable and accurate. It had the minimum average MAE (55.22) among all. The average MPE of the GA-tuned ARIMA-ETS-ELM was 0.30. It indicated that the model exhibited the least bias (in directional accuracy). It achieved the minimum average MAPE (0.79).

Table 4
Forecasting performance of the GA-tuned ARIMA-ETS-ELM model based on error metrics over an 11-month horizon

Metric	Value
ME 11-month	-52.73
Average ME	16.66
RMSE 11-month	141.79
Average RMSE	66.84
MAE 11-month	100.20
Average MAE	55.22
MPE 11-month	-0.58
Average MPE	0.30
MAPE 11-month	1.27
Average MAPE	0.79

The value was well below the commonly accepted threshold of 10% for excellent forecasting. Therefore, the GA-tuned ARIMA-ETS-ELM outperforms all its components across all evaluations. The GA-tuned ARIMA-ETS-ELM approach improved average MAPE over automatic ARIMA (74.7%), automatic ETS (61.2%), and automatic ELM (79.6%). Similar improvements were observed for average RMSE, also, i.e., 76.4% over automatic ARIMA, 62.2% over automatic ETS, and 80.9% over automatic ELM. It suggested improved forecasting. These findings highlight the advantage of combining statistical techniques with machine learning in developing a more accurate and robust forecasting framework. Table 5 presents these findings.

An ablation study was conducted to assess the individual contributions of Auto ARIMA and Auto ETS in the GA-tuned ARIMA-ETS-ELM framework. The GA-tuned ARIMA-ETS-ELM model was evaluated compared to two simplified variants: (1) the GA-tuned ARIMA-ELM model and (2) the GA-tuned ETS-ELM model. Table 6 displays the results. The GA-tuned ARIMA-ETS-ELM model outperformed both of its simplified variants, namely the GA-tuned ARIMA-ELM and GA-tuned ETS-ELM, across all evaluated metrics, achieving the lowest average ME, RMSE, MAE, MPE, and MAPE. The results indicate that incorporating ETS in the ARIMA-ELM model significantly augments average MAPE (41.2%) and average RMSE (41.1%). Adding ARIMA to the ETS-ELM combination yielded a substantial gain in average MAPE (40.2%) and average RMSE (39.8%). These findings confirm that both ARIMA and ETS individually enhance the performance of the GA-tuned ARIMA-ETS-ELM model. ARIMA captures linear and autocorrelated trends, ETS captures exponential smoothing behaviour and trend-seasonality interactions, ELM models the non-linear relationships, and GA ensures the optimal configuration of the ELM layer.

Table 5
Performance comparison of the GA-tuned ARIMA-ETS-ELM model with its component models (Auto ARIMA, Auto ETS, and Auto ELM) based on out-of-sample forecasting over an 11-month horizon

Sl.	Type	Model	Metric				
			Average ME	Average RMSE	Average MAE	Average MPE	Average MAPE
1.	Component model	Auto ARIMA	-37.28	283.7	215.33	-0.43	3.13
2.		Auto ETS	41.80	176.96	142.43	0.73	2.03
3.		Auto ELM	-230.85	349.21	260.76	-3.40	3.87
4.	Full model	GA-tuned ARIMA-ETS-ELM	16.66	66.84	55.22	0.30	0.79

Table 6
Performance comparison of the GA-tuned ARIMA-ETS-ELM model with its two simplified variants based on out-of-sample forecasting over an 11-month horizon

Sl.	Type	Model	Metric				
			Average ME	Average RMSE	Average MAE	Average MPE	Average MAPE
1.	Simplified variants	GA-tuned ARIMA-ELM	92.01	113.56	96.08	1.28	1.34
2.		GA-tuned ETS-ELM	90.48	111.10	94.08	1.27	1.32
3.	Proposed model	GA-tuned ARIMA-ETS-ELM	16.66	66.84	55.22	0.30	0.79

Performance Comparison of GA-tuned ARIMA-ETS-ELM with Others

Table 7 presents the performance of the GA-tuned ARIMA-ETS-ELM against eight other standalone models (excluding its component models) in terms of evaluated metrics. Similarly, Table 8 exhibits the performance of the GA-tuned ARIMA-ETS-ELM against six other hybrid techniques, excluding the GA-tuned hybrids used for the ablation study. For cumulative data, a naïve drift model would produce near-linear extrapolation ($\hat{Y}_{t+h} = Y_t + h.\frac{Y_t - Y_1}{t-1}$). Given the strong linear trend ($R^2 = 0.976$), such a model would perform reasonably well, but it cannot capture nonlinear growth acceleration or irregular dynamics. The proposed hybrid model demonstrates substantial RMSE and MAPE improvements over simpler linear alternatives.

The GA-tuned ARIMA-ETS-ELM model outperformed all seven standalones, a generative-AI, and six hybrid models, achieving the lowest average scores for ME, RMSE, MAE, MPE, and MAPE (Tables 7 and 8).

Table 9 provides the percentage gain in performance achieved by the GA-tuned ARIMA-ETS-ELM hybrid model compared to the nineteen (ten standalone, one generative-AI, and eight hybrid) models, based on average MAPE and average RMSE over an 11-month horizon.

Table 7
Performance comparison of the GA-tuned ARIMA-ETS-ELM model with other standalone forecasting techniques based on out-of-sample forecasting over an 11-month horizon

Sl.	Type	Model	Metric				
			Average ME	Average RMSE	Average MAE	Average MPE	Average MAPE
1.	Statistical model	STL	-322.30	408.47	322.30	-4.87	4.87
2.	Statistical model	SES	-489.15	601.48	489.15	-7.79	7.79
3.	Statistical model	HLT	41.80	176.96	142.43	0.73	2.03
4.	Statistical model	HW	-476.23	578.10	476.23	-7.52	7.52
5.	AI-based model	NNAR	-470.40	581.47	470.40	-7.46	7.46
6.	AI-based model	MLP	-25.54	206.7	161.99	-0.24	2.32
7.	AI-based model	SVR	-846.70	923.88	846.70	-14.30	14.30
8.	Generative AI model	Moirai	-373.18	462.96	373.18	-5.72	5.72
9.	Proposed Model	GA-tuned ARIMA-ETS-ELM	16.66	66.84	55.22	0.30	0.79

Table 8
Performance comparison of the GA-tuned ARIMA-ETS-ELM model with other hybrid techniques based on out-of-sample forecasting over an 11-month horizon

Sl.	Author	Technique	Metric				
			Average ME	Average RMSE	Average MAE	Average MPE	Average MAPE
1.	(Hightower et al., 2024)	STL-ARIMA	-170.00	307.58	218.07	-2.45	3.20
2.	(Hightower et al., 2024)	STL-ETS	-132.81	237.71	165.24	-1.85	2.36
3.	(Talkhi et al., 2021)	ARIMA-ANN using CV-error	-158.31	320.78	229.63	-2.29	3.39
4.	(Perone, 2021)	ARIMA-ANN using equal weight	-253.62	379.87	273.89	-3.80	4.12
5.	(Perone, 2021)	ETS-ANN using equal weight	-219.54	317.41	235.03	-3.21	3.45
6.	(Perone, 2021)	ARIMA-ETS-ANN using equal weight	-154.06	283.77	198.89	-2.19	2.89
7.	Proposed model	GA-tuned ARIMA-ETS-ELM	16.66	66.84	55.22	0.30	0.79

It also showed significant improvements in average MAPE and RMSE compared to all nineteen models (Table 9). The performance gains are notable, ranging from approximately 40% to 95% for average MAPE and from about 39% to 93% for average RMSE, indicating a considerable enhancement in accuracy.

Table 9
Percentage performance gain of the GA-tuned ARIMA-ETS-ELM model over nineteen forecasting techniques

Sl.	Technique	Performance gain (%) achieved by GA-tuned ARIMA-ETS-ELM	
		Average MAPE	Average RMSE
1.	Auto ARIMA (Component model)	74.74	76.44
2.	Auto ETS (Component model)	61.17	62.23
3.	Auto ELM (Component model)	79.57	80.86
4.	STL	83.78	83.64
5.	SES	89.86	88.89
6.	HLT	61.17	62.23
7.	HW	89.49	88.44
8.	NNAR	89.41	88.51
9.	MLP	65.94	67.66
10.	SVR	94.48	92.77
11.	Moirai	86.41	85.72
12.	GA-tuned ARIMA-ELM	41.19	41.14
13.	GA-tuned ETS-ELM	40.22	39.84
14.	STL-ARIMA (Hightower et al., 2024)	75.30	78.27
15.	STL-ETS (Hightower et al., 2024)	66.52	71.88
16.	ARIMA-ANN using CV-error (Talkhi et al., 2021)	76.69	79.16
17.	ARIMA-ANN using equal weight (Perone, 2021)	80.82	82.40
18.	ETS-ANN using equal weight (Perone, 2021)	77.13	78.94
19.	ARIMA-ETS-ANN using equal weight (Perone, 2021)	72.68	76.45

The results underscore the robust out-of-sample predictive capability of the proposed GA-tuned ARIMA-ETS-ELM approach when evaluated against traditional statistical models, contemporary generative AI methodologies, and other hybrid techniques.

Statistical Significance Testing of the GA-tuned ARIMA-ETS-ELM

The GA-tuned ARIMA-ETS-ELM approach outperformed all nineteen other techniques and was identified as the best candidate model for out-of-sample prediction of cumulative dengue cases in the USA. Statistical significance testing was conducted to assess the robustness of the forecasting performance achieved by the GA-tuned ARIMA-ETS-ELM approach. Table 10 presents the results. Both tests support the robustness of the GA-tuned ARIMA-ETS-ELM model. The Mann-Whitney U test indicates that both the distributions (forecasted values and the observed data) are close with no significant difference. The Wilcoxon Signed Rank test confirms that the model exhibits a low median forecast MAPE (≤ 1) for this 11-month forecast horizon. Therefore, these statistical pieces of evidence strengthen the GA-tuned ARIMA-ETS-ELM model's forecasting accuracy and reliability.

Table 10
Statistical significance testing of forecasting performance of the GA-tuned ARIMA-ETS-ELM model

Statistical test	Null hypothesis	Statistic	Significance	Decision
Mann-Whitney U test	The distribution of observed data and GA-tuned ARIMA-ETS-ELM forecasts is the same	61	1.00	Fail to reject the null hypothesis
Wilcoxon Signed-Rank test	Median MAPE of GA-tuned ARIMA-ETS-ELM for the forecasting horizon of 11 months ≤ 1	6.50	.9919	Fail to reject the null hypothesis

Statistical tests were applied to the 11-step forecast horizon to evaluate distributional alignment and median MAPE constraints. Nonparametric tests (Mann-Whitney/Wilcoxon) were used because forecast errors may not follow a normal distribution, the sample size is small (n = 11), and the focus was on distributional similarity and median constraints; the Diebold–Mariano test could be applied in future work for formal pairwise predictive accuracy comparison.

Theoretical Implications of the Research Findings

The theoretical implications of the findings from the present study are outlined as follows:

- Model integration: Auto ARIMA handles linear and autocorrelated patterns, while Auto ETS models trend and seasonality through exponential smoothing. ELM addresses non-linear relationships, which contribute strengths to the GA-tuned ARIMA-ETS-ELM approach.
- Evolutionary optimisation: The current work also explores GA optimisation, revealing that it enhances the model by optimally configuring the ELM component and improving forecast accuracy, emphasising evolutionary optimisation's importance in hybrid systems.
- Theoretical merit: The GA-tuned ARIMA-ETS-ELM performed better than several other models, including classical models, neural networks, and generative AI (e.g., Moirai). Statistical significance testing (Mann-Whitney U and Wilcoxon Signed Rank tests) highlights the reliability of the hybrid model, supporting its theoretical soundness and merit.
- Contribution to forecasting theory: The current study enhances the theoretical basis of hybrid forecasting by demonstrating that structured model integration and effective optimisation lead to greater accuracy in practical applications such as epidemiological surveillance.

The present study employed cumulative values of the monthly dengue cases to capture the long-term progression and growth pattern of dengue. This cumulative (monthly) data

smoothed the irregularities caused by underreporting or short-term volatility in the data. This helps to improve the model's stability. Further aid in its ability to identify structural trends. Due to larger denominators, the cumulative data may tend to produce a lower MAPE. These results primarily reflect trend-level forecast accuracy. The proposed GA-tuned ARIMA-ETS-ELM framework can be equivalently applied to non-cumulative (monthly) dengue cases for assessing short-term predictive accuracy in future studies.

Practical Implications of the Research Findings

The multifaceted empirical essences of the present work are as follows:

- Enhanced forecasting accuracy for dengue surveillance: The model demonstrates low MAPE and RMSE across various forecast horizons. It facilitates accurate early prediction of dengue outbreaks, assisting public health authorities with timely resource allocation, intervention planning, and designing awareness campaigns.
- Real-World applications: The model is quite flexible and able to capture linear as well as seasonal and nonlinear patterns that make it suitable for a wide range of health-related time series problems. Performance of the proposed model has been rigorously validated using nonparametric statistical tests, which include the Mann-Whitney U test and the Wilcoxon Signed-Rank test. Therefore, the presented model can be well-applied to the settings where the data is noisy, incomplete, or uncertain.
- Model optimisation: There is an automated and efficient model optimisation. The ELM component is automatically tuned by the GA, which minimises the need to use manual parameter selection or expert assistance. This, in its turn, makes the model easier to use in government or institutional environments.
- Generalise: The GA-tuned ARIMA-ETS-ELM model can be used to forecast dengue and can be modified to suit other infectious diseases or epidemiology-related prediction tasks. It presents a flexible model which integrates both statistical and AI-enhanced predictive approaches to make efficient real-life decisions.

The present study used monthly dengue data of the United States from the World Health Organisation (2025) dataset. The dataset includes country-wise dengue data for multiple countries. The proposed GA-tuned ARIMA-ETS-ELM framework is practically adaptable and can be extended to other countries exhibiting distinct transmission dynamics.

The proposed GA-tuned ARIMA-ETS-ELM model can act as an early-warning tool for public health authorities. In case of any potential/future outbreak, the monthly forecasts can assist in planning vector control measures, allocating medical supplies, and preparing hospital resources in advance. The same framework can be extended to other infectious diseases that exhibit similar temporal dynamics.

Limitations and Future Direction

The research studies have established many factors that play a significant role in the dynamics of disease outbreaks that can significantly deteriorate the level of performance of the forecasting model, which include selective reporting, hetero-spatial conditions, and sensitivity in the estimation of the parameters. The study acknowledges possible reporting variability. The robustness is addressed indirectly through hybrid modelling that captures both linear and nonlinear shifts, nonparametric statistical validation, and the use of cumulative transformation. These reduce abrupt reporting spikes. Explicit structural break detection (e.g., Chow test, Bai–Perron) was not incorporated in the current work and is suggested for future work. The current paper concentrated on the development of the model against the USA dataset, in which the proposed model is prepared by using lagged values and fitted values of statistical models as inputs. These inputs might not be a complete capture of the external factors like weather, mobility or demographic data that affect the dengue spread. The research centred on the dengue cases, which were cumulative monthly. The current research did not investigate the fine-grained prediction (e.g., of week or city). This study focuses on point forecasts. The current work presented distributional visualisations (density and violin plots), but formal prediction intervals are not constructed. Incorporating bootstrapped or quantile-based interval estimation within the ELM framework is an important direction for future policy-grade decision support. Irrespective of such constraints, the study can lead to a hint in making decisions on public health. There are several directions of the research. The current framework is structurally extensible: ARIMA and ETS components can incorporate exogenous regressors (ARIMAX), and ELM can include additional input features such as climate, mobility, or vector indices, which are expected to enhance short-term predictive responsiveness and mechanistic interpretability. We can add exogenous variables (e.g., climate indicators, mosquito populations, rainfall) to the ARIMA-ETS-ELM model to investigate its prediction improvement. To identify the model's merit and generality, researchers may test this hybrid model on other regions' data. This will help to assess its usefulness and generalisation. The ARIMA-ETS-ELM can be refined with other sophisticated evolutionary algorithms (e.g., Particle Swarm Optimisation and Differential Evolution). This model can be adapted to create early warning systems for a real-time decision-support system that affects the health of the population.

CONCLUSION

Dengue adversely impacts millions worldwide. Several significant contributions to the present research frontier have been accomplished. Both traditional and AI-driven forecasting models illustrate the progression of Dengue, but they possess inherent limitations. These include challenges in identifying complex non-linear associations and issues related to the interpretability of the models.

To balance the trade-offs, the proposed model incorporates the GA-tuned ARIMA-ETS-ELM method for automatically forecasting cumulative dengue cases in the USA, wherein the ELM architecture within the hybrid model is optimised using a GA. It identifies the optimal no. of hidden neurons in the ELM. While GAs is primarily used for multi-objective optimisation, this study demonstrates their effectiveness in tuning the ELM to enhance forecasting accuracy. The present study thoroughly evaluates the proposed GA-tuned ARIMA-ETS-ELM model against nineteen leading forecasting methodologies. This comparison encompasses ten standalone, one generative-AI, and eight hybrid approaches. An ablation study was conducted on the GA-tuned ARIMA-ETS-ELM model to isolate the contributions of the ARIMA and ETS components. The GA-tuned ARIMA-ETS-ELM model is identified as the most effective method among the assessed techniques for forecasting cumulative dengue cases in the United States. The performance of this model was further validated through statistical significance tests, confirming its efficacy. The GA-tuned ARIMA-ETS-ELM model demonstrates significant forecasting accuracy, achieving an average ME of 16.66, a RMSE of 66.84, a MAE of 55.22, a MPE of 0.30, and a MAPE of 0.79. The metrics clearly illustrate that the performance is well within acceptable limits, providing strong evidence of the proposed model's innovative nature and retaining its prospect of making a substantial contribution to the field of study.

ACKNOWLEDGEMENT

The authors sincerely thank the editor and anonymous referees for their constructive comments and suggestions, which greatly improved the quality of this paper.

LIST OF ABBREVIATIONS

ADF	:	Augmented Dickey-Fuller
ANN	:	Artificial neural network
ARIMA	:	AutoRegressive Integrated Moving Average
BDS	:	Brock-Dechert-Scheinkman
CV	:	Cross-validation
ELM	:	Extreme learning machine
ETS	:	Error, trend, seasonal
GA	:	Genetic algorithm
HLT	:	Holt's linear trend
HW	:	Holt-Winters
MAE	:	Mean absolute error
MAPE	:	Mean absolute percentage error
ME	:	Mean error
MLP	:	Multilayer perceptron
MPE	:	Mean percentage error
NNAR	:	Neural network AutoRegressive

RMSE	:	Root mean square error
SES	:	Simple exponential smoothing
STL	:	Seasonal-trend decomposition using LOESS
SVR	:	Support vector regression
USA	:	United States of America

REFERENCES

Ahmadoni, N. I. (2025). The role of cybersecurity governance in activating the disclosure of cybersecurity risk management and its impact on improving financial performance: An experimental study on banks listed on the Egyptian Stock Exchange. *Alexandria Journal of Accounting Research, 9*(1), 1-71. https://doi.org/10.21608/aljalexu.2025.397533

Ahmar, A. S., & Boj, E. (2021). Application of neural network time series (NNAR) and ARIMA to forecast infection fatality rate (IFR) of COVID-19 in Brazil. *JOIV: International Journal on Informatics Visualisation, 5*(1), 8-10. https://doi.org/10.30630/joiv.5.1.372

Aksu, T., Woo, G., Liu, J., Liu, X., Liu, C., Savarese, S., Xiong, C., & Sahoo, D. (2024). *GIFT-Eval: A benchmark for general time series forecasting model evaluation* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2410.10393

Ariyanto, E., & Nugraha, M. K. R. (2024). Population vs. poverty level in the future in Indonesia: Holt's linear trend method. *International Journal of Sustainable Development and Planning, 19*(7), 2763-2770. https://doi.org/10.18280/ijsdp.190733

Baker, Q. B., Faraj, D., & Alguzo, A. (2021). Forecasting dengue fever using machine learning regression techniques. In *Proceedings of the 2021 12th International Conference on Information and Communication Systems (ICICS)* (pp. 157-163). IEEE. https://doi.org/10.1109/ICICS52457.2021.9464619

Bunprom, P., Siramaneerat, I., & Bhumkittipich, P. (2026). Forecasting dengue fever incidence in Thailand using ARIMA: Implications for public health planning. *Environment and Natural Resources Journal, 24*(1), 88-97. https://doi.org/10.32526/ennrj/24/20250069

Chakraborty, A., Das, D., Mitra, S., De, D., & Pal, A. J. (2024). Forecasting adversities of COVID-19 waves in India using intelligent computing. *Innovations in Systems and Software Engineering, 20*, 821-837. https://doi.org/10.1007/s11334-022-00486-y

Das, D., & Chakrabarti, S. (2021). Forecasting non-linear WPI of manufacture of chemicals and chemical products in India: An MLP approach. *International Journal of Advanced Technology and Engineering Exploration, 8*(82), 1193-1203. https://doi.org/10.19101/IJATEE.2021.874407

Dhaked, D., Sharma, O., Gopal, Y., & Gopal, D. (2025). Prediction of dengue patients using deep learning methods amid complex weather conditions in Jaipur, India. *Discover Public Health, 22*, Article 48. https://doi.org/10.1186/s12982-025-00448-2

Flores, A., Tito-Chura, H., Guzman-Valdivia, J., Morales-Gonzales, R., Rosado-Chavez, C., & Silva-Delgado, C. (2026). Multi-step forecasting of weekly dengue cases using data augmentation and convolutional recurrent neural networks. *Computers, 15*(2), Article 114. https://doi.org/10.3390/computers15020114

Hasan, P., Khan, T. D., Alam, I., & Haque, M. E. (2024). Dengue in tomorrow: Predictive insights from ARIMA and SARIMA models in Bangladesh: A time series analysis. *Health Science Reports, 7*(12), Article e70276. https://doi.org/10.1002/hsr2.70276

Hightower, A., Ziedan, A., Guo, J., Zhu, X., & Brakewood, C. (2024). A comparison of time series methods for post-COVID transit ridership forecasting. *Journal of Public Transportation, 26*, Article 100097. https://doi.org/10.1016/j.jpubtr.2024.100097

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1-22. https://doi.org/10.18637/jss.v027.i03

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeen, F. (2024). *forecast: Forecasting functions for time series and linear models* (Version 8.23.0) [R package]. https://pkg.robjhyndman.com/forecast/

Ivanovski, Z., & Ivanovska, N. (2024). The augmented Dickey–Fuller test for the stationarity of the final public consumption and GDP time series of the Republic of North Macedonia. *UTMS Journal of Economics, 15*(2), 109-124.

Jirasakuldech, B., Emekter, R., & Bui, T. (2023). Non-linear structures, chaos, and bubbles in US regional housing markets. *Journal of Economics and Finance, 47*(1), 63-93. https://doi.org/10.1007/s12197-022-09598-4

Kourentzes, N. (2023). *nnfor: Time series forecasting with neural networks* (Version 0.9.9) [R package]. https://cran.r-project.org/package=nnfor

Li, K. (2024). Prediction of atmospheric temperature in Sacramento area based on ARIMA and ETS models. *Theoretical and Natural Science, 51*, 165-171. https://doi.org/10.54254/2753-8818/51/2024CH0197

Liu, X., Liu, J., Woo, G., Aksu, T., Liang, Y., Zimmermann, R., Liu, C., Savarese, S., Xiong, C., & Sahoo, D. (2024). *Moirai-MoE: Empowering time series foundation models with sparse mixture of experts* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2410.10469

Lu, X., Teh, S. Y., Tay, C. J., Abu Kassim, N. F., Fam, P. S., & Soewono, E. (2024). Application of multiple linear regression model and long short-term memory with compartmental model to forecast dengue cases in Selangor, Malaysia based on climate variables. *Infectious Disease Modelling, 10*(1), 240-256. https://doi.org/10.1016/j.idm.2024.10.007

Mahanty, C., Patro, S. G. K., Rathor, S., Rachapudi, V., Muzammil, K., Islam, S., Razak, A., & Khan, W. A. (2024). Forecasting of coronavirus active cases by utilising logistic growth model and fuzzy time series techniques. *Scientific Reports, 14*, Article 18039. https://doi.org/10.1038/s41598-024-67161-z

Mendoza Chirinos, J. (2024). *Spatio-temporal prediction of dengue outbreaks using neural networks with climatic and socioeconomic factors in Colombian municipalities* [Master's thesis, Universidad de los Andes]. https://hdl.handle.net/1992/75226

Mobin, M. A. (2024). Forecasting dengue in Bangladesh using meteorological variables with a novel feature selection approach. *Scientific Reports, 14*, Article 32073. https://doi.org/10.1038/s41598-024-83770-0

Mohammed, N., & Rahman, M. (2023). Forecasting dengue incidence in Bangladesh using seasonal ARIMA model: A time series analysis. In M. S. Satu, M. A. Moni, M. S. Kaiser, & M. S. Arefin (Eds.), *Machine intelligence and emerging technologies* (Vol. 491). Springer. https://doi.org/10.1007/978-3-031-34622-4_47

Nwogu, E. C., Iwueze, I. S., & Nlebedim, V. U. (2016). Some tests for seasonality in time series data. *Journal of Modern Applied Statistical Methods, 15*(2), Article 24. https://doi.org/10.22237/jmasm/1478002920

Parreño, S. J. E., & Anter, M. C. J. (2024). New approach for forecasting rice and corn production in the Philippines through machine learning models. *Multidisciplinary Science Journal, 6*(9), Article 2024168. https://doi.org/10.31893/multiscience.2024168

Patil, S., & Pandya, S. (2021). Forecasting dengue hotspots associated with variation in meteorological parameters using regression and time series models. *Frontiers in Public Health, 9*, Article 798034. https://doi.org/10.3389/fpubh.2021.798034

Pereira da Veiga, C., Pereira da Veiga, C. R., Girotto, F. M., Marconatto, D. A. B., & Su, Z. (2024). Implementation of the ARIMA model for the prediction of economic variables: Evidence from the health sector in Brazil. *Humanities and Social Sciences Communications, 11*, Article 1068. https://doi.org/10.1057/s41599-024-03023-3

Perone, G. (2021). Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalisations in Italy. *The European Journal of Health Economics, 23*, 917-940. https://doi.org/10.1007/s10198-021-01347-4

Python Software Foundation. (2025). *The Python language reference* (Version 3.11.12). https://docs.python.org/3/reference/

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahman, M., Rashid, F., Roy, S. K., & Habib, M. A. (2024). Application of extreme learning machine (ELM) forecasting model on CO_2 emission dataset of a natural gas-fired power plant in Dhaka, Bangladesh. *Data in Brief, 54*, Article 110491. https://doi.org/10.1016/j.dib.2024.110491

Rangarajan, P., Mody, S. K., & Marathe, M. (2019). Forecasting dengue and influenza incidences using a sparse representation of Google Trends, electronic health records, and time series data. *PLOS Computational Biology, 15*(11), Article e1007518. https://doi.org/10.1371/journal.pcbi.1007518

Ray, M., Barman, S., Sinha, K., & Singh, K. N. (2022). *TSSVM: Time series forecasting using SVM model* (Version 0.1.0) [R package]. https://doi.org/10.32614/CRAN.package.TSSVM

Saleh, A. Y., & Baiwei, L. (2021). Dengue prediction using deep learning with long short-term memory. In *Proceedings of the 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)* (pp. 1-5). IEEE. https://doi.org/10.1109/eSmarTA52612.2021.9515734

Sanchez-Gendriz, I., de Souza, G. F., de Andrade, I. G. M., Neto, A. D. D., de Medeiros Tavares, A., Barros, D. M. S., de Morais, A. H. F., Galvão-Lima, L. J., & de Medeiros Valentim, R. A. (2022). Data-driven computational intelligence applied to dengue outbreak forecasting: A case study at the scale of the city of Natal, RN, Brazil. *Scientific Reports, 12*(1), Article 6550. https://doi.org/10.1038/s41598-022-10512-5

Santosa, R. G., Chrismanto, A. R., Raharjo, W. S., & Lukito, Y. (2024). LQ45 stock price forecasting: A comparison study of ARIMA (p, d, q) and Holt–Winter method. *International Journal of Information Technology and Computer Science Applications, 2*(2), 115-129. https://doi.org/10.58776/ijitcsa.v2i2.150

Saragih, S. A., Munar, A., & Hasibuan, W. R. (2024). Forecasting the amount of corn production in North Sumatra based on 2017–2021 data using the single and double exponential smoothing method (case study of Central Bureau of Statistics of North Sumatra). *Journal of Artificial Intelligence and Engineering Applications, 3*(2), 614-617. https://doi.org/10.59934/jaiea.v3i2.449

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software, 53*(4), 1-37. https://doi.org/10.18637/jss.v053.i04

Scrucca, L. (2017). On some extensions to GA package: Hybrid optimisation, parallelisation and islands evolution. *The R Journal, 9*(1), 187-206. https://doi.org/10.32614/RJ-2017-008

Shaub, D., & Ellis, P. (2020). *forecastHybrid: Convenient functions for ensemble time series forecasts* (Version 5.0.19) [R package]. https://doi.org/10.1057/jors.1969.103

Sucipto, C. S., Sulandari, W., & Susanti, Y. (2025). Implementation of cross-validation on Hang Seng Index forecasting using Holt's exponential smoothing and Auto-ARIMA method. *BAREKENG: Jurnal Ilmu Matematika dan Terapan, 19*(1), 13-24. https://doi.org/10.30598/barekengvol19iss1pp13-24

Sutriyawan, A., Martini, M., Sutiningsih, D., Agushybana, F., Wahyuningsih, N. E., Adamu, V. E. V., Akbar, H., & Aba, M. (2024). Time series analysis of dengue incidence in Bandung City, Indonesia using an ARIMA model. *Journal of Microbiology, Epidemiology and Immunobiology, 101*(6), 803-811. https://doi.org/10.36233/0372-9311-570

Talkhi, N., Fatemi, N. A., Ataei, Z., & Nooghabi, M. J. (2021). Modelling and forecasting number of confirmed and deaths caused by COVID-19 in Iran: A comparison of time series forecasting methods. *Biomedical Signal Processing and Control, 66*, Article 102494. https://doi.org/10.1016/j.bspc.2021.102494

Tomov, L., Chervenkov, L., Miteva, D. G., Batselova, H., & Velikova, T. (2023). Applications of time series analysis in epidemiology: Literature review and our experience during the COVID-19 pandemic. *World Journal of Clinical Cases, 11*(29), 6974-6983. https://doi.org/10.12998/wjcc.v11.i29.6974

Torres, I., Nakano, M., Cime-Castillo, J. A., Escamilla-Hernandez, E., Lopez-Garcia, O., & Lanz Mendoza, H. (2025). Dengue-infected mosquito detection with uncertainty evaluation based on Monte Carlo dropout. *IEEE Latin America Transactions, 23*(2), 135-143. https://doi.org/10.1109/TLA.2025.10851361

Trapletti, A., & Hornik, K. (2024). *tseries: Time series analysis and computational finance* (Version 0.10-58) [R package]. https://doi.org/10.32614/CRAN.package.tseries

Wang, G., Zhuang, L., Mo, L., Yi, X., Wu, P., & Wu, X. (2023). BAG: A linear-nonlinear hybrid time series prediction model for soil moisture. *Agriculture, 13*(2), Article 379. https://doi.org/10.3390/agriculture13020379

World Health Organisation. (2025). *Global dengue surveillance*. Retrieved January 2025, from https://worldhealthorg.shinyapps.io/dengue_global/

Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., & Sahoo, D. (2024). Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)* (Vol. 235, pp. 53140-53164). PMLR. https://ink.library.smu.edu.sg/sis_research/9906

Wu, Z., Zhou, J., & Yu, X. (2025). Forecast natural gas price by an extreme learning machine framework based on a multi-strategy grey wolf optimiser and signal decomposition. *Sustainability, 17*(12), Article 5249. https://doi.org/10.3390/su17125249

Xiong, Y., Ma, Y., Ruan, L., Li, D., Lu, C., Huang, L., & National Traditional Chinese Medicine Medical Team. (2022). Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious Diseases of Poverty, 11*(1), Article 19. https://doi.org/10.1186/s40249-022-00946-4

Yadav, V., & Nath, S. (2019). Novel hybrid model for daily prediction of PM10 using principal component analysis and artificial neural network. *International Journal of Environmental Science and Technology, 16*(6), 2839-2848. https://doi.org/10.1007/s13762-018-1999-x

Yunisa, F. A., & Siregar, M. A. P. (2023). The application of seasonal trend decomposition using LOESS for export forecasting by economic commodity group in North Sumatra. *ZERO: Jurnal Sains, Matematika, dan Terapan, 7*(1), 61-67. https://doi.org/10.30829/zero.v7i1.17341

Pertanika J. Sci. & Technol. 34 (2): 1005 - 1024 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Review Article

Visualising Global Trends in Soil Infiltration Study by using Bibliometric Analysis

Nor Farah Atiqah Ahmad*, Siti Nooraiin Mohd Razali, Suhaila Sahat, and Muhammad Azraie Abdul Kadir

Department of Civil Engineering, Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, 84600 Pagoh, Johor, Malaysia

ABSTRACT

Soil infiltration plays a crucial role in the hydrological cycle, water resource management, and sustainable agricultural practices. It is important to identify the emerging trends and research gaps for this study. Therefore, this study aims to present a comprehensive and systematic academic review of soil infiltration in the field using bibliometric analysis with the aid of VOSviewer to evaluate global research trends in soil infiltration based on author keywords, affiliated countries, and co-authorships. This study retrieved 494 articles published between 2000 and 2025, as of January 16th, 2025, from the Scopus database. Results have shown a steady decrease and increased for annual number of publications, with the lowest drop in 2024 with 22 publications and the highest rise in 2021 with 30 publications. The leading contributors are from China and the USA, leading by 28% and 21% respectively. Meanwhile, Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering contributed the most publications. Although the Science of The Total Environment (STOTEN) journal ranked second after Water (Switzerland) in the top journal, the CiteScore 2023 for STOTEN is the highest with 17.6. This study underscores the effectiveness of a water management system, which includes it supports for flood risk mitigation, enhances agricultural productivity and improves groundwater recharge. Future research should focus on optimising the soil infiltration model by adapting AI tools to consider climate change factors

Keywords: Bibliometric analysis, hydrology, research trends, scopus database, soil infiltration, VOSViewer

ARTICLE INFO

Article history:
Received: 01 August 2025
Accepted: 16 March 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.17

E-mail addresses:
norfarah@uthm.edu.my (Nor Farah Atiqah Ahmad)
snooraiin@uthm.edu.my (Siti Nooraiin Mohd Razali)
suhailasa@uthm.edu.my (Suhaila Sahat)
mazraie@uthm.edu.my (Muhammad Azraie Abdul Kadir)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

In the study of the hydrological cycle and environmental science, soil water infiltration is a crucial process as it determines how the water enters and moves through the soil. The process influences groundwater recharge, surface runoff and soil moisture availability, which not only has an impact on agriculture, but it is also a key study for urban drainage and ecosystem sustainability. An adequate knowledge of soil infiltration rate is essential to ensure the effectiveness of water resource management, erosion control and flood mitigation. The factors such as soil texture, moisture, vegetation cover and land use directly influenced the soil infiltration capacity.

Bibliometrics is one of the qualitative analyses that is used to reduce the number of unrelated articles in a scholar's research area (Donohue, 1972). It can build firm foundations that enable scholars to gain an overview, identify research gaps, develop new, novel ideas for their investigation and focus on the aim of their contribution to the field (Donthu et al., 2021). Bibliometric analysis has gained immense popularity in various fields, including medicine, social science, computer science, engineering, pharmacology, neuroscience, and psychology. As of January 2025, a total of 31,536 documents focused on bibliometric analysis. The accessibility of bibliometric software tools such as Leximancer, Gephi, Bibexcel, Sci2, SciMat, Pajek, UCINET, CiteSpace and VOSViewer might have an impact on bibliometric analysis popularity. The most popular tool is the VOSViewer software, which has gained 7,797 documents recorded in the Scopus database. Visualisation of similarities (VOS) was first introduced by (Van Eck et al., 2006) as a new method to provide visualisations of the insight into the subfield relationship to another subfield, such as map co-occurrence, citations and authorship patterns. This advantage helps authors conducting research by identifying influential authors, highly cited papers and identifying research gaps quickly. Using VOSviewer, which has a user-friendly interface, free accessibility, and compatibility with Scopus and Web of Science databases, makes it valuable for authors with limited resources. The goal of this study was as follows: i) to examine journal article temporal patterns for soil infiltration, ii) to present the contributions of prominent authors, top countries, and most prolific academic institutions, and iii) to identify the dominant country based on major applications. This study is beneficial to academics, policymakers, as well as individuals who examine soil infiltration research patterns and discuss the current and possibilities for future research.

METHODS

This section explained the sequential process of data collection and analysis regarding soil infiltration extracted from the Scopus database. The procedure comprised outlining the source of data and search strategy, the screening process to choose relevant documents and the criteria for selecting the final dataset for bibliometric analysis by using VOSviewer.

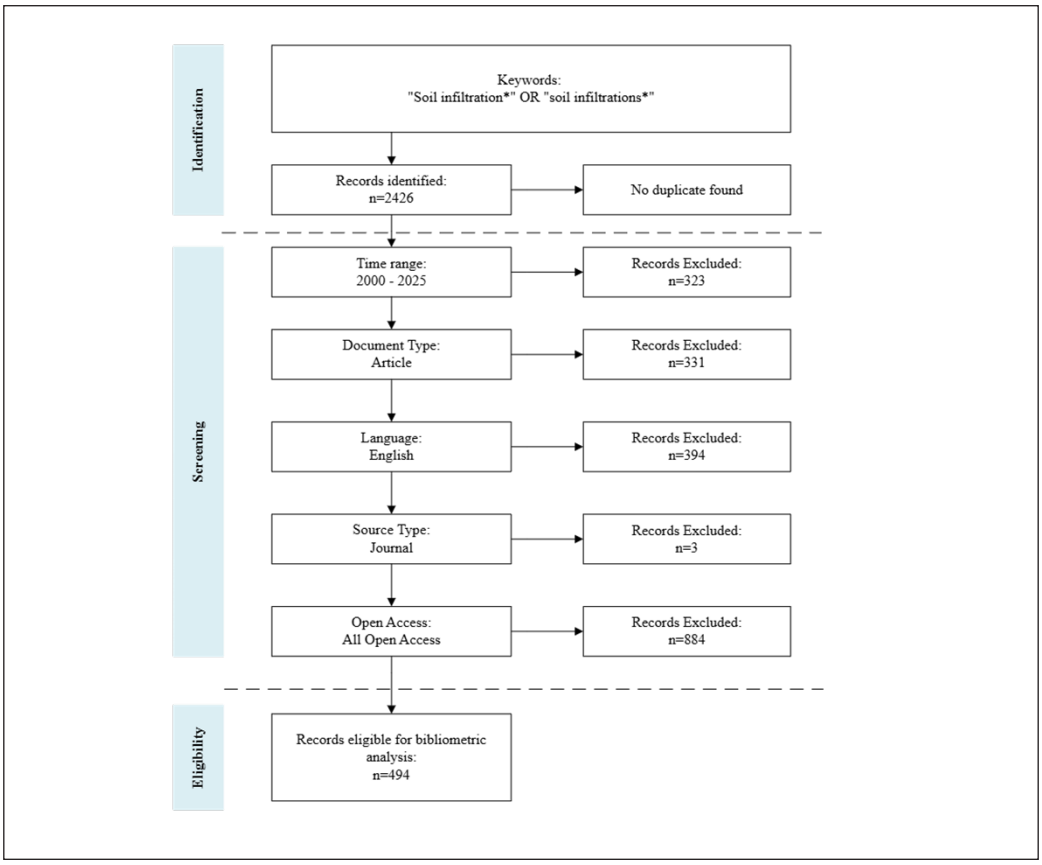


Figure 1. Systematic process of research stages and selection

The formation of a thematic cluster in VOSviewer is influenced by parameters such as keyword thresholds and normalisation methods. Although different parameter settings may slightly modify cluster boundaries, the major thematic structures identified in this study remained generally consistent.

Data Source and Search Strategy

Data mining was conducted on January 16th, 2025, using the search query: ("soil infiltration*" OR "soil infiltration*") using the Scopus database. The main theme in this study was articles containing "soil* infiltration*" in the title and abstract. The initial search retrieved a total of 2426 documents with the earliest publication dates to 1936, and more recent ones are from 2025. The search strategy focused on the keyword "soil infiltration" in the title and abstract to ensure the retrieval of studies directly related to the core topic. It is acknowledged that the other related terms, such as infiltration capacity, infiltration modelling and soil hydraulic processes, may capture additional relevant publications that

were not included in this dataset. After the screening process, as shown in Figure 1, 494 documents remained for deeper analysis, which includes duplicate keywords: singular and plural forms. This step is done by using the OpenRefine application. To focus on the relevance, the time range was set from 2000 to 2025, which excluded 323 articles. After considering papers that were classified as articles and published only in English journals, 728 documents were excluded. Consequently, 494 articles were considered suitable for the soil infiltration bibliometric analysis.

Screening Data

To ensure the only relevant and high-quality studies were included, the final stage of this study was data extraction and filtering. A total of 494 articles containing their citations, bibliographical and author keywords information were exported as comma-separated values(.csv) file to VOSviewer software (version 1.6.20). The object of interest known as items are keywords, country, authors, and co-authorship. Based on the VOSviewer user manual by (Eck & Waltman, 2023). A link that connects two objects is unique, as it cannot be shared with any other pair of objects. Only positive numerical number denotes the strength for each link, and it increases as the strength increases.

RESULTS AND DISCUSSION

Publications Output and Trend

The annual publications trends in soil infiltration from January 2000 to January 2025 are shown in Figure 2 shows an eminent increase in research interest throughout this period. The data indicates that from 2000 to 2018, the annual number of publications fluctuates and on average the increment of 5 publications annually during this period. This may be influenced by several factors beyond technological developments, such as increasing global concern regarding climate change, water resources management and sustainable agricultural practices. Furthermore, the variations may reflect in research funding priorities, environmental challenges, and extreme weather. From 2020 onwards, a prominent upward trend in 2020 the number of publications was observed, in 2020 jumped to 57 publications compared to the previous year. The increasing of interest can be linked to the emerging Artificial Intelligence (AI) applications. The AI applications allow researchers to conduct various applications, such as soil infiltration models and simulations. It is crucial to point out that most of the articles require payment to access the material, and if the publication is in an open-access journal, it is likely to receive a larger number of citations. Overall, 1375 publications on soil infiltration were published during the 2000 to 2025 period, and 35% of the publications are open access type.

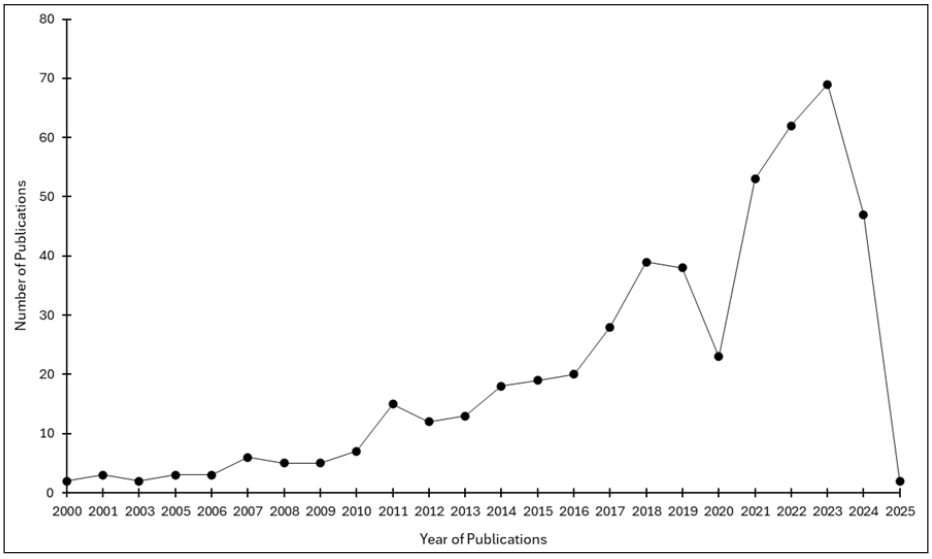


Figure 2. Number of publications and publication trend in the soil infiltration study for the last 24 years until January 16th, 2025

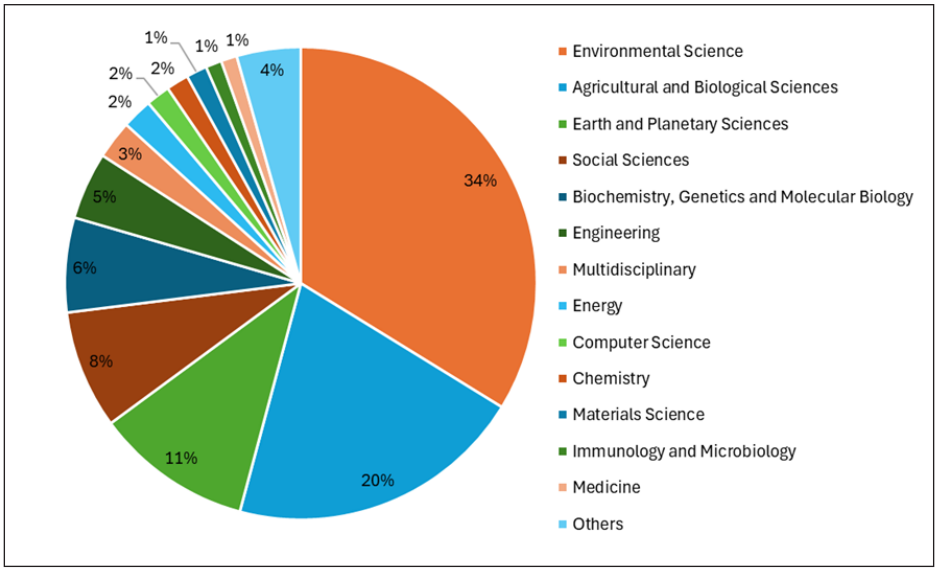


Figure 3. Distribution percentage by subject area for soil infiltration

Figure 3 demonstrates the distribution of documents by subject area regarding soil infiltration. The most representative areas for this research area are Environmental Science (308 articles), Agricultural and Biological Sciences (186 articles), Earth and Planetary Sciences (98 articles) and Social Sciences (74 articles). A total of 40 articles

were classified under Others: chemical engineering, pharmacology, toxicology, and pharmaceutics, among others. Indeed, soil infiltration is a multidisciplinary area, as one of the publications was categorised under Art and Humanities. Although it is like the rest of the soil infiltration experiments, Sittisom et al. (2022) researched designing bioretention for urban areas in Chiang Mai, Thailand, to increase the efficiency of soil infiltration in urban areas. This approach can be used to enhance the enforcement of policymakers for future development.

Results also showed that a total of 8 different languages were used to publish studies regarding soil infiltration from 2000 to 2025. English (494; 94%) was the commonly used language, followed by Portuguese (16; 3%) and Spanish (8; 1%). Other languages (10; 2%) include French, Chinese, Turkish, Japanese and Czech. Nevertheless, when a scholar submits an article in a foreign language, it is compulsory to have an English title and abstract.

Preferred Journal

Table 1 shows the top 10 most productive journals for the hydrology field. Three journals were published by the Multidisciplinary Digital Publishing Institute (MDPI) and Elsevier, respectively. Others were published by the Public Library of Science, Copernicus Publications, IWA Publishing and John Wiley & Sons. Water Switzerland, with 37 articles covering 7.5% of the total publications, has been identified as the most productive journal, followed by Science of The Total Environment (STOTEN) with 20 articles (4.0%), PLoS ONE (16, 3.2%) and Sustainability Switzerland (13, 2.6%). Although Hydrology and Earth System Sciences ranked 5th, their article ranked 2nd in highest citations with 145 after STOTEN's with 165 citations. CiteScore plays a significant role among authors in journal selection as it is the metric for measuring the annual average number of citations to the articles published in the journal. Based on the CiteScore 2023 report, eight journals had a CiteScore of 6 and above. Journals of the highest and lowest CiteScore belonged to STOTEN (17.6) and Water Science and Technology (4.9), respectively. Water Science and Technology journal ranked 6th with 11 publications in Scopus; the total citation and CiteScore were significantly lower compared to other journals. This was likely due to the primary scope of this journal being the technology of wastewater and stormwater management.

Cumulatively, over 2424 articles were published among 170 journals on soil infiltration in which 5 from the top 10 were published by Elsevier, as tabulated in Table 2. Although the most productive journal was Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, led by 0.1% compared to STOTEN with 3.4%, the CiteScore of the STOTEN journal is the highest, with 17.6 among the top 10. Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering

Table 1
Top 10 productive journals on soil infiltration study

	Journal	TP (%)	TP (%)	TC	CiteScore 2023	The most cited article	Times cited	Publisher
1	Water (Switzerland)	37 (7.5)	37 (1.7)	90,757	5.8	Effects of rainfall intensity and slope gradient on runoff and soil moisture content on different growing stages of spring maize (Mu et al., 2015).	103	Multidisciplinary Digital Publishing Institute (MDPI)
2	Science of the Total Environment	20 (4.0)	20 (3.4)	550,528	17.6	Pollution pathways and release estimation of perfluoro octane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) in central and eastern China (Liu et al., 2017).	161	Elsevier
3	PLoS ONE	16 (3.2)	16	331,118	6.2	Comparing infiltration rates in soils managed with conventional and alternative farming methods: A meta-analysis (Basche & DeLonge, 2019).	142	Public Library of Science
4	Sustainability (Switzerland)	13 (2.6)	13	432,964	6.8	Effects of super-absorbent polymer on soil remediation and crop growth in arid and semi-arid areas (Yang et al., 2020).	49	Multidisciplinary Digital Publishing Institute (MDPI)
5	Hydrology and Earth System Sciences	12 (2.4)	12	12,167	10.1	Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding (Skougaard Kaspersen et al., 2017).	145	Copernicus Publications
6	Water Science and Technology	11 (2.2)	11	9,579	4.9	A simplified approach for the design of infiltration trenches (Campisano et al., 2011).	21	IWA Publishing
7	Agronomy	11 (2.2)	11	76,029	6.2	Effect of bacterial manure on the properties of complex soil and growth of ryegrass (Zhang et al., 2021).	7	Multidisciplinary Digital Publishing Institute (MDPI)
8	Journal of Hydrology	10 (2.0)	10	64,858	11.0	A novel approach to model dynamic flow interactions between the storm sewer system and the overland surface for different land covers in urban areas (Chang et al., 2015).	132	Elsevier
9	Water Resources Research	9 (1.8)	9	21,005	8.8	The role of land use and soils in regulating water flow in small headwater catchments of the Andes (Roa-García et al., 2011)	96	John Wiley & Sons
10	Catena	9 (1.8)	9	34,802	10.5	Influence of soil moisture and plant roots on the soil infiltration capacity at different stages in arid grasslands of China (Liu et al., 2019).	130	Elsevier

Note. TP: total publications; TC: total citations.

journal was not listed under the top 10 journal after the screening process for bibliometric analysis in this study, as 99% paper from this journal were written in Chinese, resulting in it being difficult for English readers to comprehend.

Leading Countries and Top Institutions

The top 15 most productive countries that contribute to the soil infiltration study are shown in Figure 4 and Table 3. China and the USA contributed about 28% of the global publications, indicating their significant role in the progression of soil infiltration studies. With 137 publications from a total of 69 journal, China was the top-ranked country. Whereas the USA ranked second with 102 publications. Although the Chinese Academy of Sciences produced more papers than the USDA Agricultural Research Service, considering that the Chinese Academy of Sciences has 124 branches, and direct comparison might be biased (Khudzari et al., 2018). The dominance of China and the United States in this research may

Table 2
Top 10 productive journals on soil infiltration study (cumulatively)

	Journal	TP (%)	TC	CiteScore 2023	Publisher
1	Nongye Gongcheng Xuebao/ Transactions of the Chinese Society of Agricultural Engineering	85 (3.5)	14,939	4.1	Chinese Society of Agricultural Engineering
2	Science of the Total Environment	82 (3.4)	550,528	17.6	Elsevier
3	Journal of Hydrology	76 (3.1)	64,858	11.0	Elsevier
4	Water Science and Technology	46 (1.9)	9,579	4.9	IWA Publishing
5	Hydrological Processes	47 (1.9)	7,127	6.0	John Wiley & Sons
6	Catena	46 (1.9)	34,802	10.5	Elsevier
7	Water (Switzerland)	42 (1.7)	90,757	5.8	Multidisciplinary Digital Publishing Institute (MDPI)
8	Journal Of Irrigation and Drainage Engineering	35 (1.4)	853	3.2	American Society of Civil Engineers (ASCE)
9	Journal Of Environmental Management	31 (1.3)	158,269	13.7	Elsevier
10	Geoderma	29 (1.2)	24,328	11.8	Elsevier

Note. TP: total publications; TC: total citations



Figure 4. Top 15 most productive countries and academic institutions in soil infiltration publications

Table 3
Top 15 most productive academic institutions in soil studies

Rank	Country	TP	The most productive academic institutions
1	China	137	Chinese Academy of Sciences
2	USA	102	USDA Agricultural Research Service
3	Spain	35	Consejo Superior de Investigaciones Científicas
4	UK	29	UK Centre for Ecology & Hydrology
5	Germany	28	helmholtz Zentrum für Umweltforschung
6	Netherlands	23	Wageningen University & Research
7	Italy	22	Wageningen University & Research
8	Australia	21	Commonwealth Scientific and Industrial Research Organisation
9	India	20	King Saud University
10	France	19	INRAE
11	Indonesia	17	Brawijaya University
12	Brazil	16	Universidade de São Paulo
13	Sweden	15	Luleå University of Technology
14	Japan	15	Tohoku University
15	Iran	14	University of Tehran

be associated with their substantial research funding, extensive agricultural environmental research programmes and large scientific communities in soil science and hydrology.

Figure 5 The co-authorship network map reveals twelve clusters, each represented by a different colour. The size of the frame corresponds to the occurrence frequency, and a thicker line represents a strong connection between two countries. Figure 5 represents the global nature of research collaboration on soil infiltration which strong connection seen with the USA, Spain, Brazil and Colombia.

Table 4 shows the leading countries contributing to soil infiltration ranked according to total link strength (TLS). The total number of documents produced, total citations, and TLS average citations per document are taken as the primary data for this study. The USA is identified as the most affiliated country based on co-authorship data, which is linked to 25 countries. The Netherlands followed with 23, the UK (22) and China and Germany with 21 countries each. Although China ranked first in total documents and has a strong connection with other countries with 45 TLS, the average citations are significantly low (17.2). This suggests that China studies may focus on the plateau climate or be written in a non-English language. Differences in citation performance may also be influenced by factors such as journal visibility, international collaboration networks and accessibility of publications. The research that focuses on regional environmental conditions may attract fewer global citations despite its scientific importance.

Authors

Table 5 lists the 15 most prolific authors in soil infiltration based on authors' contributions to the field for the last 5 years, affiliated to 7 countries as follows: China (6 authors), USA (4 authors), Germany, Hong Kong, France, Australia and Singapore, each with 1 author. The first author's publications between 1987 and 2005, in which 2 authors had a role as the first author, 9 as co-author and 4 as the last author. The affiliation of the authors showed that soil infiltration was within fields related to geotechnical engineering, soil mechanics, environment, and water conservation.

The first prolific author is Rattan A. Lal from the USA, with the first publication in 1970, who currently has 1332 publications, 146 h-index and 95,254 citations. The second and third top authors, Harry Vereecken and Jiri Simunek, are from Germany and the USA, respectively, shared 23 co-authorship documents, and the most cited article was on modelling soil processes (Vereecken et al., 2016). Charles Wang Wei Ng from Hong Kong ranked 4th on the list with 79 h-index. Dani Or on 5th with 77 h-index also shared 3 co-authorships with Harry Vereecken and Jiri Simunek, where the minimum citations of 45 for an article on Kirkham's Model on soil infiltration (Jury et al., 2011). There are several influential authors who have long academic careers that extend beyond the analysed period. Consequently, their prominence in the dataset sustained as scholarly contributors

rather than being actively limited to the recent publication period. While a co-authorship network provides insights into international collaboration patterns, they do not necessarily distinguish between deep research integration and institutional collaboration related to field data collection or experimental activities.

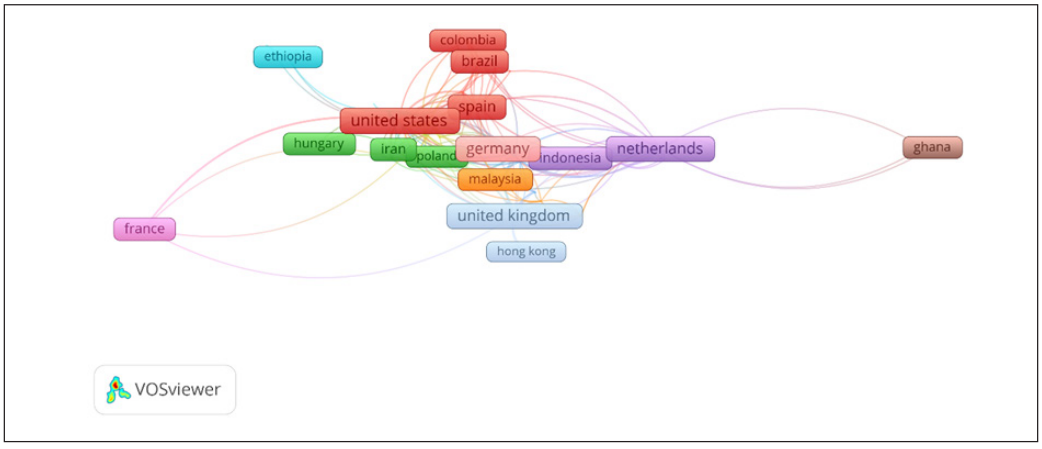


Figure 5. Co-authorship countries contributed to soil infiltration research

Table 4
Leading 15 countries publishing on soil infiltration

Country	Cluster	TLS	Documents	Citations	Avg. citations
United States	1	60	99	2983	30.1313
China	4	45	136	2340	17.2059
Netherlands	5	39	22	377	17.1364
Germany	10	34	27	1360	50.3704
United Kingdom	12	32	28	1461	52.1786
Spain	1	28	34	801	23.5588
Australia	3	25	20	377	18.85
Italy	1	24	21	325	15.4762
Sweden	3	22	15	181	12.0667
Brazil	1	14	15	110	7.3333
India	6	14	19	455	23.9474
Kenya	5	14	6	111	18.5
Switzerland	7	12	8	181	22.625
France	9	11	17	355	20.8824
Saudi Arabia	4	11	7	78	11.1429

Table 5
List of 15 most prolific authors in soil infiltration research area

	Author	Scopus ID	Year of first publication*	TP	h-index	TC	Current Affiliation	Country
1	Lal, Rattan A.	55444564800	1970[c]	1332	146	95254	The Ohio State University, United States	USA
2	Vereecken, Harry	7004082452	1988[b]	757	83	29065	Forschungszentrum Jülich GmbH,Julich, Germany	Germany
3	ŠimŮnek, Jiří	35561493700	1988[b]	459	83	28101	University of California, Riverside, Riverside, United States	USA
4	Ng, Charles Wang Wai	55454271800	1991[b]	642	79	23323	Hong Kong University of Science and Technology,Hong Kong	Hong Kong
5	Or, Dani	57204340741	1992[b]	378	77	22243	College of Engineering,Reno, United States	USA
6	Cui, Yujun	55497053200	1996[b]	545	70	19654	Laboratoire Navier, Marne-la-Vallee, France	France
7	Xia, Jun	7402327199	1987[a]	614	63	16169	Wuhan University,Wuhan, China	China
8	Fletcher, Tim D.	56251580300	1994[b]	220	65	15451	Faculty of Science, Melbourne, Australia	Australia
9	Wang, Kelin	12240170600	2003[b]	490	62	15451	Chinese Academy of Sciences,Beijing, China	China
10	Rahardjo, Harianto	7004487503	1983[c]	313	58	13420	School of Civil and Environmental Engineering, Singapore City, Singapore	Singapore
11	Blanco-Canqui, Humberto	8558145900	2002[a]	168	52	10993	University of Nebraska-Lincoln, Lincoln, United States	USA
12	Tang, Chao-Sheng	11640513400	2005[b]	269	50	10880	Nanjing University, Nanjing, China	China
13	Mu, Xingmin	56346174400	1993[c]	216	51	9274	Institute of Soil and Water Conservation Chinese Academy of Sciences,Yangling, China	China
14	Li, Xiaoyan	35205746400	1998[b]	291	50	9056	Qinghai Normal University, Xining, China	China
15	Peng, Jianbing	57204721822	2000[c]	402	48	8980	Chang'an University,Xi'an, China	China

Note. Role in publications, superscripts, [a] First author, [b] Co-author, [c] Last author.

Keywords

A total of 1614 author keywords were recorded and were used at least once; 239 (14.8%) were used twice, 101 (6.3%) were used thrice, and 31 (2%) were used more than 5 times, which were used for mapping in VOSviewer.

Keyword Cluster

It has been identified that there are four main theme clusters for soil infiltration, as shown in Figure 6. The first theme is climate change impacts on soil infiltration, with the main keyword being 'runoff' (22 occurrences, 22 links). The second theme is urbanization, land use and hydrological response with 'soil infiltration' (35 occurrences, 10 links) as their main keyword followed by soil management, conservation and adaption strategies with the main keyword is 'infiltration' (41 occurrences, 27 links) and the last theme is modelling and measurement approaches with 'hydraulic conductivity' (9 occurrences, 5 links) as its main keyword. For each theme, several other keywords contribute to the theme, as shown in Table 6.

Theme 1: Climate Change Impacts on Soil Infiltration

In this theme, it highlights a strong linkage between infiltration process and hydrological aspects represented by keywords such as precipitation intensity, temperature, soil moisture and soil erosion. The bibliometric analysis indicates an increasing presence of climate

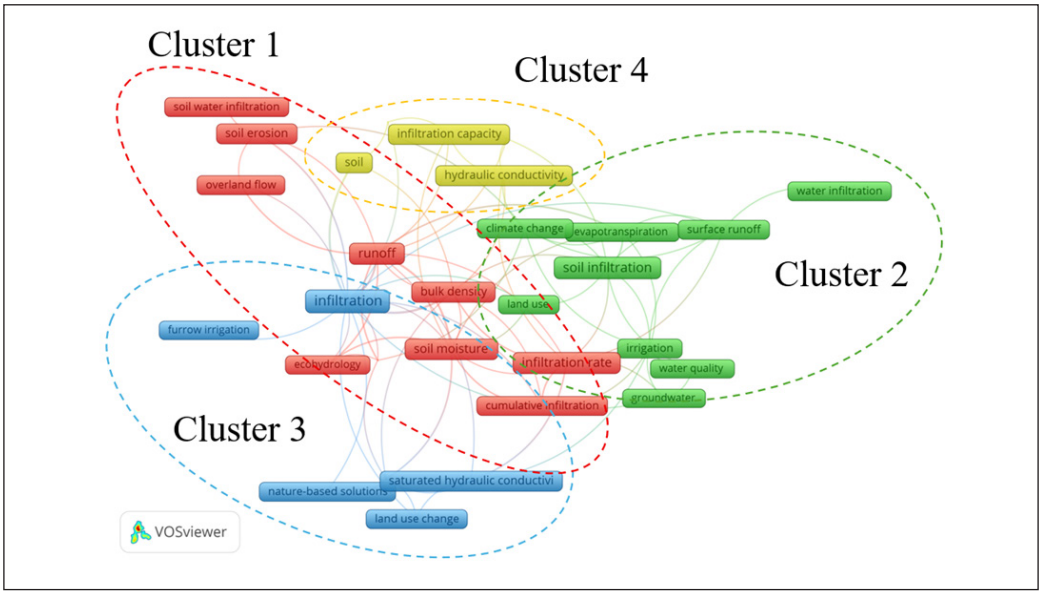


Figure 6. Author keywords co-occurrence with overlay visualisation mode, with a minimum of 5 occurrences

Table 6
Keyword cluster and its representative focus area

Theme cluster	Keywords	Research questions	References
1. Climate change impacts on soil infiltration	Cumulative infiltration Climate change Ecohydrology Infiltration rate Overland flow Rainfall intensity Runoff Soil erosion Soil moisture Soil water infiltration	How does climate change alter soil infiltration patterns, which later leads to groundwater recharge and irrigation efficiency?	(Arredondo et al., 2025; Ebrahim et al., 2019; Si et al., 2023)
2. Urbanisation, land use and hydrological response	Evapotranspiration Groundwater Infiltration models Irrigation Land use Soil infiltration Surface runoff Water infiltration Water quality	To which extend the soil properties and precipitation interact to influence infiltration-soil erosion?	(Hernández-Atencia et al., 2023; Mejía-Veintimilla et al., 2024; Senadi et al., 2025)
3. Soil management, conservation and adaptation strategies	Furrow irrigation Infiltration Land use change Nature-based solutions Saturated hydraulic conductivity Soil infiltration rate	What are the best management strategies to mitigate runoff and erosion?	(Abu-hashim et al., 2023; Lammers et al., 2022; Louis et al., 2024)
4. Modelling and measurement approaches	Hydraulic conductivity Infiltration capacity Soil Soil properties Stormwater	How can infiltration models be improved to predict infiltration-runoff?	(García Martínez et al., 2025; Layth Alrubaye et al., 2020; Sadeghi et al., 2024)

change-related keywords in recent publications, suggesting a growing focus on the interaction between soil infiltration processes and climate change impacts. A study that emphasises the relation between soil physical properties and rainfall characteristics leads to a better understanding of infiltration behaviour, which subsequently affects the overland flow and cumulative infiltration. These processes are crucial in civil engineering to have proper flood risk management, soil conservation and eco-hydrology for water resources management.

Theme 2: Urbanisation, Land Use and Hydrological Response

This cluster assesses the effects of urbanisation and land use change on infiltration capacity. Based on the keywords, it suggests an integrated perspective of infiltration across agricultural, urban and hydrological dynamics. Studies in this cluster apply the infiltration model to evaluate the effects of land use change and irrigation practices on hydrological balance. This knowledge supports strategies for sustainable water management.

Theme 3: Soil Management, Conservation and Adaptation Strategies

The third cluster closely linked the soil properties and their influence on stormwater control. With keywords such as hydraulic conductivity, infiltration rate and furrow irrigation, this highlights the role of soil infiltration in urban drainage and sustainable stormwater management systems. By understanding the soil characteristics, ensure the long-term functionality and effectiveness of stormwater infiltration practices in the design of infiltration trenches, permeable pavements and green infrastructure for civil engineering practices.

Theme 4: Modelling and Measurement Approaches

The last cluster centres on the advancement of developing and comparing infiltration and runoff models, SWMM, HEC-RAS, and HYDRUS to improve predictions of hydrological responses to climate change and land use changes. In this cluster, an integration of remote sensing, big data and machine learning to monitor and predict the soil hydrological processes has also been developed, as proven by the presence of the "remote sensing" and "machine learning" keywords in this cluster generated in VOSviewer. Although it is not as prominent as the top 5 keywords listed.

Topics of Interest

Soil water infiltration is a crucial element in the hydrological cycle, primarily in the rate of water absorption into the soil. This study is important not only to the hydrologists and irrigation engineers but also to the soil scientists. This field of study helps in determining the soil moisture levels, the dynamics of surface runoff behaviour also related to the sediment transport. Studies such as determination of the optimal empirical infiltration rate model (Jejurkar & Rajurkar, 2015; Rasool et al., 2021; Salahou et al., 2020), relationship between rainfall-runoff and soil infiltration (Chen et al., 2025; Hu et al., 2021), development of irrigation system (Zhang & Flerchinger, 2001) and groundwater (Cui et al., 2019; Tedoldi et al., 2016; Xu et al., 2025) are common among researchers. Water resources problems are increasing day by day, and research trend phase change where studies on climate change linked to land use and ecohydrology in conjunction with the United Nations Agenda

2030 for Goal 6 (Clean water and sanitation) and Goal 13 (Climate action). With that, soil infiltration studies are now more extensive, encompassing sustainability with the aid of an Artificial Intelligence (AI) technique. It is undeniable that this field is a dynamic study between soil and water, and employing AI tools can effectively demonstrate the soil physical properties as well as infiltration process, such as (Mehta et al., 2024; Sihag et al., 2020; Singh et al., 2017; Singh et al., 2021).

Limitation of Study

The result of the search may not cover all studies related to soil infiltration available on Scopus, as the search is restricted to "soil infiltration*" in titles and abstracts. Some publications with a lag publication process may also not be mentioned in this article. For future work, it is proposed that extensive studies be performed by using several databases, such as Scopus and Web of Science, to reduce the potential level of selection bias. Although, Zeng et al. (2023) mentioned that integration of multiple databases for bibliometric analysis is tedious work. Moreover, this study is limited to English-language articles and 30% of the eliminated articles were written in the Chinese language. As this study only considered English language publications, it may exclude relevant research conducted in non-English language regions. As a result, the global representation of soil infiltration research may be partially limited. This bibliometric analysis was conducted on 16th January 2025, and it is expected that the number of publications on soil infiltration will keep increasing in the days to come. Citation counts may also be influenced by the year of publication, as recently published open-access articles may not yet have accumulated their full citation potential. Therefore, citation comparison between open access and non-open access should be interpreted with caution. An alternative to journal metrics other than CiteScore values, such as Field-Weighted Citation Impact and journal percentile rankings, could also provide additional perspective for evaluating journal impacts.

CONCLUSION

This study offers a comprehensive bibliometric analysis and visualisation of the soil infiltration study domain. For the last 24 years, the growth of publications has been strong, and it is anticipated to continue to rise. Relevant literature review on the topic from 2000 to 2025 in the Scopus database employing VOSviewer software. This bibliometric analysis study not only covered countries, institutions, journals, authors, references, keywords and research trends. The most prominent research topics in the field of soil infiltration were found to be: (1) soil erosion, (2) urban storm water infiltration, and (3) irrigation systems. The findings of this study offer an overview of soil infiltration for researchers before proceeding, and to some extent help to identify research gaps.

The growth in publication output shows global recognition of infiltration as a critical process in sustainable land and water management. Thus, the research analysis for soil infiltration is becoming a fundamental discipline, integrating aspects of hydrology, agricultural, environmental engineering, and climate change adaptation. The integration of advanced technologies such as remote sensing, machine learning and field-based monitoring systems should be emphasised for future research collaboration among developed countries, which may enhance the accuracy of hydrological modelling in future research. The need to explore the implications of climate variability and land use change on infiltration capacity should also be encouraging in this field.

By identifying emerging trends, influential contributors and potential research gaps, the soil infiltration study potentially contributes to shaping the future research innovation in achieving sustainable soil and water resource management. The bibliometric analysis manages to provide a valuable reference for researchers, practitioners and policymakers to explore innovation strategies in soil infiltration management.

ACKNOWLEDGEMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot Q978).

REFERENCES

Abu-Hashim, M., Lilienthal, H., Schnug, E., Lasaponara, R., & Mohamed, E. S. (2023). Can a change in agriculture management practice improve soil physical properties? *Sustainability, 15*(4), Article 3573. https://doi.org/10.3390/su15043573

Arredondo, A. M., Mirzaei, M., Wu, G. L., Keshavarzi, A., Durán-Zuazo, V. H., Bernardo, F. S., & Rodrigo-Comino, J. (2025). Evaluating soil respiration and water infiltration in esparto grasslands: The effects of hillslope position and soil management in arid, human-affected Mediterranean environments. *Discover Sustainability, 6*(1), 1-15. https://doi.org/10.1007/s43621-025-01282-x

Basche, A. D., & DeLonge, M. S. (2019). Comparing infiltration rates in soils managed with conventional and alternative farming methods: A meta-analysis. *PLOS ONE, 14*(9), Article e0215702. https://doi.org/10.1371/journal.pone.0215702

Campisano, A., Creaco, E., & Modica, C. (2011). A simplified approach for the design of infiltration trenches. *Water Science and Technology, 64*(6), 1362-1367. https://doi.org/10.2166/wst.2011.170

Chang, T.-J., Wang, C.-H., & Chen, A. S. (2015). A novel approach to model dynamic flow interactions between storm sewer system and overland surface for different land covers in urban areas. *Journal of Hydrology, 524*, 662-679. https://doi.org/10.1016/j.jhydrol.2015.03.014

Chen, C., Li, B., Xiao, Y., Li, H., & Zhang, T. (2025). Hydrological and hydrodynamic coupling simulation under composite underlying surfaces in urban polder areas. *Journal of Hydrology: Regional Studies, 57*, Article 102154. https://doi.org/10.1016/j.ejrh.2024.102154

Cui, Z., Wu, G., Huang, Z., & Liu, Y. (2019). Fine roots determine soil infiltration potential more than soil water content in semi-arid grassland soils. *Journal of Hydrology, 578*, Article 124023.https://doi.org/10.1016/j.jhydrol.2019.124023

Donohue, C. (1972). A bibliometric analysis of certain information science literature. *Journal of the American Society for Information Science, 23*(5), 313-317. https://doi.org/10.1002/asi.4630230506

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Marc, W. (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

Ebrahim, M. N., Che Man, H., Mohamed Zawawi, M. A., & Hamzah, M. H. (2019). Prediction of groundwater contaminants from cattle farm using Visual MODFLOW. *Pertanika Journal of Science and Technology, 27*(4), 2265-2279.

García Martínez, M., Osornio Berthet, L. J., Meléndez Estrada, J., & Cruz Castro, O. (2025). Identification of flood-prone areas using the HEC-RAS 2D model in a region of the Valley of Mexico. *Chemosphere, 386*, Article 144609. https://doi.org/10.1016/j.chemosphere.2025.144609

Hernández-Atencia, Y., Peña, L. E., Muñoz-Ramos, J., Rojas, I., & Álvarez, A. (2023). Use of soil infiltration capacity and stream flow velocity to estimate physical flood vulnerability under land-use change scenarios. *Water, 15*(6), Article 1214. https://doi.org/10.3390/w15061214

Hu, C., Xia, J., She, D., Song, Z., Zhang, Y., & Hong, S. (2021). A new urban hydrological model considering various land covers for flood simulation. *Journal of Hydrology, 603*, Article 126833.https://doi.org/10.1016/j.jhydrol.2021.126833

Jejurkar, C. L., & Rajurkar, M. P. (2015). An investigational approach for the modelling of infiltration process in a clay soil. *KSCE Journal of Civil Engineering, 19*(6), 1916-1921. https://doi.org/10.1007/s12205-014-0149-3

Jury, W. A., Or, D., Pachepsky, Y., Vereecken, H., Van Genuchten, M. T., & Horton, R. (2011). Kirkham's legacy and contemporary challenges in soil physics research. *Soil Science Society of America Journal, 75*(5), 1589-1601. https://doi.org/10.2136/sssaj2011.0115

Khudzari, J., Kurian, J., Tartakovsky, B., & Raghavan, G. S. V. (2018). Bibliometric analysis of global research trends on microbial fuel cells using Scopus database. *Biochemical Engineering Journal, 136*, 51-60. https://doi.org/10.1016/j.bej.2018.05.002

Lammers, R. W., Miller, L., Bledsoe, B. P., & ASCE, M. (2022). Effects of design and climate on bioretention effectiveness for watershed-scale hydrologic benefits. *Journal of Sustainable Water in the Built Environment*. https://doi.org/10.1061/JSWBAY.0000993

Layth Alrubaye, Y., Yusuf, B., & Hamad, S. N. (2020). Development of a self-regulated bubble irrigation system to control the size and shape of wetting fronts. *Pertanika Journal of Science & Technology, 28*(4), 1297-1313. https://doi.org/10.47836/pjst.28.4.09

Liu, Y., Cui, Z., Huang, Z., López-Vicente, M., & Wu, G.-L. (2019). Influence of soil moisture and plant roots on the soil infiltration capacity at different stages in arid grasslands of China. *Catena, 182*, Article 104147. https://doi.org/10.1016/j.catena.2019.104147

Liu, Z., Lu, Y., Wang, P., Wang, T., Liu, S., Johnson, A. C., Sweetman, A. J., & Baninla, Y. (2017). Pollution pathways and release estimation of perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) in central and eastern China. *Science of the Total Environment, 580*, 1247-1256.https://doi.org/10.1016/j.scitotenv.2016.12.085

Louis, P., Delgado-Gonzalez, L., Lassabatère, L., Czarnes, S., Aubert, J., Imig, A., & Clément, R. (2024). Woodchip-filled trenches: A solution to enhance urban water infiltration capacity? *Geoderma, 451*, Article 117057. https://doi.org/10.1016/j.geoderma.2024.117057

Mehta, V., Hasanvand, S., Sepahvand, A., Sihag, P., & Beiranvand, N. (2024). A benchmark comparison of AI-based modelling of soil infiltration rates. *Journal of Hydroinformatics, 26*(12), 3060-3079. https://doi.org/10.2166/hydro.2024.086

Mejía-Veintimilla, D., Ochoa-Cueva, P., & Arteaga-Marín, J. (2024). Evaluation of the hydrological response to land use change scenarios in urban and non-urban mountain basins in Ecuador. *Land, 13*(11), Article 1907. https://doi.org/10.3390/land13111907

Mu, W., Yu, F., Li, C., Xie, Y., Tian, J., Liu, J., & Zhao, N. (2015). Effects of rainfall intensity and slope gradient on runoff and soil moisture content on different growing stages of spring maize. *Water, 7*(6), 2990-3008. https://doi.org/10.3390/w7062990

Rasool, T., Qayoom, D. A., & Wani, M. (2021). Comparative evaluation of infiltration models under different land covers: Comparison of infiltration model parameter estimation techniques under different land use/land covers. *International Journal of Hydrology Science and Technology, 12*(4), 448-476. https://doi.org/10.1134/S0097807821040175

Sadeghi, M., Shayannejad, M., Sadraddini, A. A., & Salehi, S. (2024). Estimation of soil infiltration coefficient in furrow irrigation using the combination of the hydrodynamics model and Richard's equation. *Physics and Chemistry of the Earth, Parts A/B/C, 135*, Article 103649. https://doi.org/10.1016/j.pce.2024.103649

Salahou, M. K., Jiao, X. Y., & Lu, H. S. (2020). Assessment of empirical and semi-empirical models for estimating a soil infiltration function. *Transactions of the ASABE, 63*(4), 833–845. https://doi.org/10.13031/trans.13639

Senadi, I., Zeroual, A., & Meddi, H. (2025). From urban sprawl to aquifer strain: The hydrological consequences of land use/land cover changes in the Mitidja plain, Algeria. *Physics and Chemistry of the Earth, Parts A/B/C*, Article 103951. https://doi.org/10.1016/j.pce.2025.103951

Si, J., Li, J., Lu, S., Qi, X., Zhang, X., Bao, W., Zhang, X., Zhou, S., Jin, C., Qi, L., Qi, Y., Zheng, X., Gong, Y., & Wang, Z. (2023). Effects of climate change on surface runoff and soil moisture in the source region of the Yellow River. *Water, 15*(11), Article 2104. https://doi.org/10.3390/w15112104

Sihag, P., Kumar, M., & Singh, B. (2020). Assessment of infiltration models developed using soft computing techniques. *Geology, Ecology, and Landscapes, 5*(4), 241-251.https://doi.org/10.1080/24749508.2020.1720475

Singh, B., Ebtehaj, I., Sihag, P., & Bonakdari, H. (2021). An expert system for predicting infiltration characteristics. *Water Science and Technology: Water Supply, 22*(3), 2847-2862. https://doi.org/10.2166/ws.2021.430

Singh, B., Sihag, P., & Singh, K. (2017). Modelling of the impact of water quality on the infiltration rate of soil by random forest regression. *Modelling Earth Systems and Environment, 3*(3), 999-1004. https://doi.org/10.1007/s40808-017-0347-3

Sittisom, P., Tangsongsuwan, R., Munlikawong, S., Wongsapai, W., Sitthikankun, S., & Rinchumphu, D. (2022). The determination of soil infiltration rate of urban bioretention design process in Chiang Mai, Thailand. *Nakhara Journal of Environmental Design and Planning, 21*(3), Article 228. https://doi.org/10.54028/NJ202221228

Skougaard Kaspersen, P., Høegh Ravn, N., Arnbjerg-Nielsen, K., Madsen, H., & Drews, M. (2017). Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding. *Hydrology and Earth System Sciences, 21*(8), 4131-4147. https://doi.org/10.5194/hess-21-4131-2017

Tedoldi, D., Chebbo, G., Pierlot, D., Kovacs, Y., & Gromaire, M. (2016). Impact of runoff infiltration on contaminant accumulation and transport in the soil/filter media of sustainable urban drainage systems: A literature review. *Science of the Total Environment, 569-570*, 904-926. https://doi.org/10.1016/j.scitotenv.2016.04.215

Van Eck, N., Waltman, L., Berg, J. D., & Kaymak, U. (2006). Visualising the computational intelligence field. *IEEE Computational Intelligence Magazine, 1*(4), 6-10. https://doi.org/10.1109/CI-M.2006.248043

Vereecken, H., Schnepf, A., Hopmans, J., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M., Amelung, W., Aitkenhead, M., Allison, S., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., … Young, I. (2016). Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal, 15*(5), 1-57. https://doi.org/10.2136/vzj2015.09.0131

Xu, Y., Luo, L., Guo, W., Chen, Z., Tian, P., & Wang, W. (2025). Vegetation roots-induced changes of loess infiltration: Greatly promoting lateral flow. *Catena, 250*, Article 108803. https://doi.org/10.1016/j.catena.2025.108803

Yang, F., Cen, R., Feng, W., Liu, J., Qu, Z., & Miao, Q. (2020). Effects of super-absorbent polymer on soil remediation and crop growth in arid and semi-arid areas. *Sustainability, 12*(18), Article 7825. https://doi.org/10.3390/su12187825

Zeng, Y., Cao, S., & Yang, H. (2023). Globalresearch trends on COVID-19 and stroke: A bibliometric analysis. *Frontiers in Neurology, 14*, Article 1147867. https://doi.org/10.3389/fneur.2023.1147867

Zhang, N., Yan, J., & Liu, P. (2021). Effect of bacterial manure on the properties of complex soil and growth of ryegrass. *Agronomy, 11*(3), Article 568. https://doi.org/10.3390/agronomy11030568

Zhang, X., & Flerchinger, G. N. (2001). Infiltration into freezing and thawing soils under differing field treatments. *Journal of Irrigation and Drainage Engineering, 127*(3), 176-182. https://doi.org/10.1061/(ASCE)0733-9437(2001)127:3(176)

Pertanika J. Sci. & Technol. 34 (2): 1025 - 1052 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Integrating Time Lag Effects in Predictive Modelling Using Random Forest for Early Detection of Bagworm Outbreaks

Yi Peng Wang[1,5,6], Nurul Hawani Idris[1,2]*, Norhayu Asib[3], Mohamad Hafis Izran Ishak[4], and Alvin Lau Meng Shin[1,2]

[1]TropicalMap Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

[2]Geoinformation Department, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

[3]Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[4]Control and Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

[5]Department of Hydraulic Engineering, Hebei University of Water Resources and Electric Engineering, 061001 Cangzhou, Hebei, China

[6]Hebei Technology Innovation Centre for Coastal Wetland Water Resources Allocation and Ecological Protection, 061001 Cangzhou, Hebei, China

ABSTRACT

The palm oil bagworm inflicts substantial economic losses on the palm oil industry in Malaysia annually. At present, integrated pest management remains the primary method for managing pests and diseases. However, early prediction of pest occurrence can help to assist in managing sustainable palm oils more effectively. Understanding the time window as well as the lag associated with environmental changes at the plot level is fundamental to such pest management strategies. However, few studies consider historical environmental data and recognise the biological and ecological time lags in pest population responses in their predictive models. This study aims to develop a predictive model based on spatiotemporal environmental data to support early warning systems for managing palm oil bagworm infestations. Therefore, this study examines the best time lags from one to six cycles before the census lagged variables, ranging from period 1 to 6 and trains the model using a random forest algorithm under 12 different time window configurations. Key driving factors influencing bagworm occurrences, including rainfall, land surface temperature, humidity, and road density, are investigated. Experimental results indicate that the combination of the first three cycles

ARTICLE INFO

Article history:
Received: 01 August 2025
Accepted: 16 March 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.18

E-mail addresses:
yipeng1992@graduate.utm.my (Yi Peng Wang)
hawani@utm.my (Nurual Hawani Idris)
norhayuasib@upm.edu.my (Norhayu Asib)
hafis@utm.my (Mohamad Hafis Izran)
alvinlau@utm.my (Alvin Lau)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

(Lag 1, 2, and 3) of temporal data before the census date yields the best model performance with a recall rate of 0.76 and an AUC value of 0.68. Among these, relative humidity (RH_lag2; two cycles before the census) and surface temperature (LST_lag1; one cycle before the census) emerges as the most influential predictors of bagworm outbreaks. Finally, the visualisation of model predictions enables plot-based targeted preventive measures in high-risk areas, thereby supporting the Sustainable Development Goal (SDG) in palm oil management.

Keywords: Bagworm, Geographic Information System (GIS), prediction model, random forest, spatio-temporal, time lag

INTRODUCTION

Palm oil is currently the most consumed and traded vegetable oil in the world. The global palm oil market in 2019 is estimated at 74.6 million tonnes. This figure is expected to further increase to 111.3 million tonnes by 2025 (Sundaraja et al., 2021). Malaysia is the world's second-largest producer of palm oil. In 2020, the country's palm oil production and exports accounted for 25.8% and 34.3% of the world's palm oil production and exports, respectively (Zakaria et al., 2024). Malaysia's palm oil industry has a long history of bagworm infestations, and the substantial yield reduction due to bagworm attacks is becoming a main threat in the palm oil industry (Sulaiman & Talip, 2021; Thaer et al., 2021).

The bagworm (Lepidoptera: Psychidae) is a major foliar pest capable of causing severe leaf damage and yield losses of up to 43% if left untreated (Sulaiman & Talip, 2021), as reported by the Malaysian Palm Oil Board (MPOB) in 2016. The rapid expansion and widespread monoculture of palm oil plantations provide abundant food resources, creating ideal conditions for the pest's growth and spread. Several factors influence bagworm population dynamics, including environmental conditions (both biotic and abiotic), the balance of predator-prey interactions, the availability of alternate host plants, and human activities. When environmental conditions are favourable, these factors often trigger localised infestations at high densities, which can escalate into widespread outbreaks with serious consequences for crop productivity.

Precision agriculture, utilising technologies such as remote sensing, IoT, and LiDAR visualised through Geographic Information Systems (GIS), has become increasingly common globally (Sabtu et al., 2018). The integration of these digital agriculture tools with Artificial Intelligence (AI) enhances a country's ability to modernise farming, significantly accelerating agricultural processes, improving productivity, and streamlining the food supply chain. Furthermore, digital agriculture actively supports efforts toward achieving carbon neutrality by optimising resource use, reducing greenhouse gas emissions and chemical pesticide use, and improving agricultural resilience to climate change (Balasundram et al., 2023).

Bagworms (Lepidoptera: Psychidae) are primary leaf-eating pests of palm oil in Malaysia, which can cause about 33%- 40% yield losses (Sulaiman & Talip, 2021; Thaer et al., 2021). They feed on leaves at a rapid pace. There are some kinds of species of bagworms like *Metisa plana*, *Pteroma pendula* and *Mahasena corbetti* (Hoong & Hoh, 1992; Lever, 2008; Wahid et al., 1988). Studies have shown that *Metisa plana* is the most widely distributed palm oil insect in Peninsular Malaysia, followed by *Pteroma pendula*. The most severely affected areas are in Perak, Sabah and Johor (Kamarudin & Mohd, 2007; Napiah et al., 2023).

Chemical control is still the major control mechanism in managing bagworm outbreaks in most plantations, as compared to smallholding plantations (Kamarudin et al., 2017; Salim & Hamid, 2012). However, over-reliance and excessive use of chemical insecticides to control bagworm has often led to the development of other, more persistent problems. For example, the resistance of pests to treatment, the abundance of harmful chemical residues in the environment, and the disruption of beneficial insect populations (Kamarudin et al., 2017) are not conducive to sustainable development.

One way to improve the efficiency of pesticide use and reduce the use of insecticides is to make predictions in advance about where and when those pests and diseases are likely to occur, which requires analysing and modelling pest and disease outbreak patterns from both spatial and temporal perspectives, and accordingly using pesticides within a reasonable spatiotemporal range. Studying the problem from only a temporal or only a spatial perspective is not feasible. The primary rationale is that the essence of pest and disease outbreak dynamics is spatio-temporal in nature (Li et al., 2021), and considering time and space separately during modelling fails to capture the key ecological mechanisms involved: temporal-only modelling overlooks spatial relationships between outbreak locations (e.g., areas surrounding an outbreak site are more likely to be affected due to dispersal), while spatial-only modelling is likely to ignore the temporal cumulative effects inherent in outbreak mechanisms (Guo et al., 2020). From a practical application perspective, spatio-temporal models can simultaneously answer the question of "when" and "where" risks are high.

The occurrence of observed diseases or pests is not directly influenced by current meteorological conditions, but rather by those of a previous period, indicating a time lag between the emergence of diseases or pests and environmental fluctuations (Holloway et al., 2018). At the same time, the occurrence of diseases and pests is the result of the cumulative influence of environmental variables over a period of time windows (Hamer et al., 2020; Pol et al., 2016). This characteristic of disease and pest research requires that the impact of time lag and time windows be considered when modelling. The time window refers to the period of past days or weeks to understand the environment before a pest outbreak occurs. Whereas the time lag indicates how long it takes for environmental changes to show an effect on pests.

Although abundant environmental data have been utilised in establishing the relationship between ecological characteristic variables and pests and diseases, most studies have not explicitly considered the influence of time lag and cumulative environmental variables. In their research, immediate or period-averaged environmental variables are usually used to explain the spatial distribution or risk prediction of pests and diseases (Gachoki et al., 2024; Lee et al., 2023; Mahmoodi et al., 2024; Makori et al., 2024; Rouabah et al., 2022), without incorporating the lagged impact of historical variables on the dynamics of pests and diseases into the model.

As the biological characteristics (growth cycles and patterns) of different pests vary, the optimal time lag may also vary for different pests, and the time windows over which to gather environmental variables in a prediction model are likely to be highly species-specific (Holloway et al., 2018; Pol et al., 2016). In terms of the palm oil pest *Metisa plana*, there is currently a lack of systematic research determining the optimal time lag and time windows of environmental variables to help figure out its growth and development.

In fields of crop disease and pest prediction, the decisions made by prediction models can affect the efficiency of field operations. At such times, a good model can ensure the establishment of trust, meet regulatory requirements, and reduce decision-making errors (Dwivedi et al., 2023; Rojat et al., 2021). Currently, the interpretability of most model results related to agricultural pests and diseases is not detailed enough. Most model results only demonstrate the significance of environmental variables and whether their impacts are positive or negative (Hamer et al., 2020; Mahmoodi et al., 2024). However, there is a lack of models that consider when and where to act to prevent an outbreak before it occurs. Therefore, this study develops a prediction model for bagworm occurrence using Random Forest, a common machine learning algorithm (Guo et al., 2020; Rouabah et al., 2022), which considers the time lag before the effect and time windows of spatio-temporal environmental data. This study includes 3 objectives: (i) to develop an interpretable prediction model, (ii) to quantify time lag effects of environmental variables, and (iii) to identify key drivers. The time lag effects were examined by dividing environmental conditions into 12-time window schemes. The final model was quantitatively interpreted, and the key drivers and the most informative time lags for bagworm occurrence were identified through Shapley Additive exPlanations (SHAP) analysis (Lundberg et al., 2018).

MATERIALS AND METHODS

This section describes the methodology applied in this study. The following Figure 1 is the research framework that starts with problem identification. The modelling stage is for research objectives (i) and (ii), the interpretability stage and the performance stage are for objective (iii).

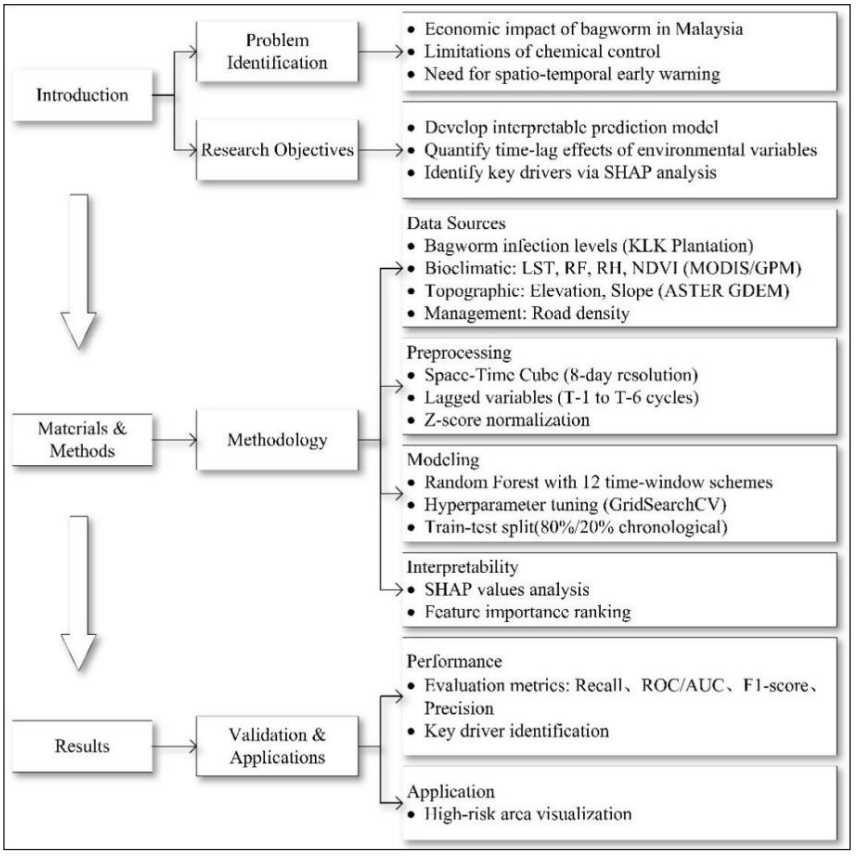


Figure 1. Research framework

Study Area and Dataset

The research area is in Lekir, situated in the Perak State, with geographic coordinates ranging from 4° 4' 48" N to 4° 10' 48" N and from 100° 45' 36" E to 100° 50' 24" E. The site covers an area of approximately 33.29 km², as shown in Figure 2.

The study area was divided into different plots and numbered. The plantation area is well-structured, divided into 42 plots, each assigned a unique identification number. The northern part of the palm plantation has higher terrain compared to the southern part. The elevation of this study area ranges between 4m and 37m above sea level. The main palm oil pest in the study area is *Metisa plana*. The statistical data indicating the severity of the pest and disease was obtained from the KLK palm oil plantation. The original census location data was collected using a handheld GPS. This data marked the geographical coordinates (latitude and longitude) of each census tree and the infection status of *Metisa plana* at specific dates. The overall period of the data is from April to December 2020, January to July 2021, and the entire year of 2022.

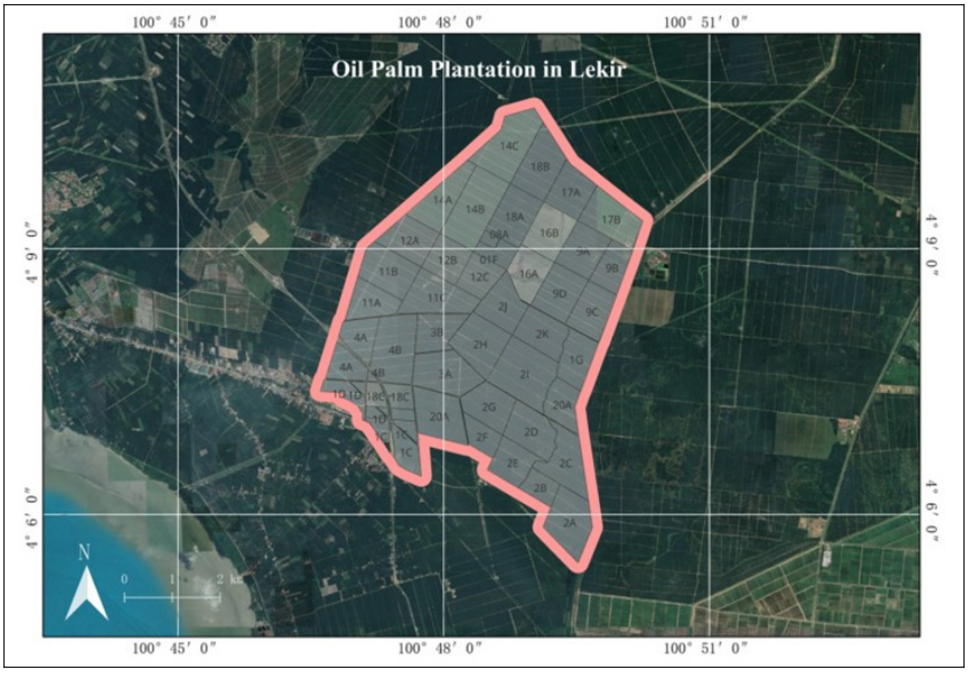


Figure 2. Location of the study site in Lekir, Perak

The census data were collected by manually counting the number of pests and assessing the severity of infestation in the plot. Some of the data collection cycles did not fully cover all plots in the study area. The infection levels were determined by the above regulation in Table 1.

The bioclimatic data used in this study include Land Surface Temperature (LST) from the MOD11A2 (Wan et al., 2021), Normalised Difference Vegetation Index (NDVI) from MOD13Q1 (Didan, 2021), and Relative Humidity (RH) derived from MOD05/06/07 (MODIS, A. S. T., 2017a, 2017b, 2017c), all of which were obtained from NASA's Earth science data at https://search.earthdata.nasa.gov/search. Rainfall (RF) data were obtained from the NASA Global Precipitation Measurement (GPM) dataset (Huffman, 2023), available at GES DISC Earth data (https://disc.gsfc.nasa.gov). The datasets covered the period from 2020 to 2022, aligning with the bagworm field census data. In addition, the predictor variables included spatial data such as road density, elevation, and slope. These topographic variables represent static features and were obtained from the official KLK plantation records and the ASTER GDEM V003 (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model, Version 003) dataset (NASA/METI/AIST/Japan, 2019), available at https://search.earthdata.nasa.gov/search/. Detailed information is provided in Table 2.

Table 1
Infection level regulation

Explanation	Symbol	Indicator
Larvae (<1cm, 0 pcs)	Green flag	OK
Larvae (<1cm, 1-4 pcs)	Yellow flag	Monitor carefully
Larvae (<1cm, >5 pcs)	Red flag	Immediate treatment
Larvae (>1cm, >5 pcs)	Red block	Monitor carefully
Larvae (>1cm, <5 pcs)	Green block	No need to treat

Table 2
The data used in this study

Data layers	Data sources	Year
Bagworm Infectious Level	Official data-KLK Plantation	April 2020 to Dec 2020. Jan 2021 to July 2021; 2022
LST / (℃)	MODIS (MOD11A2)	2020/2021/2022
RF (mm/8days)	GPM	2020/2021/2022
RH	MODIS (MOD05/06/07)	2020/2021/2022
NDVI	MODIS (MOD13A1)	2020/2021/2022
Road density (km/km²)	Official data-KLK Plantation	2022
Elevation / m	ASTER GDEM V3	2022
Slope /（ ）	ASTER GDEM V3	2022

Data Preprocessing

The bagworm infestation data were applied to a Space Time Cube (STC) model to statistically analyse the spatio-temporal pattern of each plot. The STC was constructed using the ArcGIS Pro tool. The constructed STC is based on an 8-day unit to align with the satellite-based data cycle of bioclimatic data, with the data representing *Metisa plana* in each plot. Therefore, the severity of infected plots, from April 2020 to July 2021 and the entire year of 2022, within 8 days, can be assessed. The final dataset comprises 5,208 records of pest infection at 124 time points across 42 plots, with 842 records containing valid data and the remaining 4,366 records being missing or invalid. The main causes of this data missingness are that the bagworm census data were collected manually (Tailliez & Ballo Koffi, 1992). Ideally, bagworm counts should be collected and recorded regularly in both time and space. However, as the spatial and temporal coverage of the survey expands, the workload and difficulty of data collection increase exponentially, making it inevitable that some data will be missing. As this research concentrates on the relationship between the presence of bagworms and habitat factors, rather than on the relationship between the absence of bagworm data and habitat factors, here we exclude the missing parts.

The spatio-temporal distribution of bagworm infestation severity is presented in the form of a heat map as shown in Figure 3. The vertical axis on the left represents different plot numbers, the horizontal axis at the bottom represents time, and the colour band on the right, ranging from yellow to green, indicates the increasing severity of bagworm infestation.

In this study, the dependent variable, "bagworm infectious level" data, was standardised and converted into binary categories, represented by 0 and 1 to indicate the absence or presence of bagworms. Another reason for this approach is that data collection on bagworm infection levels in the lekir study area has not been uniformly conducted over time and space. Consequently, the spatio-temporal cube (data points aggregated within a time step interval) cannot distinguish the degree of infestation but can only distinguish the presence or absence of bagworm infestation.

Besides, the bioclimatic independent variables (i.e. bioclimatic data) also need to be pre-processed and normalised, so that they can keep and transform their temporal resolution into the exact resolution, then the model can more easily interpret the data and identify potential features within it.

In terms of spatial scale, all bioclimatic data (raster format) were resampled using bilinear interpolation to increase their spatial resolution, ensuring that it is finer than the management units of the study area. Zonal statistical analysis was then applied to ensure spatial alignment for subsequent analyses. Thus, the responses of environmental variables identified in this model should be interpreted as indicative of plot-level averages pertinent to plantation management.

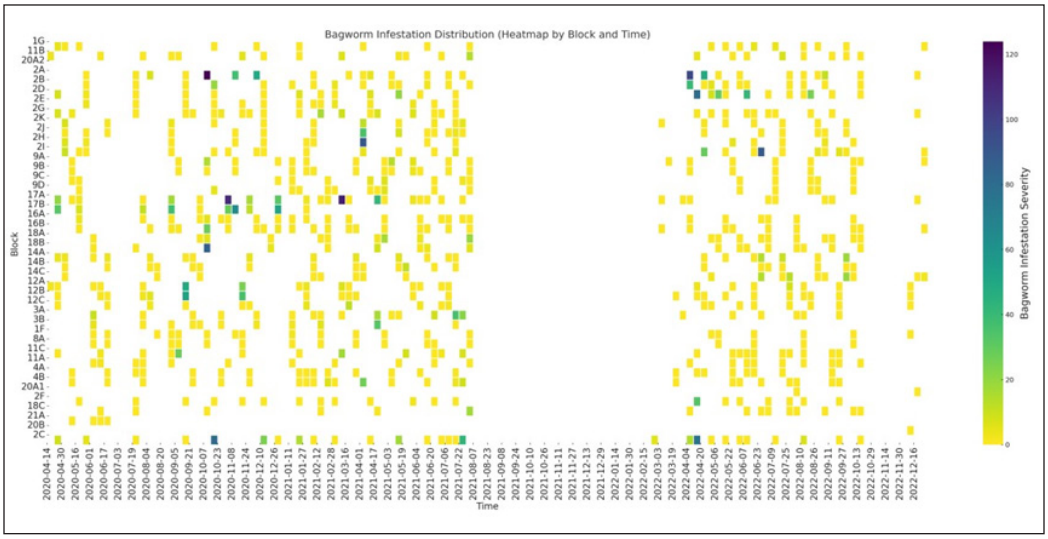
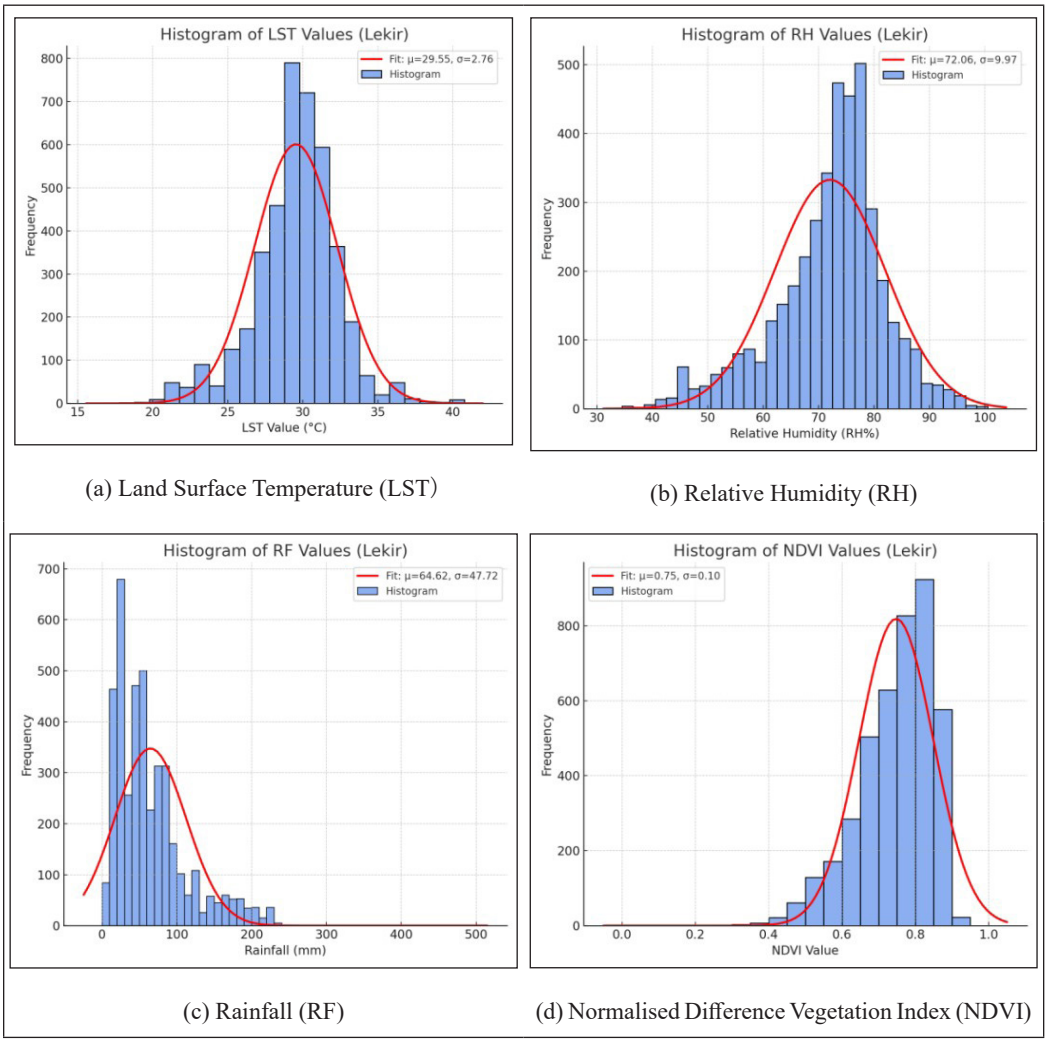


Figure 3. Heat map of bagworm infestation severity

After outliers are removed, the weather data are summarised per plot per 8-day intervals, and the frequency of the data range for each plot every 8 days is calculated. This yields the frequency distribution histogram (Figure 4). In addition, the data for each plot every 8 days are summarised in terms of mean, standard deviation, maximum value, and minimum value, resulting in the weather data statistics for the Lekir study area (Table 3).

In the pest and disease spatial-temporal prediction work, these two contexts - the time lag and time windows- should also be considered, as the effects of certain environmental factors on the bagworm at the current moment do not immediately manifest but occur after a period.



(a) Land Surface Temperature (LST)

(b) Relative Humidity (RH)

(c) Rainfall (RF)

(d) Normalised Difference Vegetation Index (NDVI)

Figure 4. Frequency distribution histograms of (a) LST; (b) RH; (c) RF; and (d) NDVI in Lekir

Table 3
The statistics of Lekir weather data

Weather data	Average	Standard deviation	Max	Min
LST	29.55℃	2.76℃	40.73℃	16.78℃
NDVI	0.75	0.10	0.93	0.22
RH	72.06%	9.97%	99.82%	34.52%
RF	64.62 mm	47.72 mm	232.91 mm	3.08 mm

Note. LST = land surface temperature; NDVI = normalised difference vegetation index; RH = relative humidity; RF = rainfall

In this study, a time lag (Lag T) feature represents the value of an environmental variable at a specific time period in the past (Ravljen et al., 2018), relative to the observation period of the bagworm census. This present study defined the time lag as Equation 1:

$$X_{lag,k}(T) = X(T - k * \Delta t) \quad [1]$$

Where X is an environmental variable (e.g., RH, LST), T is the time of the bagworm census, k is the lag number (1,2,3…), and Δt is the time step (in our case, 8 days interval). While the time window (N cycle window) refers to including multiple consecutive lagged features to form a time window, thereby capturing recent historical conditions (Ravljen et al., 2018). In this study, it is defined as the set of features with Equation 2:

$$Window_N(T) = \{X(T - \Delta t), X(T - 2\Delta t), \cdots, X(T - N\Delta t)\} \quad [2]$$

Where N is the window size. In this research, a "3-cycle window" for variable X means including 3 individual lagged environmental variables (X_Lag1, X_Lag2, X_Lag3) as stacked, independent predictors in the model.

With an 8-day interval, here we pair the feature values of LST/NDVI/RH/RF lagged by 1 to 6 cycles with bagworm statistics based on the time relationship. After data processing, the data used for modelling are presented in the form of a table, including the bagworm data and other bioclimatic data components, all of which contain variable values with 1 to 6 cycle lags. The bioclimatic variables are shown in Table 4. After removing the data of invalid bagworm statistical periods, we finally get a total of 784 samples.

Table 4
Independent variables (bioclimatic data) in the experiment

Independent variable	Current [a]	Lag T [b]					
		1	2	3	4	5	6
LST	√	√	√	√	√	√	√
RF	√	√	√	√	√	√	√
RH	√	√	√	√	√	√	√
NDVI	√	√	√	√	√	√	√
Road density	√	-	-	-	-	-	-
Elevation	√	-	-	-	-	-	-
Slope	√	-	-	-	-	-	-

Note. [a] Current = the period of the bagworm census [b]Lag T = the period that T (=1,2,3,4,5,6) cycle before the bagworm census period. For example, Lag1 indicates the period that is 1 cycle before the bagworm census period

Model Construction

The model construction employed the Random Forest algorithm (Cutler et al., 2012), a supervised learning approach that integrates multiple decision trees to generate predictions based on the collective output of all individual models. This ensemble method utilises a resampling technique known as "bagging," in which each decision tree is independently constructed using bootstrap samples drawn randomly from both the dataset and the feature space (Cutler et al., 2012; Fox et al., 2017). The introduction of randomness significantly reduces the correlation among constituent trees, thereby minimising the risk of overfitting and enhancing generalisation performance (Guo et al., 2020). As the number of diverse and independent trees increases, the predictive accuracy of the ensemble improves accordingly. Furthermore, the Random Forest algorithm demonstrates robustness and effectiveness in scenarios involving high-dimensional data with relatively few informative features, owing to its reliance on stochastic sampling with replacement (Salman et al., 2024). Figure 5 shows the workflow of data processing and random forest model training.

In the process of model development, four critical hyperparameters were optimized to enhance predictive performance: (i) the maximum depth of each decision tree (max_depth), (ii) the minimum number of samples required to perform an internal node split (min_samples_split), (iii) the minimum number of samples required at a leaf node (min_samples_leaf), and (iv) the total number of trees in the ensemble (n_estimators) (A Ilemobayo et al., 2024). To systematically identify the optimal parameter configuration, grid search was implemented in conjunction with k-fold cross-validation, ensuring a balance between model complexity and generalisation capability.

To prevent the influence of feature variable values and scales on model building and subsequent analysis, when using the tree-based models, there is a need to normalise the

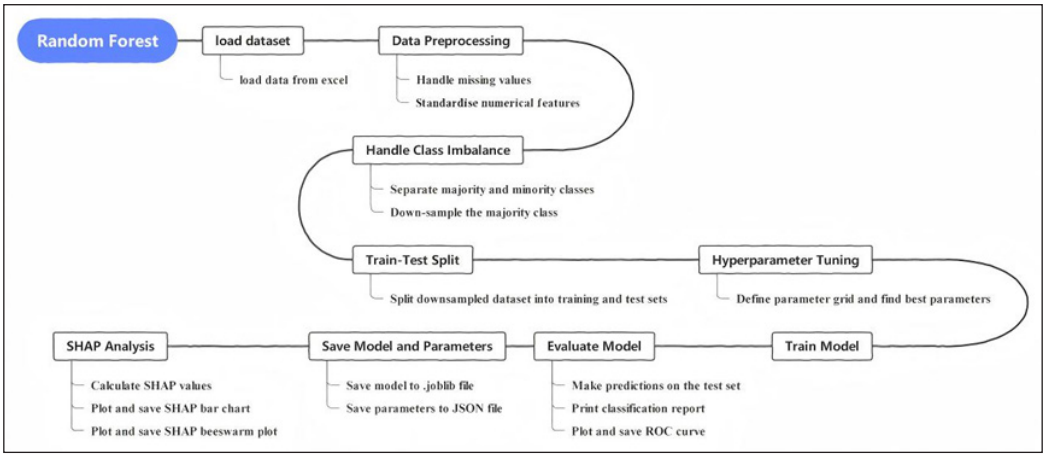


Figure 5. Workflow of the random forest modelling method

feature variables (Rajakumaran et al., 2024). Here we used z-score, which was implemented using the *sklearn.preprocessing.StandardScaler* function in Python. This method adjusts the distribution of each feature variable to a standard normal distribution with a mean of 0 and a standard deviation of 1.

At the same time, we matched the dependent variable and independent variables for different experimental schemes, as shown in the following Table 5. The data were divided into three different time window lengths, corresponding to 1 cycle, 2 cycles, and 3 cycles, to examine the temporal dynamics of environmental factors influencing bagworm infestation. Within each time window length, four different levels of time lag were incorporated to capture the delayed effects of environmental variables. The specific time window and time lag combination is shown in Figure 6.

Finally, the experimental design includes 12 experimental schemes. Considering that lagged environmental variables may exhibit correlations, we quantified these correlations to ensure the robustness of model evaluation. Specifically, for the environmental variables (including all lag periods) involved in the 12 experimental schemes, we calculated the Pearson correlation coefficient matrices and variance inflation factors (VIFs), as shown in Figure 7 and Figure 8, respectively. The results show that, although there is some correlation among different lag periods of the same variable, their VIF values all fall within acceptable thresholds (all less than 5), indicating a low risk of severe multicollinearity. Therefore, we ultimately retained all lagged variables to capture the complete temporal dynamics.

All variables were then matched with the bagworm census target variable based on the time and plot attributes to ensure accurate alignment between the feature variables and the target labels for model training and evaluation. Then the data were divided into training sets and test sets, with 80% and 20%, respectively (Raschka, 2018).

Table 5
Experiment design schemes

Time window length [a]	Time lag [b]	Variable used
1 cycle	1 lag	Lag 1
	2 lag	Lag2
	3 lag	Lag3
	4 lag	Lag4
2 cycle	1 lag	Lag 1-2
	2 lag	Lag 2-3
	3 lag	Lag 3-4
	4 lag	Lag 4-5
3 cycle	1 lag	Lag1-3
	2 lag	Lag2-4
	3 lag	Lag3-5
	4 lag	Lag4-6

Note. [a] Time window length means the number of cycles of environmental conditions used in model building [b] Time lag means the time delay between a statistic in environmental conditions and the resulting impact on pest populations

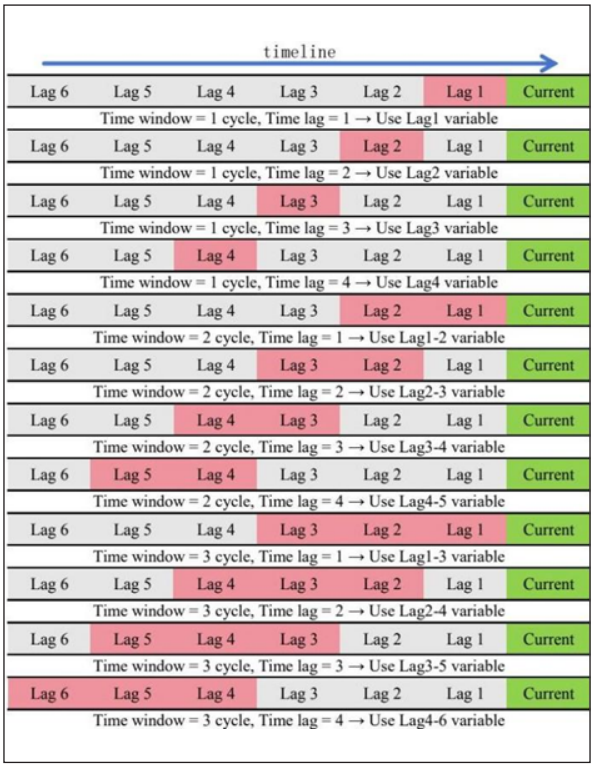


Figure 6. Time window & time lag for bagworm prediction

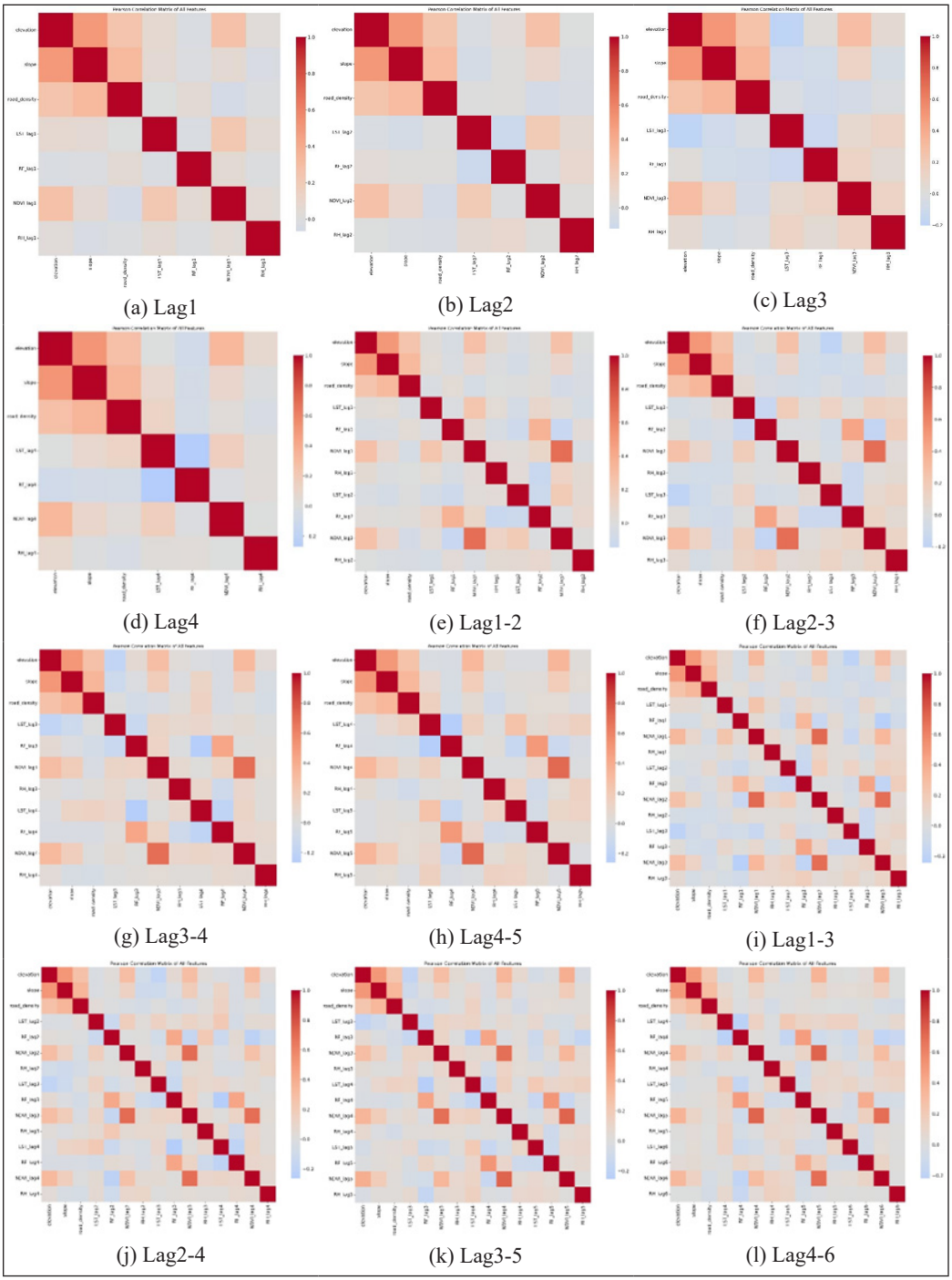


Figure 7. Pearson correlation coefficient matrices of 12 experiment schemes: (a) Lag1; (b) Lag2; (c) Lag3; (d) Lag4; (e) Lag1-2; (f) Lag2-3; (g) Lag3-4; (h) Lag4-5; (i) Lag1-3; (j) Lag2-4; (k) Lag3-5; (l) Lag4-6

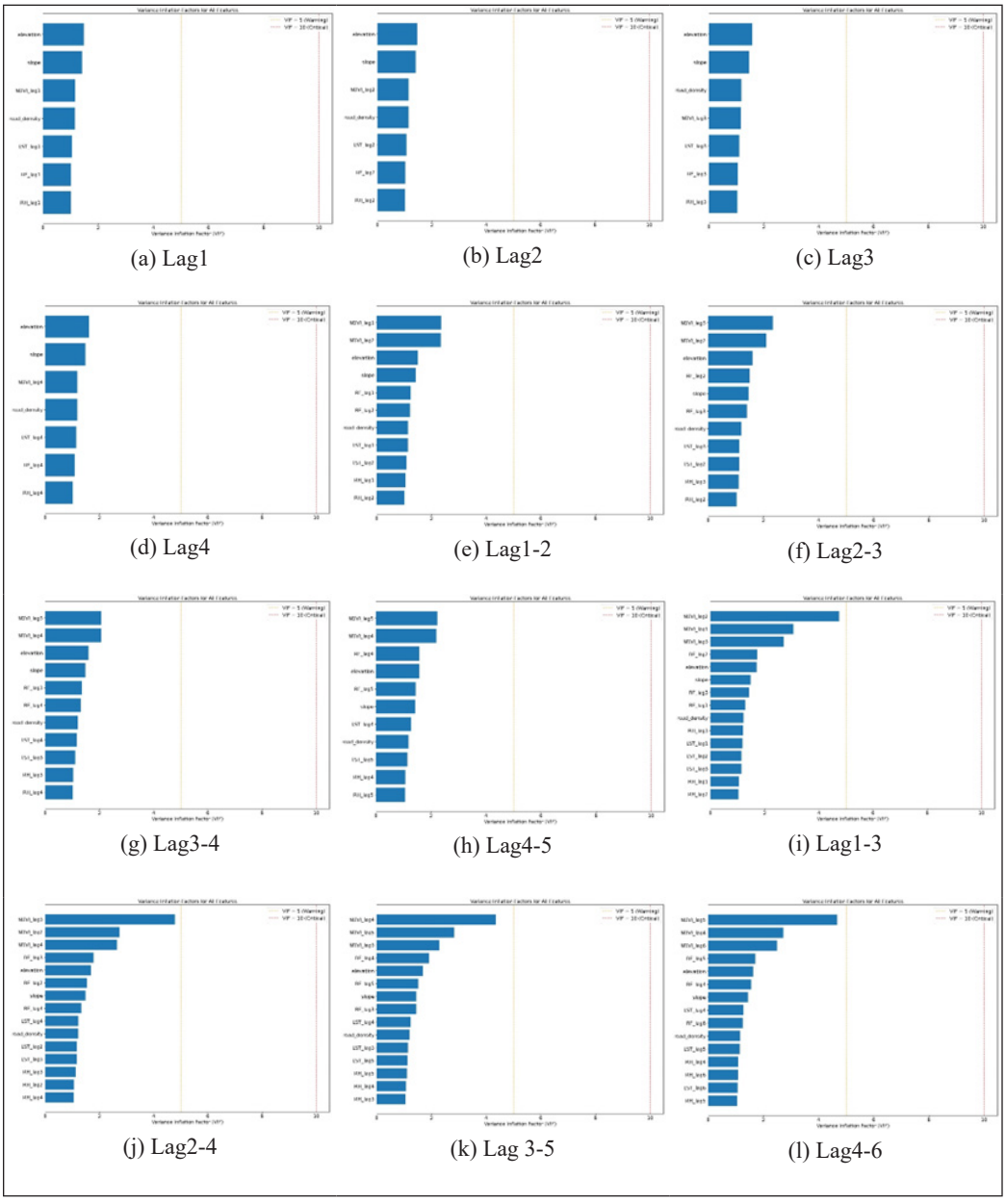


Figure 8. Variance inflation factors (VIFs) of 12 experiment schemes: (a) Lag1; (b) Lag2; (c) Lag3; (d) Lag4; (e) Lag1-2; (f) Lag2-3; (g) Lag3-4; (h) Lag4-5; (i) Lag1-3; (j) Lag2-4; (k) Lag3-5; (l) Lag4-6

Considering that the initial data has a time attribute, i.e., the data is time-sequenced, dividing the training set and test set according to the time sequence is a more reasonable choice. This can ensure that the model training and evaluation process conforms to the actual logic of time series analysis and avoids data leakage (Cerqueira et al., 2020).

We used 80% of the data with earlier time stamps as the training set and 20% of the data with later time stamps as the test set. Within the 80% training set, we adopted K-fold cross-validation (Wong & Yeh, 2020) to robustly tune the model parameters (K=5 in this study), and to get reliable estimates of the model performance. This method involved splitting the training set into K consecutive folds and iteratively training the model on K-1 folds, while using the remaining fold for validation, and finally averaging the results to guide model parameter selection. Then the remaining 20% data were used to test and make a prediction, thus obtaining the final evaluation. The final model represents the contribution of feature variables to the target variable by calculating SHAP values and thereby indicates the importance of features.

Evaluation Metrics

In the model evaluation process, we adopt multiple methods, including precision, recall and F1-score, and also use the ROC-AUC value (Marković et al., 2021; Naidu et al., 2023).

Precision: Precision represents the number of samples with a true positive value among those predicted as positive. In this research, it refers to the number of samples that truly have bagworms among those predicted to have them. It's calculated with the Equation 3:

$$\text{Precision} = \frac{TP}{TP + FP} \quad [3]$$

Recall: Recall is employed to gauge the proportion of positive instances that the model successfully predicts as positive among all the positive samples. The higher the recall, the stronger the model's ability to recognise positive instances. It's calculated with Equation 4:

$$\text{Recall} = \frac{TP}{TP + F\text{N}} \quad [4]$$

In the field of agricultural pest infestation prediction, recall is particularly important. The primary goal is to accurately identify plots with pest infestations, minimising the risk of overlooking any actual occurrences of pests. Missing an infestation could lead to severe crop damage, economic losses, and the rapid spread of pests to surrounding areas.

F1-score: The F1-score is the harmonic mean of precision and recall, comprehensively considering the accuracy and identification ability of the model. The higher the F1-score, the better the comprehensive performance of the model. It's calculated with Equation 5:

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{Recall}}{\text{precision} + \text{Recall}} \quad [5]$$

ROC-AUC: This metric stands for "Area Under the Receiver Operating Characteristic Curve", which is a performance metric used to evaluate the quality of binary classification models(Fawcett, 2006; Hanley & McNeil, 1982). This quantitative indicator ranges from 0.5 to 1, summarising the overall ability of the classifier to rank positive examples ahead of negative ones across all possible decision thresholds. We compute it by using the standard implementation in *scikit-learn* (Pedregosa et al., 2011).

RESULTS AND DISCUSSION

This study develop an interpretable prediction model and quantify the time lag effects of environmental variables, then identify the key drivers for bagworm occurrence. The following section provides an analysis of model performance and feature importance.

Model Performance

Performance comparison across different time window configurations reveals distinct patterns in model effectiveness. As has been mentioned before, the time window refers to the length of segments of consecutive time steps used to structure environmental variables data for prediction, while time lag means the time between the response of pest emergence and environmental fluctuation. The experiment results are summarised in Table 6.

Among the 3-time window schemes, the Lag1-3 configuration demonstrates superior performance with a recall rate of 0.76 and an ROC-AUC value of 0.68, indicating its relatively high accuracy in predicting bagworm occurrences. For the 2-time window setting, the Lag1-2 configuration achieves the highest recall (0.71) and an ROC-AUC of 0.66. In contrast, under the single time window condition, the Lag2 configuration yields the best recall rate of 0.71 and an ROC-AUC of 0.74. Overall, the random forest model performs most effectively under the Lag1-3 configuration with three-time windows, followed by the Lag2 configuration under a single time window.

Feature Importance Analysis

Given that the random forest model achieved the highest predictive performance under the Lag1-3-time condition, we selected this configuration for further feature importance analysis. SHAP values were calculated separately for each individual lagged feature variable within the Lag1-3 window, and corresponding visualisations were generated, including a standard bar chart (Figure 9) and a beeswarm plot (Figure 10). In Figure 9, each bar in the bar plot represents the mean absolute SHAP value for a specific feature, and in Figure 10, each point in the beeswarm plot represents a SHAP value for a specific feature for a single record in the test set.

Table 6
Experiment result

Evaluation index	Precision	Recall	F1-score	ROC-AUC
Lag1-3	0.31	0.76	0.44	0.68
Lag2-4	0.28	0.6	0.38	0.6
Lag3-5	0.23	0.45	0.31	0.51
Lag4-6	0.32	0.69	0.44	0.66
Lag1-2	0.3	0.71	0.42	0.66
Lag2-3	0.23	0.55	0.32	0.59
Lag3-4	0.28	0.63	0.38	0.56
Lag4-5	0.33	0.7	0.45	0.62
Lag1	0.31	0.68	0.42	0.64
Lag2	0.34	0.71	0.46	0.74
Lag3	0.25	0.52	0.33	0.58
Lag4	0.3	0.58	0.4	0.63
Maximum	0.34	0.76	0.46	0.74
Average	0.29	0.63	0.40	0.62

As illustrated in Figures 9 and 10, RH_lag2 and LST_lag1 emerged as the two most influential predictors of bagworm outbreaks, ranked first and second in terms of feature importance. The swarm plot reveals nonlinear relationships between habitat-related variables and outbreak likelihood. Specifically, RH_lag2 (relative humidity from two periods prior) exhibits a strong negative influence on bagworm occurrence. When RH_lag2 is below 60% (indicated by dense blue dots), the risk of an outbreak two periods later increases (positive SHAP values), which aligns with the biological observation that bagworm larvae hatch more readily under low-humidity conditions (Napiah et al., 2023). Conversely, when RH_lag2 exceeds 75%, SHAP values become negative, suggesting that high humidity may suppress adult activity. LST_lag1 (land surface temperature from one period prior) demonstrates optimal predictive power within the range of 28-32°C, where SHAP values peak, indicating favourable conditions for bagworm development. Outside this range (either too high or too low), the predictive strength diminishes, implying less favourable conditions for outbreaks one period later, which follows the conclusion of publication before (Enting & Latip, 2021). Road density ranks third in feature importance and shows notable influence. It is calculated as the total road length within a block divided by the block area and is used as a feature variable to represent the road dust level. The insecticidal effect of inert dusts like road dust and powdered clay has long been known; the mode of action is from its desiccant and adsorptive functions (Ebeling, 1971; Stejskal et al., 2021). However, according to Ho. (2002), the bagworm can survive on high fronds of tall palms.

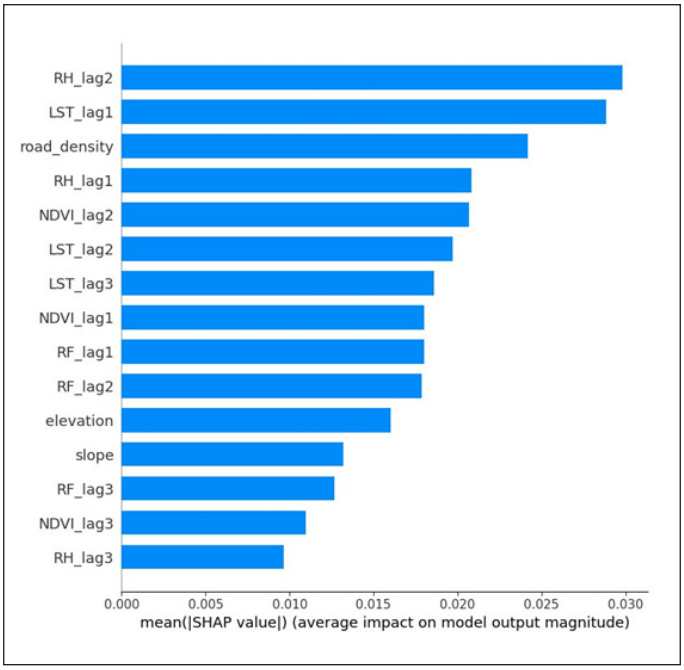


Figure 9. Bar plot of each variable SHAP values

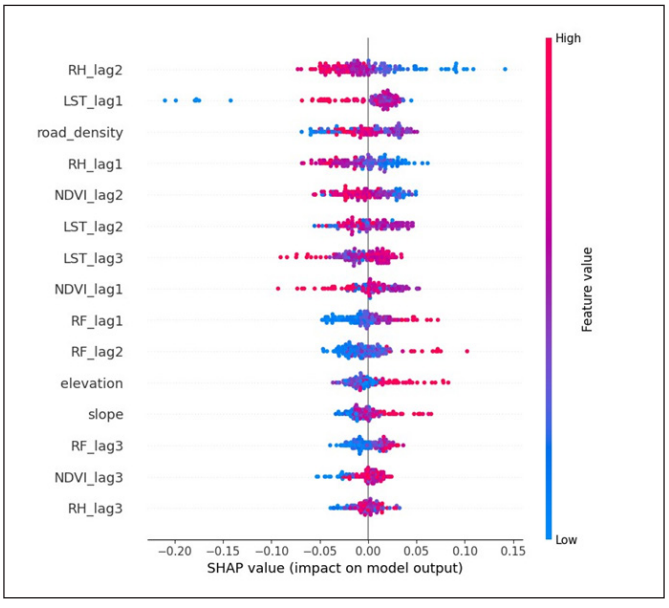


Figure 10. Beeswarm plot of each variable SHAP values (the colour ranging from blue to red corresponds to the magnitude of the independent feature value)

Its SHAP distribution lacks a consistent directional trend, which suggests that the relationship between road density and bagworm outbreaks varies spatially, potentially reflecting differences in land management intensity. RH_lag1 (relative humidity from one cycle prior) displays a nonlinear pattern like RH_lag2. RF_lag1, RF_lag2, and RF_lag3 exhibit relatively lower feature importance overall, indicating the impact of rainfall on the population of bagworm is very limited and is not the main driving factor (Napiah et al., 2023). They share a consistent positive influence mechanism whereby increased precipitation across lagged periods correlates with higher likelihoods of bagworm outbreaks 1-3 periods later.

Temporal and Spatial Patterns

The data from the Lekir experimental zones were partitioned into two distinct subsets. To prevent data leakage, we implemented a chronological split method, wherein the initial 80% of the dataset was allocated for model training and validation, while the remaining 20% from a later time was reserved for prediction testing. Subsequently, we visualised the performance of the Random Forest model on the test set. The results are presented in Figure 11. In the visualisation, red-marked plots indicate areas predicted to experience bagworm outbreaks, green plots represent areas predicted to remain unaffected, and blank-coloured plots correspond to regions where no data were available in the test dataset for that specific period.

This study achieved an interpretable prediction of bagworm pest outbreaks in palm oil plantations by integrating multi-temporal remote sensing satellite data with the random forest model. Key findings revealed that lagged effects and nonlinear responses of environmental variables (such as relative humidity (RH), land surface temperature (LST), normalised difference vegetation index (NDVI), and rainfall (RF)) provided novel insights into the mechanisms underlying bagworm outbreaks. Compared to traditional models that rely solely on instantaneous meteorological data or averaged environmental variables over fixed periods (Gachoki et al., 2024; Lee et al., 2023; Mahmoodi et al., 2024) The incorporation of time-lagged environmental factors aligns more closely with the biological and temporal logic of pest development.

The SHAP analysis further uncovered significant nonlinear threshold effects of key predictors. Specifically, when RH_lag2 fell below 60%, the risk of outbreak increased markedly, which is consistent with the biological requirement of bagworm eggs for dry conditions during hatching (Napiah et al., 2023). Additionally, LST_lag1 exhibited an optimal temperature window of 28-32°C, within which the likelihood of outbreak peaked. Deviations from this range resulted in reduced SHAP values, indicating less favourable conditions for bagworm development, which corroborates laboratory observations of larval metabolic activity (Enting & Latip, 2021).

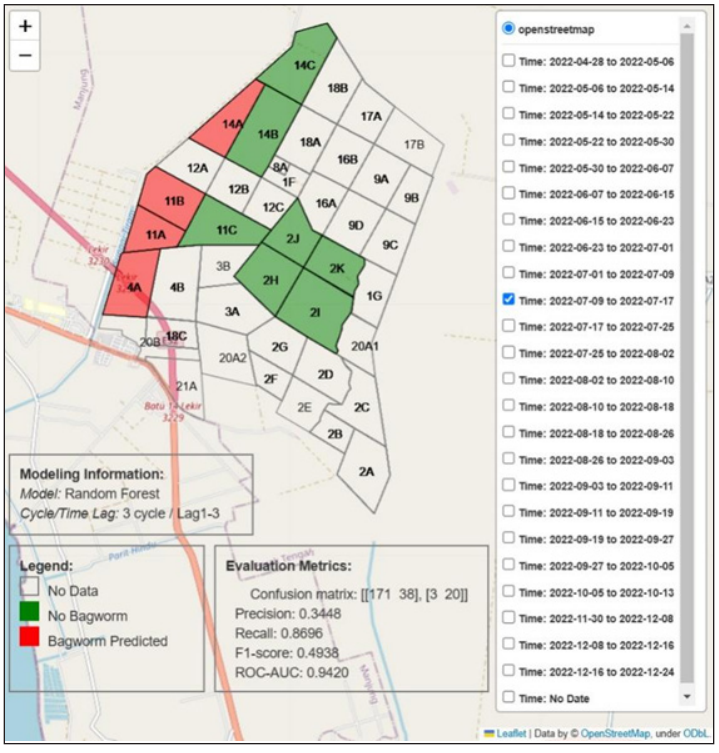


Figure 11. Visualisation of random forest prediction result on test data (randomly choose 2022-07-09 to 2022-07-17 as an example)

The application of SHAP analysis in this research played a crucial role in shifting from correlational insights to explanatory insights. Unlike traditional feature importance measures in Random Forests (such as Gini importance), which only reflect the overall contribution of a variable (Archer & Kimes, 2008; Loecher, 2022). SHAP values quantify both the direction and magnitude of each feature's influence on every individual prediction. This enables us to identify specific environmental thresholds that trigger high risk. Although traditional logistic regression models are interpretable (Carvalho et al., 2019), but they typically assume linear relationships and would struggle to reveal such complex, threshold-driven interactions without pre-specifying interaction terms.

In contrast to conventional pest models that focus exclusively on climatic variables (Napiah et al., 2023), this study identified road density as a non-climatic factor with substantial predictive power, highlighting the ecological implications of agricultural infrastructure. Prior research has indicated that dust generated by roads can influence bagworm survival (Ebeling, 1971; Ho, 2002; Lever, 2008). Our findings quantify this relationship, revealing complex but meaningful patterns in SHAP value distributions. These suggest potential management-related influences; for instance, high road density may

facilitate rapid pesticide application, whereas low road density may support parasitoid wasp populations. Future studies could explore the multilevel interactions between 'landscape structure, microclimate, and pest outbreaks' to deepen understanding of these ecological linkages.

From a practical perspective, the developed model can be operationalised into a practical application. For example, when habitat conditions such as RH_lag2 < 60% and LST_lag1 > 28°C are met, a high-risk alert is triggered. Subsequently, control strategies can be tailored based on road density, like drone-based spraying in low-density areas. High-risk spatial predictions can also guide targeted interventions, for instance, by integrating the spatio-temporal predictions shown in Figure 8 to reduce the use of pesticides, thereby directly supporting sustainable development goals in Malaysia's palm oil industry (Parveez et al., 2021).

Despite its contributions, this study has several limitations. The pest dataset used in this research contains a total of 5,208 records after preprocessing, of which 842 records contain valid data, while the remaining 4,366 records are missing or invalid, which means that nearly 84% of the records are unusable. Such a high proportion of missing data makes it likely that the valid subset cannot fully represent the conditions of the entire area. As a result, the trained model may inherently exhibit systematic selection bias (Phillips et al., 2009), with poor spatial generalisation: its predictions are most reliable for areas that resemble the well-monitored subset used for training, whereas greater uncertainty exists for areas with insufficient monitoring.

The time-series split approach adopted during validation to assess predictive performance on unseen time periods within the same spatial domain provides an evaluation of temporal generalisation to some extent (Cerqueira et al., 2020). However, its applicability to other regions or future climate scenarios requires additional scrutiny.

In terms of data, other factors not considered in this study include: pest occurrence was treated as a binary classification problem rather than a regression task to reflect infestation intensity; the bagworm pests are also not differentiated by species or sex (Loong & Chong, 2012); other potential independent variables such as wind direction, speed (Rhainds et al., 2002) and solar radiation is not included in this study; the influence of soil type and other static environmental variables was not controlled; finally, the landscape structure data used were static and did not account for the impact of temporary agricultural paths, which may affect local dispersal patterns of bagworms.

Future research will specifically concentrate on assessing the model's spatial scalability by implementing it in plantations with varying management practices and microclimates, as well as its temporal scalability by integrating multi-year data to evaluate its resilience under changing climatic conditions. Another direction for future work is to systematically compare the Random Forest model used here with other advanced algorithms (such as XGBoost, LSTM).

Such studies would require larger datasets to ensure reliable training and validation and would further refine the way managers select the most appropriate predictive tools in application.

CONCLUSION

This study constructed a spatio-temporal prediction model for palm oil bagworm by integrating remote sensing imagery with the random forest algorithm. The model demonstrated effective performance in predicting the spatial and temporal patterns of bagworm outbreaks, achieving a maximum recall rate of 0.76. Visualisation of the prediction results enables targeted management of high-risk areas, thereby enhancing the practical applicability of the model. These outcomes contribute to improving pest control efficiency and support the sustainable development goals of Malaysia's palm oil industry. Future research could incorporate additional habitat-related variables, such as soil type, to further refine model accuracy. Moreover, integrating weather forecasting data into predictive modelling and exploring advanced ensemble techniques may enhance the robustness and generalisability of the prediction framework.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support from the Ministry of Education Malaysia under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/WAB04/UTM/02/2) and Universiti Teknologi Malaysia under R.J130000.7852.5F485. The authors also thanked Lekir Plantation, Kuala Lumpur Kepong Berhad, for the data used in this study. We also acknowledge the insight of this study from Mr Muhammad Idrus Bin Shukor.

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
FN	:	False negative
FP	:	False positive
GIS	:	Geographic information system
GPM	:	Global precipitation measurement
GPS	:	Global positioning system
IoT	:	Internet of Things
LiDAR	:	Light detection and ranging
LST	:	Land surface temperature
LSTM	:	Long short-term memory
MODIS	:	Moderate resolution imaging spectroradiometer
MPOB	:	Malaysian Palm Oil Board
NDVI	:	Normalised difference vegetation index
RF	:	Rainfall
RH	:	Relative humidity

ROC-AUC : Area under the receiver operating characteristic curve
SDG : Sustainable development goal
SHAP : Shapley additive exPlanations
STC : Space-time cube
TP : True positive
VIF : Variance inflation factor
XGBoost : eXtreme gradient boosting

REFERENCES

A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., & Ifeanyi, A. (2024). Hyperparameter tuning in machine learning: A comprehensive review. *Journal of Engineering Research and Reports, 26*(6), 388-395.https://doi.org/10.9734/jerr/2024/v26i61188

Archer, K. J., & Kimes, R. V. (2008). Empirical characterisation of random forest variable importance measures. *Computational Statistics & Data Analysis, 52*(4), 2249-2260. https://doi.org/10.1016/j.csda.2007.08.015

Balasundram, S. K., Shamshiri, R. R., Sridhara, S., & Rizan, N. (2023). The role of digital agriculture in mitigating climate change and ensuring food security: An overview. *Sustainability, 15*(6), 5325. https://doi.org/10.3390/su15065325

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832. https://doi.org/10.3390/electronics8080832

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning, 109*(11), 1997-2028. https://doi.org/10.1007/s10994-020-05910-7

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 157-175). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_5

Didan, K. (2021). MODIS/Terra vegetation indices 16 days L3 global 250 m SIN/Grid Version V061 [Data set]. https://doi.org/10.5067/MODIS/MOD13Q1.061

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., & Morgan, G. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys, 55*(9), 1-33. https://doi.org/10.1145/3561048

Ebeling, W. (1971). Sorptive dusts for pest control. *Annual Review of Entomology, 16*(1), 123-158. https://doi.org/10.1146/annurev.en.16.010171.001011

Enting, C. E., & Latip, S. N. H. M. (2021). Life cycle of palm oil bagworm, *Metisa plana* Walker (Lepidoptera: Psychidae) at different temperatures under controlled environment. *Serangga, 26*(2), 151-165.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., & Weber, M. H. (2017). Assessing the accuracy and stability of variable selection methods for random forest modelling in ecology. *Environmental Monitoring and Assessment, 189*(7), 316. https://doi.org/10.1007/s10661-017-6025-0

Gachoki, S., Groen, T. A., Vrieling, A., Skidmore, A., & Masiga, D. (2024). Towards accurate spatial prediction of *Glossina pallidipes* relative densities at country-scale in Kenya. *Ecological Informatics, 81*, Article 102610. https://doi.org/10.1016/j.ecoinf.2024.102610

Guo, X., Bian, Z., Wang, S., Wang, Q., Zhang, Y., Zhou, J., & Lin, L. (2020). Prediction of the spatial distribution of soil arthropods using a random forest model: A case study in Changtu County, Northeast China. *Agriculture, Ecosystems & Environment, 292*, 106818. https://doi.org/10.1016/j.agee.2020.106818

Hamer, W. B., Birr, T., Verreet, J.-A., Duttmann, R., & Klink, H. (2020). Spatio-temporal prediction of the epidemic spread of dangerous pathogens using machine learning methods. *ISPRS International Journal of Geo-Information, 9*(1), 44. https://doi.org/10.3390/ijgi9010044

Ho, C. T. (2002). Ecological studies on Pteroma pendula Joannis and Metisa plana Walker (Lepidoptera: Psychidae) towards improved integrated management of infestations in palm oil [Doctoral dissertation, Universiti Putra Malaysia]. Universiti Putra Malaysia Publications. https://core.ac.uk/download/pdf/43000616.pdf

Holloway, P., Kudenko, D., & Bell, J. R. (2018). Dynamic selection of environmental variables to improve the prediction of aphid phenology: A machine learning approach. *Ecological Indicators, 88*, 512-521.https://doi.org/10.1016/j.ecolind.2017.10.032

Hoong, H. W., & Hoh, C. (1992). Major pests of palm oil and their occurrence in Sabah. *Planter, 68*(793), 193-210.

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2023). *GPM IMERG final precipitation L3 1 day 0.1 degree × 0.1 degree Version V07* [Data set]. Goddard Earth Sciences Data and Information Services Centre (GES DISC). https://doi.org/10.5067/GPM/IMERGDF/DAY/07

Kamarudin, N., Ali, S. R. A., Masri, M. M. M., Ahmad, M. N., Manan, C. A. H. C., & Kamarudin, N. (2017). Controlling *Metisa plana* Walker (Lepidoptera: Psychidae) outbreak using *Bacillus thuringiensis* at an palm oil plantation in Slim River, Perak, Malaysia. *Journal of Palm oil Research, 29*(1), 47-54. https://doi.org/10.21894/jopr.2017.2901.05

Kamarudin, N., & Mohd, B. W. (2007). Status of common palm oil insect pests in relation to technology adoption. *Planter, 83*(975), 371-385.https://doi.org/10.56333/tp.2007.005

Lee, D. S., Lee, T. G., Bae, Y. S., & Park, Y. S. (2023). Occurrence prediction of western conifer seed bug (*Leptoglossus occidentalis*: Coreidae) and evaluation of the effects of climate change on its distribution in South Korea using machine learning methods. *Forests, 14*(1), 14. https://doi.org/10.3390/f14010117

Lever, R. J. A. W. (2008). Pests of palm oils in Malaysia and their control by Brian J. Wood, Kuala Lumpur: Incorporated Society of Planters (1968), pp. 204, $25 Malaysian. *Experimental Agriculture, 6*(1), 79-80. https://doi.org/10.1017/S0014479700005950

Li, J.-Y., Chen, Y.-T., Shi, M.-Z., Li, J.-W., Xu, R.-B., Pozsgai, G., & You, M.-S. (2021). Spatio-temporal distribution patterns of *Plutella xylostella* (Lepidoptera: Plutellidae) in a fine-scale agricultural landscape based on geostatistical analysis. *Scientific Reports, 11*(1), 13622. https://doi.org/10.1038/s41598-021-92562-9

Loecher, M. (2022). Unbiased variable importance for random forests. *Communications in Statistics-Theory and Methods, 51*(5), 1413-1425. https://doi.org/10.1080/03610926.2020.1764042

Loong, C., & Chong, T. (2012). Understanding pest biology and behaviour for effective control of palm oil bagworms. The Planter, 88(1039), 699-715.https://doi.org/10.56333/tp.2012.008

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualised feature attribution for tree ensembles. *arXiv* (arXiv:1802.03888). https://doi.org/10.48550/arXiv.1802.03888

Mahmoodi, S., Ganje, M. B., Ahmadi, K., Dalvand, Y., Naghibi, A., & Newlands, N. K. (2024). Modelling spatiotemporal distribution of yellow rust wheat pathogen using machine learning algorithms: Insights from environmental assessment. *Environmental Technology & Innovation, 36*, 103865.https://doi.org/10.1016/j.eti.2024.103865

Makori, D. M., Abdel-Rahman, E. M., Odindi, J., Mutanga, O., Landmann, T., & Tonnang, H. E. Z. (2024). Multi-pronged abundance prediction of bee pests' spatial proliferation in Kenya. *International Journal of Applied Earth Observation and Geoinformation, 128*, 103738. https://doi.org/10.1016/j.jag.2024.103738

Marković, D., Vujičić, D., Tanasković, S., Đorđević, B., Ranđić, S., & Stamenković, Z. (2021). Prediction of pest insect appearance using sensors and machine learning. *Sensors, 21*(14), 4846. https://doi.org/10.3390/s21144846

MODIS, A. S. T. (2017a). *MODIS/Terra clouds 5-min L2 swath 1 km and 5 km* [Data set]. NASA LANCE MODIS at the MODAPS. https://doi.org/10.5067/MODIS/MOD06_L2.NRT.061

MODIS, A. S. T. (2017b). *MODIS/Terra temperature and water vapour profiles 5-min L2 swath 5 km* [Data set]. NASA LANCE MODIS at the MODAPS. https://doi.org/10.5067/MODIS/MOD07_L2.NRT.061

MODIS, A. S. T. (2017c). *MODIS/Terra total precipitable water vapour 5-min L2 swath 1 km and 5 km* [Data set]. NASA LANCE MODIS at the MODAPS. https://doi.org/10.5067/MODIS/MOD05_L2.NRT.061

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. In *Artificial Intelligence application in networks and systems*. Springer. https://doi.org/10.1007/978-3-031-35314-7_2

Napiah, N. R. A. M. A., Kamarudin, N., Bakeri, S. A., Zainuddin, N., Keni, M. F., & Masri, M. M. M. (2023). Impact of environmental factors on the larval population of bagworm, *Metisa plana* Walker (Lepidoptera: Psychidae) in palm oil smallholdings. *Serangga, 28*(2), 149-161.

NASA/METI/AIST/Japan, U.S./Japan ASTER Science Team. (2019). *ASTER global digital elevation model version V003* [Data set]. NASA Land Processes Distributed Active Archive Centre.https://doi.org/10.5067/ASTER/ASTGTM.003

Parveez, G. K. A., Tarmizi, A. H. A., Sundram, S., Loh, S. K., Ong-Abdullah, M., Palam, K. D. P., Salleh, K. M., Ishak, S. M., & Idris, Z. (2021). Palm oil economic performance in Malaysia and R&D progress in 2020. *Journal of Palm oil Research, 33*(2), 181-214. https://doi.org/10.21894/jopr.2021.0026

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.https://doi.org/10.48550/arXiv.1201.0490

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications, 19*(1), 181-197. https://doi.org/10.1890/07-2153.1

Pol, M., Bailey, L. D., McLean, N., Rijsdijk, L., Lawson, C. R., Brouwer, L., & Gimenez, O. (2016). Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution, 7*(10), 1246-1257. https://doi.org/10.1111/2041-210X.12590

Rajakumaran, M., Arulselvan, G., Subashree, S., & Sindhuja, R. (2024). Crop yield prediction using multi-attribute weighted tree-based support vector machine. *Measurement: Sensors, 31*, 101002. https://doi.org/10.1016/j.measen.2023.101002

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* (arXiv:1811.12808). https://doi.org/10.48550/arXiv.1811.12808

Ravljen, M., Hovelja, T., & Vavpotič, D. (2018). Immediate, lag and time window effects of meteorological factors on ST-elevation myocardial infarction incidence. *Chronobiology International, 35*(1), 63-71. https://doi.org/10.1080/07420528.2017.1381847

Rhainds, M., Gries, G., Ho, C. T., & Chew, P. S. (2002). Dispersal by bagworm larvae, *Metisa plana*: Effects of population density, larval sex, and host plant attributes. *Ecological Entomology, 27*(2), 204-212.https://doi.org/10.1046/j.1365-2311.2002.00389.x

Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., & Díaz-Rodríguez, N. (2021). Explainable Artificial Intelligence (XAI) on time series data: A survey. *arXiv* (arXiv:2104.00950). https://doi.org/10.48550/arXiv.2104.00950

Rouabah, A., Meiss, H., Villerd, J., Lasserre-Joulin, F., Tosser, V., Chabert, A., & Therond, O. (2022). Predicting the abundances of aphids and their natural enemies in cereal crops: Machine learning versus linear models. *Biological Control, 169*, 104866. https://doi.org/10.1016/j.biocontrol.2022.104866

Sabtu, N., Idris, N., & Ishak, M. (2018). The role of geospatial in plant pests and diseases: An overview. *IOP Conference Series: Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/169/1/012013

Salim, H., & Hamid, N. H. (2012). Evaluation of several chemical control approaches against bagworm, *Metisa plana* Walker (Lepidoptera: Psychidae) in Felda palm oil plantations. *The Planter, 88*, 785-799.https://doi.org/10.56333/tp.2012.009

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning, 2024*, 69-79. https://doi.org/10.58496/BJML/2024/007

Stejskal, V., Vendl, T., Aulicky, R., & Athanassiou, C. (2021). Synthetic and natural insecticides: Gas, liquid, gel and solid formulations for stored-product and food-industry pest control. *Insects, 12*(7), 590. https://doi.org/10.3390/insects12070590

Sulaiman, M. N., & Talip, M. S. A. (2021). Sustainable control of bagworm (Lepidoptera: Psychidae) in palm oil plantation: A review paper. *International Journal of Agriculture, Forestry and Plantation, 11*, 47-55.

Sundaraja, C. S., Hine, D. W., & Lykins, A. D. (2021). Palm oil: Understanding barriers to sustainable consumption. *PLoS ONE, 16*(8), e0254897. https://doi.org/10.1371/journal.pone.0254897

Tailliez, B., & Ballo Koffi, C. (1992). A method for measuring palm oil leaf area. *Oléagineux, 47*(8-9), 537-545.

Thaer, S., Kassim, F. A., Hasbullah, N. A., & Al-Obaidi, J. R. (2021). Evaluation of bagworm, *Metisa plana* (Lepidoptera: Psychidae) infestation and beneficial parasitoid in an palm oil plantation, Perak, Malaysia. *Journal of Science and Mathematics Letters, 9*(1), 19-35. https://doi.org/10.37134/jsml.vol9.1.3.2021

Wahid, M. B., Hassan, H. A. H., & Masijan, Z. (1988). Bagworms (Lepidoptera: Psychidae) of palm oils in Malaysia. *PORIM Occasional Paper, 23*, 1-38.

Wan, Z., Hook, S., & Hulley, G. (2021). *MODIS/Terra land surface temperature/emissivity 8-day L3 global 1 km SIN grid version V061* [Data set]. https://doi.org/10.5067/MODIS/MOD11A2.061

Wong, T.-T., & Yeh, P.-Y. (2020). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering, 32*(8), 1586-1594. https://doi.org/10.1109/TKDE.2019.2912815

Zakaria, K., Hassan, N. A. M., Azam, A. H. M., Salleh, K. M., Kamil, N. N., & Abdullah, N. (2024). Factors affecting export demand for Malaysian palm-based finished products. *Asian Academy of Management Journal, 29*(2), 109-132. https://doi.org/10.21315/aamj2024.29.2.5

Pertanika J. Sci. & Technol. 34 (2): 1053 - 1075 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Design of an Integrated Entropy-based Framework for Dynamic Risk Weight Allocation and Mitigation Strategy Prioritisation in High-speed Rail Projects

Yogesh P. Kherde[1]*, Uday P. Waghe[1], Radhika S. Thakre[1], Rajesh M. Bhagat[1], Sanskruti M. Padmawar[1], Yogeshwari D. Waje[1], Rishi S. Lohia[1], Anup K. Chitkeshwar[2], and Vaibhav Dhawale[3]

[1]Department of Civil Engineering, Yeshwantrao Chavan College of Engineering, 441110 Nagpur, Maharashtra, India
[2]Department of Civil Engineering, P. R. Pote College of Engineering, 444602 Amravati, Maharashtra, India
[3]Department of Civil Engineering, Prof Ram Meghe College of Engineering and Management, 444701 Amravati, India

ABSTRACT

High-Speed Rail (HSR) projects involve complex systems, huge investments and multidimensional risks. Effective risk reduction requires context-sensitive mitigation strategies which are accurate. The traditional entropy-based models are static and based on the subjective judgment of humans, and they lack adaptability in real-time modelling of the complexity of the HSR environment. The paper proposes an innovative integrated entropy-based risk assessment and mitigative framework, which is related specifically to HSR systems and will centre around dynamic gradations of risk weights and mitigation strategy. The framework presented is a series of 5 interconnected methodology blocks. First Dynamic Spatio-Temporal Entropy Weighting (DSTEW) generates its entropy weights, which are real-time adaptable timeline or temporal, e.g. seasonal, operational time-line factors, and spatial, i.e. zone-related variability. For checking the robustness of the procedure, Multi-Entropy Cross Validation (MECV) is used to check the consistency of subjective entropy estimates relating to the various spatial zones and time zones to retain only statistically consistent weight vectors. Entropy-Driven Bayesian Risk Adjustment (EDBRA) then adjusts and modifies these weights by including the history of past risk occurrences by means of Bayesian updating of risk. Subjective uncertainties are also resolved by using Hybrid Fuzzy-Entropy

ARTICLE INFO

Article history:
Received: 24 October 2025
Accepted: 15 December 2025
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.19

E-mail addresses:
ypkherde@ycce.edu (Yogesh P. Kherde)
upwaghe@ycce.edu (Uday P. Waghe)
radhikathakre13@gmail.com (Radhika S. Thakre)
rmbhagat@ycce.edu (Rajesh M. Bhagat)
anupkchitkeshwar@gmail.com (Anup K. Chitkeshwar)
dhawalevr31@gmail.com (Vaibhav Dhawale)
sanskrutipadmawar8@gmail.com (Sanskruti M. Padmawar)
yogeshwariwaje11@gmail.com (Yogeshwari D. Waje)
rishilohiya17@gmail.com (Rishi S. Lohiya)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Multi-Criteria Prioritisation (HFEMCP), which uses the system of fuzzy logic in its consideration of the recalibrated weights and thus generates risk matrices of prioritised risks. Lastly, Entropy Resilient Networks Modelling (ERNM) produces inter-risk influence networks in which mitigation strategies are determined in a global way based on the various centralities which are derived from the entropy weighting procedures. The simulation results feature a quantum leap in improvements that include 10% to 15% improvements in weighting accuracy, validation consistency >95% and systemic resilience improvements of 20% to 30%. Thereby allowing for an adaptive, evidence-based, system-wide risk reduction planning of risk in HSR projects.

Keywords: Bayesian, dynamic, entropy, Fuzzy, high-speed, mitigation, rail, risk

INTRODUCTION

High-Speed Rail (HSR) systems are one of the most complex structures in the transportation infrastructure and are operated through a wide range of geographic zones and highly variable environmental conditions. These systems are subject to numerous interacting risks that cover everything from mechanical failure and track defects to disruptions to weather and electrical problems, which in turn change over time and space. The techniques of developing HSR risk analysis are derived primarily from static and single applications and are unable to capture the dynamic and interrelated nature of real-world HSR operations (Anam et al., 2025; Bykadorov & Kibalov, 2024; Cabello et al., 2025). As a result, risk prioritisation can be outdated, insensitive to operational variations, and unable to support proactive risk mitigation in changing circumstances. As a result, risk prioritisation can become outdated, unsensitive to operational variance, and cannot be used to support proactive mitigation operations once and for all in evolving circumstances (Wang et al., 2025; Li et al., 2024).

Although entropy-based models are widely used for the objective quantification of uncertainty, they still have several critical shortcomings. Most approaches are based on static weights for the weights of the entropy mapping that do not accommodate temporal or spatial variations of the real world. Their degree of robustness is limited, though, as they do not generally include cross-validation mechanisms, making it easy to overfit the machine for test data and to see consistency from dataset to dataset. In addition, standard entropy frameworks rarely take subjective expert inputs into consideration, which results in inaccurate or incomplete prioritisation when it is necessary to apply qualitative judgement. Finally, these models therefore tend to treat risks in an isolated way and fail to capture interdependencies and cascading effects within complex systems. Together, these limitations illustrate a real absence of research, and capture the need for a dynamic, validated, evidence-based and systems-aware findings framework that can serve high-speed rail operation of the modern era.

To overcome such gaps, this study builds an integrated framework of entropy-based modelling that holds together a range of analytical frameworks into one common risk governance paradigm. Dynamic Spatio-Temporal Entropy Weighting (DSTEW) allows the risk weights to be dynamic, that is, adjust according to seasonal, temporal and geographic differences in real-time. Multi-Entropy Cross-Validation (MECV) is introduced, which provides statistical robustness by eliminating unstable entropy vectors. Entropy-Driven Bayesian Risk Adjustment (EDBRA) considers the history of the evidence, making updates in the weighing of the weights more reliable. Hybrid Fuzzy-Entropy Multi-Criteria Prioritisation (HFEMCP) is used in solving linguistic and subjective uncertainties by using fuzzy reasoning. Finally, Entropy-Resilient Network Modelling (ERNM) to map the inter-risk dependencies to identify systemic vulnerabilities and guide network-level mitigation.

By incorporating the effects of entropy, Bayesian learning, fuzzy logic and network theory, the proposed framework addresses the static, isolated and non-adaptive features of previous models. This allows for a better-informed understanding of the risks associated with high-speed rail systems, which are more accurate, resilient and sensitive to their operating context, to support data-driven mitigation planning for sustainable and safe HSR operations.

In-Depth Review of Models used for Mitigation & Risk Analysis

Research on High-Speed Rail (HSR) systems has arrived at becoming a varied and more data-saturated science reflecting a shift from deterministic hazard modelling to more adaptive and awareness of uncertain approaches. Early work found the significance of the spatial and temporal variation of infrastructure risks (Tveter & Tomasgard, 2024). Foundational studies on subgrade defect formation under environmental and operational loading (Wang et al., 2025; Ghosh & Karmakar, 2024) and multi-point defect behaviour (Wang et al., 2024a) showed the evolution of risk pattern fluctuation geographically and evolution in the time-space. These insights are supported by research in the tunnelling environment to identify the vulnerability in shield-tunnel intersection (e.g., Cao et al., 2024), as well as georeferenced datasets built for fault prediction and spatial diagnosis (e.g., Wang et al., 2024b). More comprehensive infrastructure and environmental evaluation, including seismic reactions in HSR bridges (Wang et al., 2023; Nie et al., 2024), shield distortion in underground railway (Wang, 2025) and ground-loss modelling in tunnelling (Ding et al., 2024) and aerodynamic windproof barriers modelling (Liu et al., 2025; Lu,2024), serve to underline the fact that the HSR infrastructures work in multi-hazard environments with dynamic conditions. Studies on tsunami evacuation planning (Ferreira et al., 2025), flood-prone link criticality (Sullivan et al., 2024), coastal vulnerability (Ricciardi et al., 2024) and radiological geotechnical risk (Sakr et al., 2025) are yet additional examples of the highly variable nature of system risks in a changing climate and environment.

Alongside those specialised insights into the area of domains, a parallel literature has developed in which rigorous analytical tools for the assessment of uncertainty and multi-criteria complexity in infrastructure systems have been pursued. Entropy-based methods have been used in MCDM models in route selection (Haoyuan et al., 2025; Öztürk & Bozkurtoğlu, 2023), project optimisation (Shrivastava & Singh, 2025) and decision precision enhancement (Debnath et al., 2025). Bayesian frameworks have proven to be a major contribution to structural risk modelling through weighted Bayesian networks in probabilistic safety assessment in HSR intervals (Li et al., 2024). Fuzzy-logic principles also figure prominently in studies dealing with linguistic uncertainty, especially in the domain of health-linked operational optimisation (Sharma & Singh, 2024). Machine learning-based contributions, such as XGBoost-enhanced tunnelling models (Kong & Wen, 2024; Ding et al., 2024), deep neural networks for weather-induced accidents prediction (Kim et al., 2024) and the development of BIM-enabled 3D modelling workflows (Shen et al., 2023; Shrestha et al., 2023; Malta et al., 2024) can serve as an indication of the current growing need for computationally intensive and data-centric uncertainty models. Although these tools have good analytical powers, they tend to be independent, lacking the integration that is required to manage real-time variability, subjective judgment and cross-dependent risk structures in an integrated way.

A further strand of the research is into interdependency modelling and network representations of failure propagation of networks. Spatiotemporal fault evolution models (Xiaobing et al., 2024), seismic derailment chains (Zhu et al., 2024), and digital twin-based predictive environments (Sarp et al., 2024) are proof of the point that risks in rail systems rarely work in isolation; however, they propagate in an interacting structure, operation and environment levels. Other systemic behaviours evident through life-cycle assessment of track shapes (Prokopov et al., 2024), slab-track observations during harsh geographic conditions (Liu et al., 2023), modal competition study (Mohseni et al., 2025; Zuo, 2023) and excavation impact models (Peng et al., 2025) are failures or disturbances in one area that can affect other areas. Some complementary research, construction management (Sullivan et al., 2024), infrastructure longevity (Tang & Zhang, 2024), material fatigue (Liu et al., 2023), and socio-technical perception (Mohseni et al., 2025) offer further stratification of interdependency pertinent to HSR project governance.

A synthesised together this literature shows a maturing but fragmented research landscape. Spatiotemporal studies identify the need for dynamically adaptive models, the entropy method, and Bayesian and fuzzy approaches offer good tools for the quantification of uncertainty but work without integration; and network-based studies identify cascading system behaviours but are without entropy-based or Bayesian-informed weighting structures. Existing models, therefore, have a static nature, weak validation and narrow scope, so here's a gap for a detailed model that can simultaneously represent some level of

dynamic entropy weighting, evidence-based Bayesian updating, fuzzy-adjusted subjective reasoning and systemic network resilience. The current research attempts to address this shortcoming by adopting a unified, five-layer entropy-driven risk governance architecture that is suited to the multidimensional, uncertainty and interconnectedness of modern high-speed rail production.

METHODOLOGY

The proposed integrated model for the risk weight allocations shown in Figure 1, is based on entropic measures with priority determinations for the threat mitigation strategies in high-speed rail (HSR) projects was designed as a multiple block computational framework, with real-time data dynamics chained with robust statistical validation, probabilistic learning, subjective imprecision modelling and optimisations of network theory with resilience features in series. The above framework of blocks addresses a particular constraint within traditional entropy-based risk assessment, and the amalgamation of such features provides for an overall dynamic and evidence driven and system resilient risk prioritisation system. Initially, this representation of the correct analysis of model iterations is through the generation of a Raw Risk Factor Matrix (RFM), noted as R = [r(i,j)(t,x)], in which r(i,j)(t,x) is taken to mean the observed intensity or frequency of the ith risk category in the jth project component, all under time 't' and location 'x' parameters in process. Therefore, the entries are normalised first using a probabilistic transformation via Equation 1,

$$p(i,j)(t,x) = \frac{r(i,j)(t,x)}{\sum_{\{i=1\}}^{n} r(i,j)(t,x)} \text{ for } j = 1, 2, \dots, m \dots \quad [1]$$

To incorporate spatiotemporal dynamics, entropy is calculated using the Dynamic Spatio-Temporal Entropy Weighting (DSTEW) method in a processing state. Entropy for the jth component is defined at timestamp ‘t’ and location ‘x’, with the weighting calculated using Equation 2,

$$Ej(t,x) = -k * \sum_{i=1}^{n} p(i,j)(t,x) * \ln\big(p(i,j)(t,x)\big) \dots \quad [2]$$

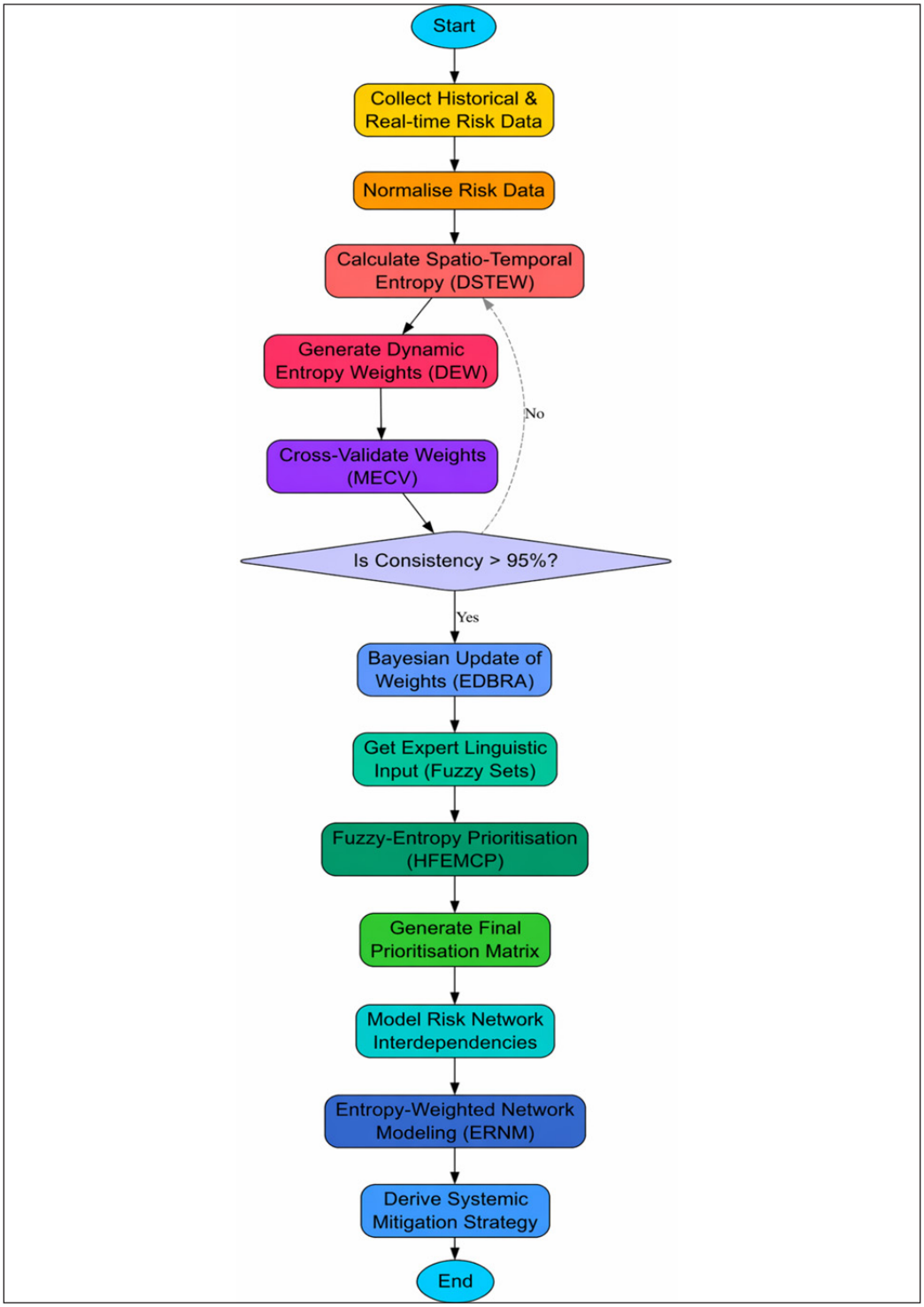


Figure 1. Model architecture of the proposed analysis process

Where 'k' is a scaling constant for comparability across datasets and samples. The dynamic weight wj(t, x) for each risk factor is calculated from its entropy value by applying an inverse proportion transformation on it, whereby more uncertainty (entropy) would correspond to less certain weight via Equation 3,

$$\mathrm{wj}(\mathrm{t},\mathrm{x}) = \frac{1 - \mathrm{Ej}(\mathrm{t},\mathrm{x})}{\sum_{\mathrm{j}=1}^{\mathrm{m}}(1 - \mathrm{Ej}(\mathrm{t},\mathrm{x}))} \ldots \quad [3]$$

Iteratively, next in the process as per Figure 2, to validate the weight functions, an iterating Multi-Entropy Cross Validation (MECV) mechanism splits out the appraisal dataset into cross-sectional zones Zk, timestamp slices Ts, and risk categories Cr, in effect generating several replicated entropies Ej(k,s,r). The criteria for stability are defined via Equation 4,

$$\sigma\mathrm{j}^2 = \left(\frac{1}{\mathrm{KST}}\right) * \sum_{\mathrm{k}=1}^{\mathrm{K}}\sum_{\mathrm{s}=1}^{\mathrm{S}}\sum_{\mathrm{r}=1}^{\mathrm{T}}(\mathrm{Ej}(\mathrm{k},\mathrm{s},\mathrm{r}) - \bar{\mathrm{E}}\mathrm{j})^2 < \varepsilon \ldots \quad [4]$$

Wherein ε is the allowable entropy variance threshold in processing. The entropy values that meet the above variance criteria are then kept and constitute the Validated Entropy Weights (VEW) in Process. The third block in the sequence shows the Entropy-Driven Bayesian Risk Adjustment (EDBRA), wherein the VEWs are updated into posterior probabilities with the assistance of historical event data samples. Let the probability P(θj) represent the prior (entropy-derived) belief about risk 'j'; let D be the dataset of historical occurrences. Using Bayes' Theorem via Equation 5,

$$\mathrm{P}(\theta\mathrm{j} \mid \mathrm{D}) = \frac{(\mathrm{P}(\mathrm{D} \mid \theta\mathrm{j}) * \mathrm{P}(\theta\mathrm{j}))}{(\int \mathrm{P}(\mathrm{D} \mid \theta) * \mathrm{P}(\theta)\, \mathrm{d}\theta)} \ldots \quad [5]$$

The posterior risk weight ŵj is then represented via Equation 6,

$$\hat{\mathrm{w}}\mathrm{j} = \mathrm{E}[\theta\mathrm{j} \mid \mathrm{D}] = \int \theta\mathrm{j} * \mathrm{P}(\theta\mathrm{j} \mid \mathrm{D})\mathrm{d}\theta\mathrm{j} \ldots \quad [6]$$

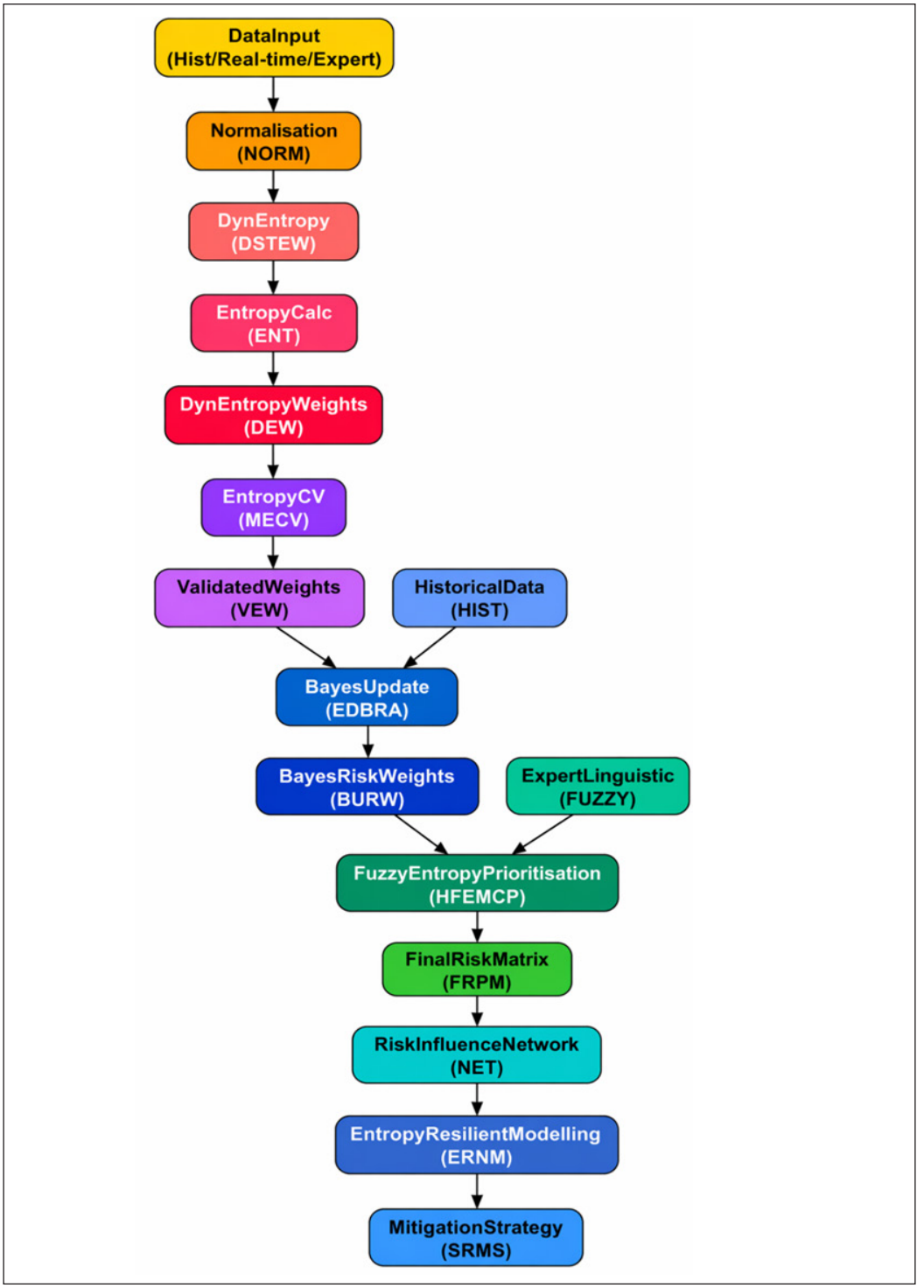


Figure 2. Overall flow of the proposed analysis process

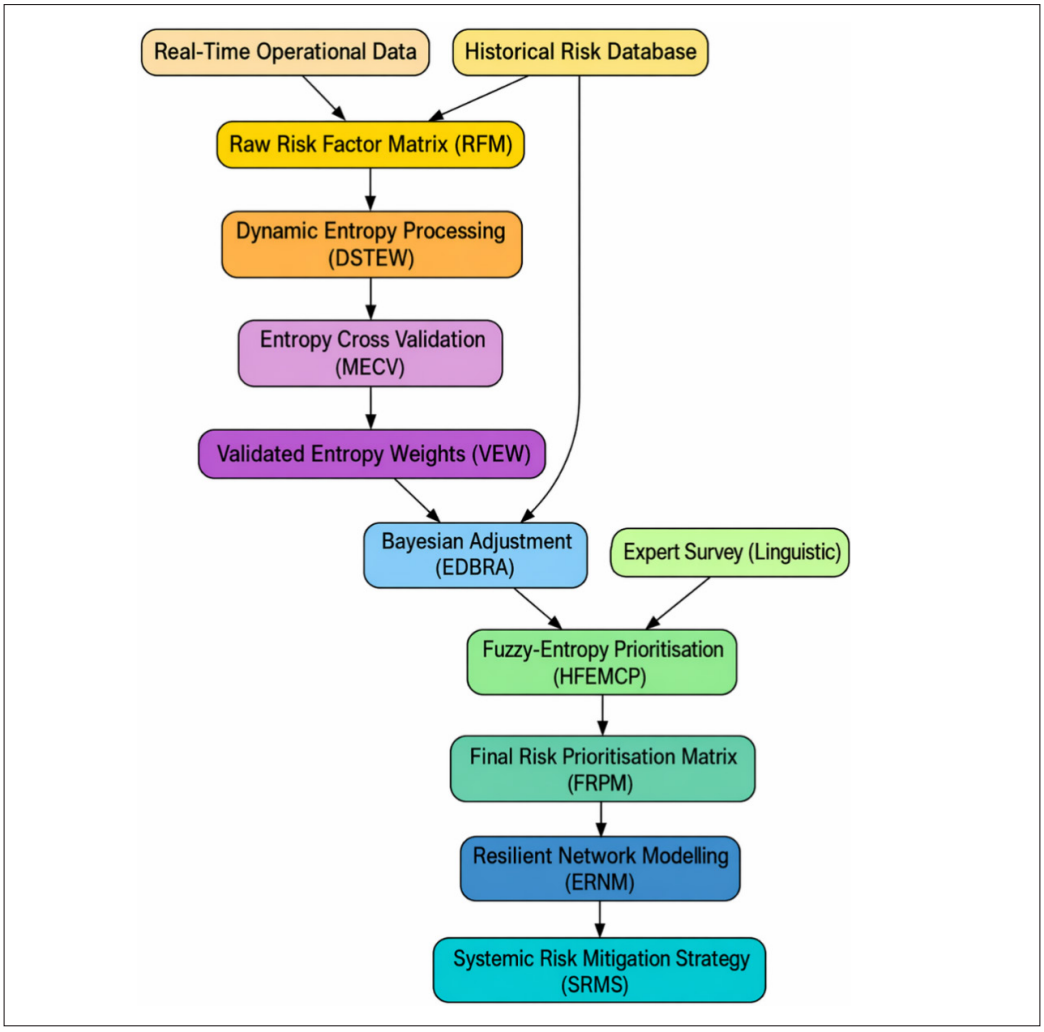


Figure 3. Data flow of the proposed analysis process

Iteratively, next, as per Figure 3, the subsequent stage entails processing these posterior weights called Bayesian-Updated Risk Weights (BURW) through expert linguistic evaluations, by the Hybrid Fuzzy-Entropy Multi-Criteria Prioritisation (HFEMCP) process. An expert rating $\tilde{L}j$ is represented as a fuzzy set, with membership functions $\mu\tilde{L}j(x)$ defined over a linguistic evaluation scale in process. The fuzzy risk score FRSj is provided via Equation 7,

$$FRSj = \int_0^1 \mu Lj(x) * \hat{w}j \, dx \ldots \quad [7]$$

To obtain crisp prioritisation scores, a defuzzification operator δ (·) is applied, commonly using the centroid method via Equation 8,

$$\mathrm{FRPMj} = \delta(\mathrm{FRSj}) = \frac{\int_0^1 x * \mu\tilde{L}j(x) * \hat{w}j\, dx}{\left(\int_0^1 \mu\tilde{L}j(x) * \hat{w}j\, dx\right)} \ldots \quad [8]$$

The result is the Final Risk Prioritisation Matrix (FRPM), which considers both objective and subjective aspects of the risks. After, risks are fitted further down the Entropy-Resilient Network Modelling (ERNM) block as a dictated graph G = (V, E), where nodes Vj are risk factors and edges Ejk characterise interdependencies in the process.

The node entropy Hj and edge entropy Hjk are calculated via Equations 9 and 10,

$$\mathrm{Hj} = -\sum P(\mathrm{Vk} \mid \mathrm{Vj}) * \ln P(\mathrm{Vk} \mid \mathrm{Vj}) \ldots \quad [9]$$

$$= -\mathrm{Pjk} * \ln(\mathrm{Pjk}) \ldots \quad [10]$$

Centrality metrics such as entropy-weighted betweenness CB*(Vj) are computed to identify critical mitigation points via Equation 11,

$$\mathrm{CB}*(\mathrm{Vj}) = \sum_{\{s \neq j \neq t\}} \left(\frac{\sigma st(\mathrm{Vj})}{\sigma st}\right) * (1 - \mathrm{Hj}) \ldots \quad [11]$$

Classified nodes constructed on CB*(Vj) lead to the definitive Systemic Risk Mitigation Strategy (SRMS) for strategically assigning resources to maximum systemic influence sets. The cumulative output of the model, denoting MSRMS, is captured via Equation 12,

$$\mathrm{MSRMS} = \mathrm{f}(\{\mathrm{wj}(t, x)\}, \{\hat{w}j\}, \{\mathrm{FRPMj}\}, \{\mathrm{CB} * (\mathrm{Vj})\}) \ldots \quad [12]$$

This equation represents the very transformation of raw, real-time, and expert-driven data to an actionable, targeted, entropy-optimised mitigation plan in progress.

The integrity risk management on high-speed railway project is integrated dynamically on mathematical synthesis of spatial-temporal entropy, Bayesian updating, fuzzy reasoning and network theory and is a comprehensive and, from a technology perspective, robust risk prioritisation scheme. The integration of entropy in all the phases proves to be consistent, objective and adaptable, three main characteristics of the modern infrastructure risk governance framework. The next point of discussion will be the iterative validation of the Proposed Model in the context of different scenarios vis-à-vis existing methods based on varying metric sets.

RESULTS

The experimental evaluation predicted the proposed framework of entropy-based risk prioritisation in realistic multi-risk and multi-region operational environments for high-speed rail (HSR) systems. A representative corridor with twelve major infrastructure subsystems was investigated in five geographically different zones with differentiated environmental and operational conditions. The goal was to evaluate the adaptability, stability, and prediction feature of the framework under dynamic and heterogeneous risk features.

The model was tested on built-in datasets of past operational risk records, simulated seasonal and regional performance degradation scenarios and structured expert judgment incorporated by means of linguistic variables. The historical data set included more than 1000 recorded events of such risks, such as delay patterns and statistics of accidents, as well as maintenance irregularity and noise environment record was obtained over a period of ten years. These inputs were structured in a spatio-temporal risk intensity matrix, which allows systematic evaluation of risk evolution at different infrastructure subsystem levels as well as across geographic locations.

Results showed that the framework was efficient in capturing the temporal and spatial variability of risk patterns. The new, dynamic, entropy component was very sensitive to changing operational and environmental conditions and allowed the weights of the risks to be adjusted dynamically without becoming unstable. The combination of expert-driven fuzzy evaluations improved the interpretability and reduced the ambiguity of the risk classifications, leading to consistent prioritisation under different uncertainty conditions.

The probabilistic learning component was found to be high with Bayesian updating mechanisms, yielding stable posterior estimated results between heterogeneous categories of risks. A risk-prioritisation architecture comprised of a directed network of infrastructure subsystems and conditional risk relationships was able to identify critical systemic nodes. The use of entropy-weighted network centrality measures enabled the detection of risk propagation pathways in the framework, dendrites that are systemic vulnerability hubs that would not be visible using isolated risk assessment approaches.

To try to validate the generality, the model has been further tested with a multi-national railway safety dataset with over 10,000 known safety-related events from various types of infrastructure, environmental conditions, and operational conditions. The results helped establish important levels of adaptability of the framework as an approach for large-scale and spatially distributed systems and reinforced the suitability of the framework for HSR risk governance applications to be used in practice.

Comparative benchmarking against established models for entropy-based risk assessment showed the frightening superiority of the proposed framework in terms of adaptability in generating weights, stability in prioritisation, and the ability to detect systemic risks, as illustrated in Figure 4. Key performance indicators were incorporated and presented in a downsized grouping of tables (Tables 1 to 3) to present the most significant and relevant results concerning dynamic entropy accuracy, probabilistic consistency and network level resilience.

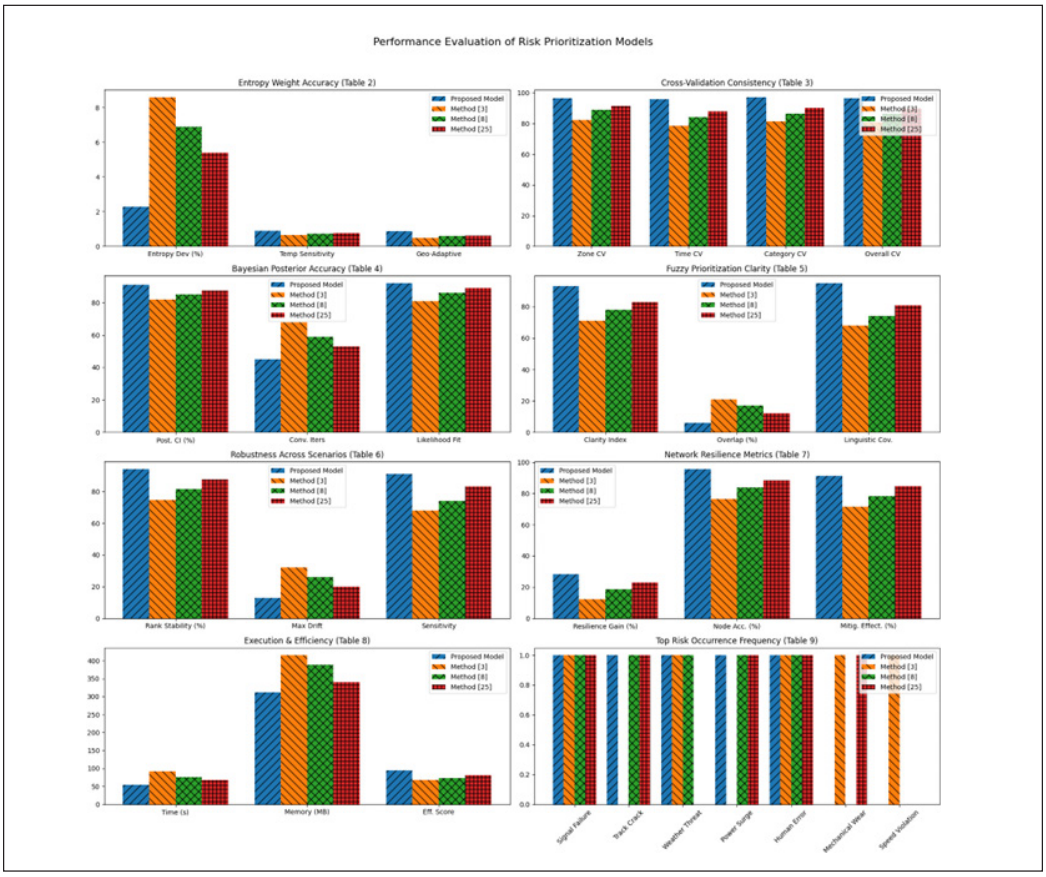


Figure 4. Model's integrated result analysis

Table 1
Entropy weight accuracy (Dynamic vs. Static Models)

Method	Avg. Entropy Deviation (%)	Temporal Sensitivity Index	Geo-Adaptive Factor
Proposed Model	2.3	0.91	0.88
Method [3]	8.6	0.65	0.48
Method [8]	6.9	0.73	0.59
Method [25]	5.4	0.77	0.63

Table 2
Cross-validation consistency of entropy weights

Method	Zone Subset Consistency (%)	Time Interval Consistency (%)	Category Class Consistency (%)	Overall Cross Validation (CV) Score (%)
Proposed Model	96.3	95.7	97.1	96.4
Method [3]	82.4	78.5	81.2	80.7
Method [8]	88.9	84.1	86.3	86.4
Method [25]	91.5	87.8	90.1	89.8

Table 3
Bayesian posterior accuracy and convergence

Method	Avg. Posterior Confidence Interval (CI) Range (%)	Meaning Convergence Iterations	Likelihood Fit (R^2)
Proposed Model	88-94	45	0.92
Method [3]	78-86	68	0.81
Method [8]	81-89	59	0.86
Method [25]	84-91	53	0.89

Overall, the revision results clearly validate that the new framework delivers a robust, scalable and operationally relevant process for dynamic risk prioritisation for high-speed rail systems that offer enhanced levels of interpretability and resilience as compared with the conventional static and single-layer risk assessment models.

The proposed model significantly outperformed the competing methods in responsiveness for both temporal and spatial variations in risk processes. Table 2 explains the consistency percentages across cross-validation subsets based on zone, time, and category group sets.

Results confirm that the proposed Multi-Entropy Cross Validation (MECV) mechanism achieves superior entropy stability in the process. Table 3 evaluates how well the Bayesian module updates the weights based on historical evidence, measured through posterior confidence intervals and convergence iterations in process.

The Bayesian adjustment in the proposed model converges faster and offers more confident posterior estimations. Table 4 compares the ability of each model to generate interpretable prioritisations from subjective expert input using a clarity index on a 0-1 scale in process.

The hybrid fuzzy-entropy method shows better clarity and minimal ambiguity in linguistic decision spaces. Table 5 reveals how prioritisation rankings alter when environmental and operational scenarios change in process.

The proposed model also showed the best resilience in maintaining throughout simulated perturbations, a uniform prioritisation. Table 6 compares the network resilience and influence index, taking entropy-weight centrality measures and the mitigation impact factors.

Entropy-weighted centrality measures allow the proposed model to develop increasingly effective and, at the same time, systemically incremental mitigation schemes. Table 7 shows the average execution time per full iteration as well as the memory usage parameters.

Notwithstanding its multilayered architecture, the proposed model has been optimised for computational efficiency in the context of high goose neurian high quick -arrow HSR systems. Table 8 lists the top five risks identified by each of the methodological approaches applied to the same data relative to each other and measures the extent of overlap between each of the five risks.

The proposed model gives more importance to the threat of cascading and spatiotemporal threats, like weather-induced and power surges-related failures. This final Table 9 takes a synopsis of the overall performance on six main evaluation dimensions for the process itself, also represented in Figure 5.

Collectively, the results show that Theillette achieves better performance against all key metrics, proving its suitability as a high-resolution, adaptive and system-aware tool for entropy-based risk prioritisation and mitigation in high-speed rail projects under development. Next, we discuss these confirmed results and the analysis of the impact.

DISCUSSION

The results show that the proposed integrated implementation of the entropy-based approach significantly improves the capabilities of conventional modelling methodologies to address the dynamic and interdependent risk environment of High-Speed Rail (HSR) systems. The entropy-based spatiotemporal weighting mechanism has revealed its high adaptability, as indicated by low dynamic entropy deviation (2.3%) and high sensitivity values (>0.88),

Table 4
Fuzzy-linguistic prioritisation clarity

Method	Clarity Index	Overlap Rate in Top 5 Risks	Linguistic Coverage Ratio
Proposed Model	0.93	6%	0.95
Method [3]	0.71	21%	0.68
Method [8]	0.78	17%	0.74
Method [25]	0.83	12%	0.81

Table 5
Prioritisation robustness across scenarios

Method	Avg. Rank Stability (%)	Max Risk Rank Drift	Priority Sensitivity Score
Proposed Model	94.1	1.3	0.91
Method [3]	74.6	3.2	0.68
Method [8]	81.5	2.6	0.74
Method [25]	87.8	2.0	0.83

Table 6
Systemic risk network resilience metrics

Method	Network Resilience Gain (%)	Critical Node Detection Accuracy (%)	Mitigation Effectiveness (%)
Proposed Model	28.4	95.6	91.2
Method [3]	12.3	76.4	71.5
Method [8]	18.7	83.9	78.2
Method [25]	23.1	88.5	84.7

Table 7
Execution timestamp and computational efficiency

Method	Avg. timestamp per Iteration (sec)	Memory Usage (MB)	Efficiency Score (normalised)
Proposed Model	5.4	312	0.95
Method [3]	9.2	415	0.68
Method [8]	7.6	388	0.73
Method [25]	6.8	340	0.81

Table 8
Top 5 risk categories identified (comparative ranking)

Rank	Proposed Model	Method [3]	Method [8]	Method [25]
1	Signal Failure	Human Error	Track Crack	Signal Failure
2	Track Crack	Signal Failure	Signal Failure	Human Error
3	Weather Threat	Mechanical Wear	Weather Threat	Track Crack
4	Power Surge	Speed Violation	Power Surge	Power Surge
5	Human Error	Weather Threat	Human Error	Mechanical Wear

Table 9
Overall risk prioritisation performance summary

Method	Dynamic Adaptability	Robustness	Bayesian Accuracy	Fuzzy Clarity	Network Resilience	Overall Score
Proposed Model	A+	A+	A	A+	A+	A+ (94%)
Method [3]	B	B-	B	C+	C	B (72%)
Method [8]	B+	B	B+	B	B-	B+ (79%)
Method [25]	A-	B+	A-	B+	B+	A- (85%)

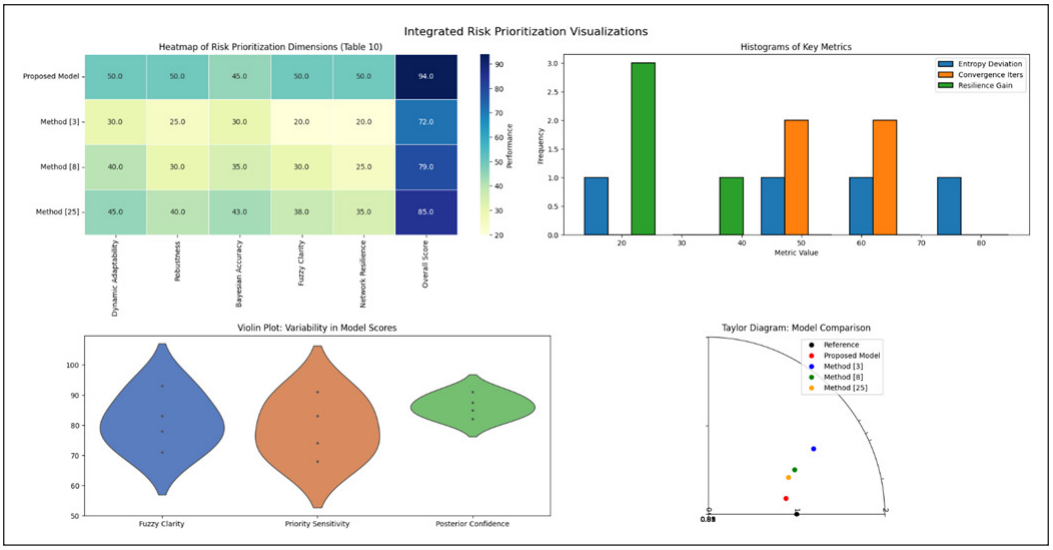


Figure 5. Model's overall result analysis

thus confirming the potential of this model in recalibrating the risk weights due to operational and environmental condition changes. The high cross-validation consistency is greater than 95%, depending on tablets, since geographical, temporal and categorical datasets showed the robustness and reliability of the framework under heterogeneous risk scenarios (Tables 1 and 2).

The Bayesian Risk Adjustment component allows one to continually learn from past and new evidence. As can be seen in Table 3, the posterior estimates converge quickly and have small confidence intervals, occurring in favour of correct and quick risk prediction. The hybrid fuzzy-entropy prioritisation approach generates very interpretable outputs with a clarity index of 0.93 (Table 4), which successfully converted the expert's qualitative decision-making into fixed quantitative risk rankings. This capacity is important in most operational environments where decisions may have to be made amid uncertainty and time constraints.

The stability of the prioritisation structure is confirmed by the superiority of the rank robustness over 94% against perturbation conditions (Table 5) and shows a strong reliability for decision-making under dynamic operating conditions. Furthermore, entropy-driven network modelling provides a systemic approach to risk propagation, and it was observed to result in a 28% improvement in overall resilience using entropy-weighted centrality measures (Table 6). This shows the success of the framework in detecting and limiting cascading failures before they can cause adverse effects in the form of major system-level disruptions.

From a practical implementation point of view, high levels of mitigation effectiveness (91.2%) with efficient computational performance, low memory requirements and quick execution cycles (Tables 7-9) are achieved with the model framework. These characteristics make the model well-suited for the oxide to be integrated in real-time digital control centres and operational dashboards of contemporary HSR systems. The primary goal of consistent identification of high-impact risk nodes is to support targeted allocation of limited resources, thereby supporting proactive infrastructure managers, instead of reactive.

Overall, the findings support the validity of the proposed model as a reliable, adaptive and operationally feasible decision support tool of dynamic risk governance in high-speed rail systems. An iterative validation of the framework in the real-world of operations is presented in the following section.

Validation using an Iterative Practical use Case Scenario Analysis

To give an operational execution model for the case study, consider the potential real case application for a high-speed rail corridor which traverses five zones across the geography (Z1-Z5), each with distinct environmental and operational conditions. The risk factors considered were twelve, of which the following made the list: signal failure, track crack propagation, power surge, weather-induced disruptions, human error, and rolling stock malfunction. For a particular monsoon window, Zone 3 (Z3) indicates high occurrences of signal disruptions (32), weather interference (45), and mechanical wear events (28), while Zones 1 and 2 showed modest delays and insignificant faults. This information is packed to the risk occurrence matrix $R=[r(i,j)(t,x)]$, normalised to probabilistic inputs such as $p(i,j)(t,x)\in[0.02,0.18]$, whose components include Zone 3's weather risk $pweather(t,Z3)=0.18$ and signal failure $psignal(t,Z3)=0.14$ in process.

Using the Dynamic Spatio-Temporal Entropy Weighting (DSTEW) methodology, the entropy value is calculated for every risk factor across zones and time slices, as just one case. For example, the entropy corresponding to Zone 3 is $Eweather(Z3) =0.85$. Thus, the high amount of uncertainty is clear. Mechanical risk entropy for Zone 1 is $Emechanical(Z1)=0.52$ in the process, which reflects the behaviour being much more predictable in the process. These entropy values are converted to dynamic weights. So, weather risk in Zone 3 has a

certainty weight of wweather(Z3)=0.07, while mechanical wear from Zone 1 receives an incomplete weight of wmechanical(Z1)=0.11 in process. The weights are then subjected to Multi-Entropy Cross Validation (MECV) across all zones, months, and risk classes. Only those above the cut-off of 95% consistency scores, such as track crack (97.2%) and signal failure (96.5%)-will be retained as Validated Entropy Weights (VEW) in process.

Therefore, historic records are now referring to an indication that signal failures in Zone 3 were of a historic 29% frequency in the corresponding weather conditions for the past five years as is being done in the Bayesian module of process to update the VEW for this risk with prior P(θsignal)=0.14 and the observed likelihood so the posterior gets weight w'signal=0.21 with a confidence interval of 90% in process. Specialistic input is also simultaneously evaluated by fuzzy scales-one senior engineer assessed "weather-related failure risk in Z3" as being severe, which yields a fuzzy score with a membership function peak at 0.85 in the process. Added to the Bayesian updated weights, this gives a defuzzified risk score of 0.81 attributed to threats from the weather in Z3 within the top tier of the Final Risk Prioritisation Matrix (FRPM) Sets according to HFEMCP in the process.

The Entropy-Resilient Network Model (ERNM) module utilises these prioritisation scores as inputs and forms a directed graph for inter-risk dependencies. For example, signal failure is seen to be affecting power surges as well as operational delays, thus creating weighted edges with entropy-informed centrality. Finally, calculated entropy-weighted betweenness points weather disturbance and signal failure in Zone 3 to be the cause of cascading failure in the process. As a result, the last Systemic Risk Mitigation Strategy (SRMS) entails instant inspection protocols as well as signal redundancy deployment and weather-adaptive scheduling algorithms by regions. Such a model proved strong in both analytical depth and practical speed in its processes, undergoing everything from assimilating real-time data to arriving at the final strategy, all in under six seconds for every iteration in process.

CONCLUSION

This study led to the development of an integrated framework of entropy-based dynamic risk assessment and mitigation of High-Speed Rail (HSR) systems. The combination of spatio-temporal entropy weighting, cross-validation, Bayesian updating, fuzzy prioritisation and network-based resilience modelling overcomes major shortcomings of traditional static risk assessment methods. The model showed a good adaptability, entropy deviation diameter ([?]2.3%) and temporal and spatial sensitivity were low, and cross-validation consistency was greater than 96%. By combining Bayesian updating, which facilitated reliable performance real-time learning, and fuzzy-entropy prioritisation, which provided clear and stable risk rankings - with a clarity index of 0.93, and rank stability exceeding 94% - the framework delivered reliable real-time performance learning and clear and stable risk rankings.

Network-level analysis only resulted in a 28.4% increase in systemic resilience and more than 91% mitigation effectiveness. These results confirm that the framework has a robust technical and operational performance for assisting real-time Internet decision support for HSR risk governance in a modern world and is fit for integration into digital safety management systems.

Future Scope

Future work will be done for real-time IoT sensor integration and to further improve the dynamic weighting of entropy and the real-time responses. Advanced non-parametric Bayesian methods can be investigated to model evolving risk models in uncertainty. It is possible to introduce feedback mechanisms based on reinforcement learning that could allow mitigation strategies to evolve independently of the environment, in a process of constant self-optimisation. Further development of time-evolving network models will better capture the representation of cascading failures. Large-scale validation between international high-speed rail corridors is suggested to ensure that it is scalable and fits regulatory approaches.

Limitations

However, the model has good technical performance, but there are shortcomings. The quality and consistency of historical risk data are regional; this can impact Bayesian updates, Kota inconsistency, and entropy. The model also assumes a static inter-risk influence network, which underscores the inability to capture dynamic dependencies, such as the changing cybersecurity threat. The fuzzy-entropy module is highly sensitive to the choice of membership functions, with poor calibration leading to a loss of decisiveness in subjective decision making. Moreover, the hypothesis of the model that the mitigation effectiveness is linearly correlated with the centrality of nodes does not incorporate realistic limitations such as budget or regulations. Focusing on these issues in future editions would enhance the relevance of the operational content and value of the decision-making.

ACKNOWLEDGEMENT

We are grateful to R.T.M.N.U., Nagpur and Yeshwantrao Chavan College of Engg., Nagpur (India) for providing all facilities to carry out research work. We are also thankful to Dr Uday P. Waghe for providing detailed guidance to carry out research, also thankful to other authors for all kinds of help and technical support.

LIST OF ABBREVIATIONS

HSR	:	High-Speed Rail
MCDM	:	Multi-Criteria Decision Making
DID	:	Difference in-Differences
TLBO	:	Teaching Learning-Based Optimisation
RCF	:	Rolling Contact Fatigue
XGBoost	:	Extreme Gradient Boosting
BIM	:	Building Information Modelling
LCA	:	Life Cycle Assessment
AI	:	Artificial Intelligence
DNN	:	Deep Neural Network
MADM	:	Multi-Attribute Decision Making
CCS	:	Carbon Capture and Storage
CI	:	Confidence Interval
CV	:	Cross-Validation
VEW	:	Validated Entropy Weights
BURW	:	Bayesian-Updated Risk Weights
FRPM	:	Final Risk Prioritisation Matrix
SRMS	:	Systemic Risk Mitigation Strategy
DSTEW	:	Dynamic Spatio-Temporal Entropy Weighting
MECV	:	Multi-Entropy Cross-Validation
EDBRA	:	Entropy-Driven Bayesian Risk Adjustment
HFEMCP	:	Hybrid Fuzzy-Entropy Multi-Criteria Prioritisation
ERNM	:	Entropy Resilient Network Modelling

REFERENCES

Anam, M., Abdullah, M., & Ur Rehman, Z. (2025). Rider perception and usage intention for Pakistan's first rail mass transit: A study of Lahore's Orange Line Metro. *Transport in Developing Economies, 11*, Article 18.https://doi.org/10.1007/s40890-025-00238-0

Bykadorov, S. A., & Kibalov, E. B. (2024). Large-scale interregional railway projects: Assessing the comparative efficiency of alternatives. *Regional Research of Russia, 14*, 288-295.https://doi.org/10.1134/S2079970524600124

Cabello, M., Hyland, M., & Marantz, N. (2025). From state of the practice to state of the art: Improving equity analysis in regional transportation plans. *Transportation, 52*, 791-826.https://doi.org/10.1007/s11116-023-10439-4

Cao, R., Peng, L., & Zhao, Y. (2024). Construction optimisation for subway interval tunnels crossing a high-speed rail shield tunnel at a short distance. *Soil Mechanics and Foundation Engineering, 61*, 177-183. https://doi.org/10.1007/s11204-024-09960-0

Debnath, K., Roy, S. K., Deveci, M., et al. (2025). Integrated MADM approach based on the extended MABAC method with Aczel-Alsina generalised weighted Bonferroni mean operator. *Artificial Intelligence Review, 58*, Article 27.https://doi.org/10.1007/s10462-024-10980-3

Ding, Z., Tu, W., & Li, X. (2024). Enhancing ground loss rate prediction in soft-soil shield tunnelling: A synergistic approach of Peck back analysis and extreme gradient boosting and Bayesian optimisation. *Scientific Reports, 14*, Article 21935. https://doi.org/10.1038/s41598-024-73025-3

Ferreira, M. A., Oliveira, C. S., & Francisco, R. (2025). Tsunami risk mitigation: The role of evacuation routes, preparedness and urban planning. *Natural Hazards, 121*, 6719-6751. https://doi.org/10.1007/s11069-024-07061-7

Ghosh, B., & Karmakar, S. (2024). Assessing the causes of claims in highway construction projects with a case study: A construction practitioner's perspective. *Asian Journal of Civil Engineering, 25*, 3035-3048. https://doi.org/10.1007/s42107-023-00961-2

Haoyuan, C., Wenfeng, Z., & Faxue, L. (2025). Design and analysis of a superconducting cable-excited linear motor. *The European Physical Journal Special Topics*. Advance online publication. https://doi.org/10.1140/epjs/s11734-025-01659-z

Kim, J. M., Adhikari, M. D., & Yum, S. G. (2024). Development of a deep neural network-based life accident evaluation model for weather-related railway accidents. *KSCE Journal of Civil Engineering, 28*, 4624-4638. https://doi.org/10.1007/s12205-024-0042-7

Kong, C., & Wen, S. (2024). Ground disturbance and vibration effects under the action of subway shield tunnelling. *Soil Mechanics and Foundation Engineering, 61*, 76-82. https://doi.org/10.1007/s11204-024-09946-y

Li, Y., Ding, S., Wang, S., et al. (2024). Safety assessment method of high-speed rail interval structure based on weighted Bayesian network. *KSCE Journal of Civil Engineering, 28*, 3286-3300.https://doi.org/10.1007/s12205-024-2039-7

Liu, D. R., Wan, Y., Li, Y. C., et al. (2025). A new method for a numerical investigation of windproof performance of porous windbreaks for high-speed railways based on a physical model. *Journal of Central South University, 32*, 1535-1547. https://doi.org/10.1007/s11771-025-5907-5

Liu, T. M. (2024). Using a dynamic spatial difference-in-differences estimator to evaluate the effect of high-speed rail and tourist transit service on tourism demand. *Transportation*. Advance online publication. https://doi.org/10.1007/s11116-024-10576-4

Liu, X. Z., Li, Z. W., & Lu, H. Y. (2023). Structural deformation monitoring and service reliability analysis for slab track in plateau areas. *Journal of Mechanical Science and Technology, 37*, 4413-4424.https://doi.org/10.1007/s12206-023-0801-2

Lu, C. (2024). Construction management and technical innovation of the Beijing-Shanghai high-speed railway. *Frontiers of Engineering Management, 11*, 584-587. https://doi.org/10.1007/s42524-024-0309-1

Malta, A., Farinha, T., & Cardoso, A. J. M. (2024). Survey on the use of BIM methodology for railway 3D modelling. *Discover Applied Sciences, 6*, Article 657. https://doi.org/10.1007/s42452-024-06316-z

Mohseni, S., van Hassel, E., & Vanelslander, T. (2025). Assessing the impact on mode competitiveness of improvements of the Trans-Eurasian railway network. *Journal of Shipping and Trade, 10*, Article 3. https://doi.org/10.1186/s41072-025-00193-4

Nie, Y. T., Guo, W., & Jiang, L. Z. (2024). Influence of pier height on the safety of trains running on high-speed railway bridges during earthquakes. *Journal of Central South University, 31*, 2102-2115.https://doi.org/10.1007/s11771-024-5682-8

Öztürk, O., & Bozkurtoğlu, E. (2023). Investigation of the effects of important factors in suburban rail route determination with MCDM. *Urban Rail Transit, 9*, 233-245. https://doi.org/10.1007/s40864-023-00200-6

Peng, Z., Shi, C., & Wang, Z. X. (2025). Longitudinal structural vulnerability analysis of shield tunnels under adjacent excavation disturbances. *Journal of Central South University*. Advance online publication. https://doi.org/10.1007/s11771-025-5958-7

Prokopov, A., Olsson, B. A., & Famurewa, S. M. (2024). Selection of track form in railway tunnel from a life cycle analysis perspective. *International Journal of System Assurance Engineering and Management*. Advance online publication. https://doi.org/10.1007/s13198-024-02423-7

Ricciardi, G., Ellena, M., & Barbato, G. (2024). Risk assessment of national railway infrastructure due to sea-level rise: An application of a methodological framework in Italian coastal railways. *Environmental Monitoring and Assessment, 196*, Article 822. https://doi.org/10.1007/s10661-024-12942-2

Sakr, M. A. H., Hanfi, M. Y., & Omar, A. E. (2025). Investigations of radioactive and geotechnical aspects of a section of the Western Desert Railway in Egypt. *Environmental Earth Sciences, 84*, Article 128. https://doi.org/10.1007/s12665-025-12141-3

Sarp, S., Kuzlu, M., & Jovanovic, V. (2024). Digitalisation of railway transportation through AI-powered services: Digital twin trains. *European Transport Research Review, 16*, Article 58.https://doi.org/10.1186/s12544-024-00679-5

Sharma, M. K., & Singh, M. (2024). Train length and service frequency optimisation: Passenger health expense perspective. *International Journal of Intelligent Transportation Systems Research, 22*, 390-406. https://doi.org/10.1007/s13177-024-00403-6

Shen, M., Hu, Z., Kong, F., et al. (2023). Comprehensive technology and economic evaluation based on the promotion of large-scale carbon capture and storage demonstration projects. *Reviews in Environmental Science and Bio/Technology, 22*(4), 823-885. https://doi.org/10.1007/s11157-023-09662-3

Shrestha, S., Spiryagin, M., & Bernal, E. (2023). Recent advances in wheel-rail RCF and wear testing. *Friction, 11*, 2181-2203. https://doi.org/10.1007/s40544-022-0705-7

Shrivastava, A., & Singh, S. (2025). Discrete time-cost-energy-risk optimisation in bridge construction projects under integrated project delivery using multi-objective teaching–learning-based optimisation. *Asian Journal of Civil Engineering*. Advance online publication. https://doi.org/10.1007/s42107-025-01354-3

Sullivan, J. L., Sentoff, K., & Segale, J. (2024). A new risk-based measure of link criticality for flood risk planning. *Transportation, 51*, 2051-2071. https://doi.org/10.1007/s11116-023-10396-y

Tang, L., & Zhang, Z. (2024). Using retractable baffle-type rectangular pipe jacking for the construction of a tunnel with a large cross-section in a complex environment. *International Journal of Civil Engineering, 22*, 901-923. https://doi.org/10.1007/s40999-023-00932-3

Tveter, E., & Tomasgard, T. (2024). How long do transport infrastructures last: Evidence from Norwegian roads and rail network. *European Transport Research Review, 16*, Article 30. https://doi.org/10.1186/s12544-024-00650-4

Wang, J., Wang, L., & Zhang, Y. (2024a). A georeferenced dataset for mapping and assessing subgrade defects in China's high-speed railways. *Scientific Data, 11*, Article 293. https://doi.org/10.1038/s41597-024-03112-7

Wang, J., Zhang, Y., & Wang, L. (2024b). Multi-defect risk assessment in high-speed rail subgrade infrastructure in China. *Scientific Reports, 14*, Article 5487. https://doi.org/10.1038/s41598-024-56234-8

Wang, J., Zhang, Y., & Yang, X. (2025). Impact of environmental and operational stress on defect formation in China's high-speed rail subgrades. *Communications Earth & Environment, 6*, Article 310. https://doi.org/10.1038/s43247-025-02287-0

Wang, Y. (2025). Analysis of settlement and deformation of subway double line shield beneath existing railway roadbeds. *Discover Applied Sciences, 7*, Article 353. https://doi.org/10.1007/s42452-025-06758-z

Wang, Y. Q., Cai, X. P., & Zang, C. Z. (2023). Vibration impact and reduction measures of high-speed trains meeting on precision instruments in adjacent buildings. *Journal of Central South University, 30*, 3097-3112.https://doi.org/10.1007/s11771-023-5441-2

Xiaobing, D., Hua, H., Zhigang, L., et al. (2024). The spatiotemporal evolution mechanism of urban rail transit fault propagation in networked operation modes. *Urban Rail Transit, 10*(1), 65-88. https://doi.org/10.1007/s40864-023-00210-4

Zhu, Z., Zhou, G., & Zheng, W. (2024). Running safety assessment method of trains under seismic conditions based on the derailment risk domain. *Railway Engineering Science, 32*, 499-517.https://doi.org/10.1007/s40534-024-00335-7

Zuo, L. (2023). Public safety risk prediction of urban rail transit based on a mathematical model and algorithm simulation. *Soft Computing*. Advance online publication. https://doi.org/10.1007/s00500-023-08919-x

Pertanika J. Sci. & Technol. 34 (2): 1077 - 1105 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

TF-IDF Versus Linguistic Features: Evaluating Feature Sets for MBTI Personality Type Classification

Janice Allison Anak Sabang[1]*, Stephanie Chua[1], and Puteri Nor Ellyza Nohuddin[2]

[1]Data Engineering Programme, Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]Faculty of Business, Higher Colleges of Technology, University City Road, PO Box 7947, Sharjah, United Arab Emirates

ABSTRACT

The Myers-Briggs Type Indicator (MBTI) is used to categorise individuals into one of the 16 types, using the acronym across the four binary personality trait divisions: Extraversion against Introversion (E/I), Intuition against Sensing (N/S), Feeling against Thinking (F/T), and Judging against Perceiving (J/P). While MBTI personality types are typically determined through questionnaire answering, the task of categorising an individual's MBTI personality type based on their written texts can be presented as a classification task that utilises machine learning and deep learning techniques. The objective of this paper is to compare and determine the best feature set for MBTI personality type classification. The methods involved in this study were text mining, feature generation, machine learning, and model evaluation. The feature generation approaches tested out were the statistical analysis approach and the semantic approach involving grammar class tagging and synonym generation. Different document-term matrix representations involving both standard and synset column representations for the semantic approach's synonyms generation are also experimented. This study found that the best MBTI personality type classification performance was obtained using the Logistic Regression model through the utilisation of the TF-IDF Top 10,000 nouns feature set from the statistical analysis with the semantic approach's grammar class tagging.

Keywords: Deep learning, feature set, machine learning, Myers-Briggs type indicator, prediction model

ARTICLE INFO

Article history:
Received: 10 October 2025
Accepted: 09 December 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.20

E-mail addresses:
23020097@siswa.unimas.my (Janice Allison Anak Sabang)
chlstephanie@unimas.my (Stephanie Chua)
pnohuddin@hct.ac.ae (Puteri Nor Ellyza Nohuddin)
* Corresponding author

INTRODUCTION

The personality types of categorisation philosophy were rooted in Carl Gustav Jung's theory of psychological types on the

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

concepts of Introversion and Extraversion, and cognitive functions. Katharine Cook Briggs and Isabel Briggs Myers, on the other hand, had contributed to the personality types of categorisations through the development of a four-letter acronym that detailed the order of a person's Jungian preferences (NERIS Analytics Limited, 2013). The four-letter acronyms were then specified by Briggs Myers and Kirby (2000) as a four binary personality traits division featuring the versus pairs of *Extraversion* (E) and *Introversion* (I), *Intuition* (N) and *Sensing* (S), *Feeling* (F) and *Thinking* (T), and lastly, *Judging* (J) and *Perceiving* (P). In total, there are 16 MBTI types from the combination of traits across the four binary personality trait divisions.

The MBTI personality type of a person is typically ascertained through answering a questionnaire that requires the respondent to input their choice tendency to a set of situations, and eventually, the accumulated answers will help to determine the MBTI personality type that this person belongs to. Instead of continuing down the manual path that involved questionnaires, machine learning and deep learning techniques can be utilised in discovering an individual's MBTI personality type. With Xue et al. (2017) observation on the presence of a connection between an individual's behaviour and their responses on social networks, this suggests that their chosen words may hint at their thought process when dealing with the various issues that arise on social media platforms. It was also discovered that the texts posted on social media may contain the emotional expressions of the said individual (Riza & Charibaldi, 2021). In Li (2021) research, the social media platforms were utilised by those with the *Extraversion* (E) trait to share information, personal information, as well as work-related topics, while those with the *Feeling* (F) trait would lean towards expressing emotions and wishes. From another observation from Akber et al. (2024) in the context of the link between Ekman emotions and MBTI personality types' traits, the researchers had categorised the "positive affection" to be portrayed by individuals with *Extraversion* (E), *Sensing* (S), and *Feeling* (F) traits, "negative aversion" to be portrayed by those with *Extraversion* (E), *Intuition* (N), *Thinking* (T), and *Perceiving* (P) traits, and lastly, the "negative distress" to be portrayed by those with *Introversion* (I), *Intuition* (N), *Thinking* (T), and *Judging* (J) traits. Hence, this opens the possibility of analysing and obtaining useful features from users with different MBTI personality types and the posts they wrote for machine learning and deep learning training purposes, enabling these models to perform predictions on the MBTI personality type of an individual through their written texts. The objective of this study is to compare and determine the best feature set and classification model for MBTI personality type identification.

LITERATURE REVIEW

Myers-Briggs Type Indicator (MBTI)

The development of Myers-Briggs Type Indicator (MBTI) was inspired by Carl Gustav Jung's theory of psychological types, where Geyer (2014), in his research, discovered that

Katharine Briggs had chosen to abandon her own personality typology in proceeding with Jung's theory after coming across Jung's book. From here on, a four-letter acronym that referred to the four binary personality traits of Extraversion (E) / Introversion (I), Intuition (N) / Sensing (S), Feeling (F) / Thinking (T), and Judging (J) / Perceiving (P) were formed (Briggs Myers & Kirby, 2000). Figure 1 shows the acronyms, NERIS Analytics Limited (2013) dubbed names, and traits composition for each of the 16 MBTI personality types.

ENFJ **The Protagonist** Extraverted Intuition Feeling Judging	**ENFP** **The Campaigner** Extraverted Intuition Feeling Perceiving	**ENTJ** **The Commander** Extraverted Intuition Thinking Judging	**ENTP** **The Debater** Extraverted Intuition Thinking Perceiving
ESFJ **The Consul** Extraverted Sensing Feeling Judging	**ESFP** **The Entertainer** Extraverted Sensing Feeling Perceiving	**ESTJ** **The Executive** Extraverted Sensing Thinking Judging	**ESTP** **The Entrepreneur** Extraverted Sensing Thinking Perceiving
INFJ **The Advocate** Introverted Intuition Feeling Judging	**INFP** **The Mediator** Introverted Intuition Feeling Perceiving	**INTJ** **The Architect** Introverted Intuition Thinking Judging	**INTP** **The Logician** Introverted Intuition Thinking Perceiving
ISFJ **The Defender** Introverted Sensing Feeling Judging	**ISFP** **The Adventurer** Introverted Sensing Feeling Perceiving	**ISTJ** **The Logistician** Introverted Sensing Thinking Judging	**ISTP** **The Virtuoso** Introverted Sensing Thinking Perceiving

Figure 1. 16 MBTI personality types of acronyms, dubbed names and their traits composition

Aligning with the initial purpose of the MBTI personality type questionnaire in assisting the suitability between an individual's personality type and their job during the World War II era, as highlighted by Geyer (2014), Fatima et al. (2022) direction of study was to suggest a list of suitable careers for individuals through text classification means as based by their determined personality profiles. Whereas in noting the trend that involved the usage of the MBTI personality types, Wang, C. et al. (2024) proved that the emergence of the MBTI personality type testing as a cultural and social phenomenon had led to an influence on the Chinese youth's behaviour and social interactions. The increase in the stated trend may be contributed by a few factors. Based on Hua and Zhou (2023), the increase in the *Barnum effect*, which was defined as "a psychological phenomenon in which people easily accept general and ambiguous personality interpretations", due to the vagueness that came with this personality testing had reduced the anxiety and depression levels among individuals, and thus left a positive effect on the adolescent mental health, as observed in this study's findings. In seeking self-recognition, the MBTI personality type testing has also helped in reaching great satisfaction in terms of an individual's personal identity and collective social interaction (Fan, 2024). In a more personal take, the MBTI personality type test allowed an individual to understand themselves along with people within their circle (Lee & Shin, 2024). Other advantages of discerning one's personality type include obtaining

better language-acquiring tactics as observed in Gu and Sharil (2023) research, as well as tailoring one's language learning method to maximise the learning task, as emphasised in Jiang (2024) research counterpart. A positive effect contributed by the MBTI personality types in the context of academic performance and learning motivation was also discovered in Ke (2024) study.

Myers-Briggs Type Indicator Classification

Dataset

In an earlier study on MBTI personality type classification by Keh and Cheng (2019), the researchers performed manual data scraping on the Personality Café forum for their research dataset. Instead of implementing the same method, a public dataset created by Mitchell (2017) on Kaggle containing over 8600 rows of data on a person's personality type and the last 50 things they wrote in Personality Café forum was mostly opted by several researchers through the expense of different years, namely, Amirhosseini and Kazemian (2020), Khan et al. (2020), Mushtaq et al. (2020), Vaddem and Agarwal (2020), Basto (2021), Choong and Varathan (2021), Kaushal et al. (2021), Maulidah and Pardede (2021), Ren et al. (2021), Ontoum and Chan (2022), Ryan et al. (2023), Zhang (2023), and Wang, Y. (2024). As based on Choong and Varathan (2021) emphasis, the factors that led to utilising the Mitchell (2017) dataset were due to its availability and size, not consisting of microblogs, and had been cited in other research at least twice to enable comparison. A concern from Choong and Varathan (2021) on a different handling of mircroblogs' texts would be due to a few factors emphasised by Stajner and Yenikent (2021) in their study, where these texts were composed of incorrect form of grammar as well as punctuations uses that caused the post author to appear *Extraverted* (E) as based on the text due to the repeated usage of exclamation marks in these platforms. Despite opting for another personality typing known as the Big Five Model, Garg and Garg (2021) had utilised Mitchell (2017) dataset in their research by mapping the MBTI personality type's binary traits to the four personality terms of the Big Five Model, with an exception to the *Neuroticism* trait. Alsini et al. (2024) had adapted the Mitchell (2017) dataset by focusing on the Big Five Model's *Agreeableness* trait by basing the trait on the *Feeling* and *Thinking* binary traits. Adawadkar and Gandhi (2023), on the other hand, chose a different public dataset from Kaggle, where an MBTI personality type-based KPMI dataset that contained survey responses reflecting the respondents' psychological choices was used instead. Some researchers, on the other hand, have utilised different datasets in their studies. A crowd-sourced database from Rev. Emmy Kegler's personality sets was used in Shafi et al. (2021) study, while a combination of datasets featuring Mitchell (2017) dataset, Reddit MBTI9K datasets, and PANDORA dataset was utilised in Cerkez et al. (2021) study. A summary of the datasets utilised by different authors mentioned is shown in Table 1.

In countering the concern about the presence of disproportioned MBTI personality types in classes, the undersampling technique was implemented by Ren et al. (2021),

while both undersampling and oversampling techniques were applied by Khan et al. (2020), Maulidah and Pardede (2021), and Ontoum and Chan (2022) in their respective research. Choosing a different option, Ryan et al. (2023) utilised the Synthetic Minority Oversampling Technique (SMOTE) in their study.

Table 1
Datasets used by different authors

Dataset	Author
Manual data scraping	(Keh & Cheng, 2019)
Single public dataset	(Amirhosseini & Kazemian, 2020) (Khan et al., 2020) (Mushtaq et al., 2020) (Vaddem & Agarwal, 2020) (Basto, 2021) (Choong & Varathan, 2021) (Kaushal et al., 2021) (Maulidah & Pardede, 2021) (Ren et al., 2021) (Shafi et al., 2021) (Ontoum & Chan, 2022) (Adawadkar & Gandhi, 2023) (Ryan et al., 2023) (Zhang, 2023) (Wang, Y., 2024)
Multiple public datasets	(Cerkez et al., 2021)

Data Preprocessing

All the research mentioned in the previous section had performed data preprocessing except for Basto (2021) due to the researcher's take on the limitation posed by data preprocessing and text representation tasks in identifying patterns within the abstracted data. Shafi et al. (2021) study did not detail any preprocessing steps done on the dataset either. A summary of the removed elements and varied preprocessing steps included in different research is shown in Figures 2 and 3, respectively.

Authors	Year	Removed Elements							
		Links	Stopwords	Special Characters	Emoticons	Numbers	One-letter words	Nouns	MBTI profile strings
Keh & Cheng	2019			✓					✓
Amirhosseini & Kazemian	2020	✓	✓						✓
Khan et al.			✓						
Mushtaq et al.		✓							✓
Vaddem & Agarwal		✓	✓	✓					✓
Cerkez et al.	2021	✓	✓	✓		✓	✓		
Choong & Varathan		✓	✓					✓	✓
Garg & Garg			✓						✓
Kaushal et al.		✓	✓						
Maulidah & Pardede		✓	✓	✓	✓				
Ren et al.		✓	✓	✓					
Ontoum & Chan	2022	✓	✓						
Ryan et al.	2023	✓	✓	✓					
Zhang			✓	✓					
Ashraf et al.	2024	✓	✓	✓	✓				
Y. Wang		✓	✓	✓		✓			
Alsini et al.		✓	✓	✓					
Hartono et al.	2025		✓						
Yang et al.			✓						

Figure 2. Removed elements in the preprocessing step

Authors	Year	Preprocessing Steps							
		Links & Usernames Conversion	Lowercase Conversion	Shortforms Conversion	Emoticons Conversion	Numeric Digits Conversion	Text Tokenisation	Text Lemmatisation	Text Stemming
Keh & Cheng	2019		✓	✓			✓		
Amirhosseini & Kazemian	2020							✓	
Khan et al.							✓		✓
Mushtaq et al.			✓					✓	
Vaddem & Agarwal								✓	
Choong & Varathan	2021				✓	✓	✓	✓	
Garg & Garg								✓	
Kaushal et al.							✓	✓	✓
Maulidah & Pardede			✓				✓	✓	
Ontoum & Chan	2022						✓	✓	
Ryan et al.	2023		✓				✓	✓	
Zhang		✓					✓		
Ashraf et al.	2024						✓	✓	
Y. Wang			✓				✓	✓	
Alsini et al.			✓				✓		
Hartono et al.	2025							✓	
Yang et al.			✓						

Figure 3. Further preprocessing steps taken

Feature Selection

The Term Frequency-Inverse Document Frequency (TF-IDF) technique allowed a vocabulary dictionary to be learned by the machine learning models from a token counts matrix, as well as output a term-document matrix (Amirhosseini & Kazemian, 2020). In highlighting this technique's usefulness, Khan et al. (2020) had implied on the TF-IDF score's ability to adjust the weights between the most regular, general words, and less utilised words, while Mushtaq et al. (2020) inputted on the contribution in terms of the capability of detecting words based on its importance to the increase in the machine learning models' performance efficiency. This technique was implemented by several researchers in their respective studies, namely, Amirhosseini and Kazemian (2020), Khan et al. (2020), Vaddem and Agarwal (2020), Mushtaq et al. (2020), Kaushal et al. (2021), Ontoum and Chan (2022), Zhang (2023), and Alsini et al. (2024). Ontoum and Chan (2022) had an additional utilisation of the BOW approach alongside TF-IDF.

Word embeddings were also applied in a few of the observed studies' feature selection step. Ryan et al. (2023) chose the Continuous Bag of Words (CBOW) model upon taking Mikolov et al. (2013) discovery on the quicker training time and better frequent words representation of the CBOW model into consideration. In a comparison between CBOW and Skip-gram models, Ryan et al. (2023) observed that the CBOW model utilised a series of context words in a phrase to predict a single word, while the opposite approach of utilising a single provided word in predicting a series of context words was done by

the Skip-gram model. Apart from that, Ren et al. (2021) had implemented a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model for a sentence-level embedding task, where multiple successive sentences from each sample data were included instead of only separating each sentence from the sample data as done in Keh and Cheng (2019) study. Adapting Keh and Cheng (2019) approach, Zhang (2023) and Ashraf et al. (2024) respective research had utilised the BERT model to perform word embedding, where the latter researchers observed that BERT had better performance than GloVe in the said task. Cerkez et al. (2021) had opted for FastText in performing the word embedding task in their research, while Maulidah and Pardede (2021) utilised an embedding layer with a vector length of 100 for word representation. Besides that, Alsini et al. (2024) experimented with Word2Vec, GloVe, and sentence embedding.

Table 2
Feature selection method utilised by different authors

Feature Selection Method	Author
Term Frequency-Inverse Document Frequency (TF-IDF)	(Amirhosseini & Kazemian, 2020) (Khan et al., 2020) (Vaddem & Agarwal, 2020) (Mushtaq et al., 2020) (Kaushal et al., 2021) (Ontoum & Chan, 2022) (Zhang, 2023) (Alsini et al., 2024)
Word embeddings	(Keh & Cheng, 2019) (Cerkez et al., 2021) (Maulidah & Pardede, 2021) (Ren et al., 2021) (Ryan et al., 2023) (Zhang, 2023) (Ashraf et al., 2024) (Alsini et al., 2024)
Sentiment analysis	(Basto, 2021) (Ren et al., 2021)
Grammar analysis	(Basto, 2021)

In a separate experiment by Basto (2021), the researcher implemented sentiment analysis to capture additional patterns in texts, aspect analysis to pinpoint critical aspects in the dataset, and grammar analysis to calculate the percentage of occurrences of syntactic terms in the personality types of traits. From this research, it was found that the sentiment analysis feature had helped in reaching the best accuracy score in the binary traits' classes of E/I, N/S, and F/T, with the scores of 78.96%, 85.95%, and 75.43%, respectively (Basto, 2021). Whereas for the J/P binary class in the same study by Basto (2021), the highest score of 64.29% was achieved through a feature generated by a BERT model. Similarly, Ren et al. (2021) had also implemented the sentiment analysis as part of the research's feature selection step through the utilisation of a common-sense knowledge extraction tool named SenticNet5.The feature selection methods discussed are summarised in Table 2.

Comparative Analysis

The MBTI personality type can be divided into binary classification and multiclass classification approaches, where the binary classification task was viewed in the context

of separated binary personality traits of E/I, N/S, F/T, and J/P, as based on Briggs Myers and Kirby (2000) research, while the multiclass classification task was viewed in the context where each MBTI personality type was captured as a single class, disregarding the composition of the four binary personality traits of the said MBTI personality type. Hence, unlike the binary classification's multiple outcomes that came from the four binary categories, there will only be one outcome from each model in the multiclass classification approach. An analysis of the classification models for both MBTI personality type binary classification and multiclass classification will be done to discover the best-performing model utilised in different research.

It was observed that the same dataset provided by Mitchell (2017) was consistently used in all the binary classification research discussed. In terms of the utilisation of the XGBoost classification model, Khan et al. (2020) model proved to obtain the best performance as compared to the researchers who utilised the same model, with an achieved score of over 90% in all binary personality traits categories. Whereas for a similar implementation of the RNN model across Ontoum and Chan (2022) and Amirhosseini and Kazemian (2020) studies, Ontoum and Chan (2022) model obtained a better score as compared to the latter. The LR model utilised in Zhang (2023) research scored better accuracy scores in comparison to Vaddem and Agarwal (2020), Basto (2021), Kaushal et al. (2021), Maulidah and Pardede (2021), and Wang, Y. (2024), respectively. Khan et al. (2020) SVM and RF models had proved to perform the best among the similar models implemented across different studies. Being the only research that utilised the CNN model for MBTI personality type classification, Ren et al. (2021) model scored an accuracy of over 80% for all categories. On the other hand, Basto (2021) BERT model had an overall better performance than the researcher's LSTM model. Focusing on the F/T category in observing the Big Five Model's equivalent of the *Agreeableness* trait, Alsini et al. (2024) found that the SVM model had outperformed the other classification models with scores of 84% in all evaluation metrics. Thus, it can be concluded that Khan et al. (2020) XGBoost model had the best accuracy performance in the binary MBTI personality type approach, with the respective scores of 0.9937, 0.9992, 0.9455, and 0.9553 in the E/I, N/S, F/T, and J/P categories. The accuracy scores from the discussed studies are all compiled in Table 3.

A few of the researchers had also included the F_1 score as their model evaluation metric alongside the accuracy score, while Ryan et al. (2023) had opted to only focus on the F_1 score metric in their research. In comparison, Ashraf et al. (2024) RF model was the best-performing classification model due to achieving better F_1 scores than Khan et al. (2020) XGBoost model, Maulidah and Pardede (2021) RF model, Ryan et al. (2023) LR model, Basto (2021) BERT model, and Alsini et al. (2024) LSTM model with sentence embeddings for the F/T category only. The F_1 scores from the mentioned researchers are compiled in Table 4.

Table 3
Accuracy scores compilation for the binary classification approach

Classification Models	Author	MBTI Binary Personality Traits			
		E/I	N/S	F/T	J/P
Extreme Gradient Boosting (XGBoost)	(Amirhosseini & Kazemian, 2020)	0.7817	0.8606	0.7178	0.6570
	(Khan et al., 2020)	0.9937	0.9992	0.9455	0.9553
	(Mushtaq et al., 2020)	0.8901	0.8593	0.8419	0.8542
	(Vaddem & Agarwal, 2020)	0.7621	0.8562	0.7501	0.6376
	(Maulidah & Pardede, 2021)	0.7192	0.7727	0.7050	0.6509
	(Alsini et al., 2024)	-	-	0.8300	-
Recurrent Neural Network (RNN)	(Amirhosseini & Kazemian, 2020)	0.6760	0.6200	0.7780	0.6370
	(Ontoum & Chan, 2022)	0.8359	0.9322	0.8000	0.7740
Logistic Regression (LR)	(Vaddem & Agarwal, 2020)	0.7907	0.8668	0.7787	0.6717
	(Basto, 2021)	0.7896	0.8595	0.7543	0.6333
	(Kaushal et al., 2021)	0.8225	0.8801	0.7935	0.7367
	(Maulidah & Pardede, 2021)	0.7625	0.8634	0.7233	0.6473
	(Zhang, 2023)	0.8937	0.9026	0.9023	0.8121
	(Wang, Y., 2024)	0.8580	0.8930	0.8620	0.7910
	(Alsini et al., 2024)	-	-	0.8400	-
Support Vector Machine (SVM)	(Vaddem & Agarwal, 2020)	0.7639	0.8658	0.7676	0.6842
	(Khan et al., 2020)	0.9594	0.9808	0.9263	0.9137
	(Kaushal et al., 2021)	0.8428	0.8820	0.8345	0.7856
	(Maulidah & Pardede, 2021)	0.8738	0.9759	0.7417	0.7191
	(Ontoum & Chan, 2022)	0.8215	0.8732	0.8049	0.7270
	(Alsini et al., 2024)	-	-	0.8400	-
Random Forest (RF)	(Khan et al., 2020)	0.9836	0.9945	0.8215	0.9162
	(Maulidah & Pardede, 2021)	0.9495	0.9893	0.7119	0.7406
	(Kaushal et al., 2021)	0.7709	0.8612	0.7635	0.6547
	(Alsini et al., 2024)	-	-	0.7900	-
Convolutional Neural Network (CNN)	(Ren et al., 2021)	0.8146	0.9251	0.8357	0.8236
BERT	(Basto, 2021)	0.7646	0.8636	0.6829	0.6429
LSTM	(Basto, 2021)	0.7669	0.8541	0.4897	0.6063
	(Alsini et al., 2024)	-	-	0.9017	-

As opposed to the binary classification, the researchers who attempted the multiclass classification approach had mainly utilised the accuracy score as the evaluation metric. It was also observed that there was a variation in the MBTI personality types of datasets used in different research for this approach. In one of the earliest studies done by Keh and Cheng (2019), the BERT classification model utilised in this research had only achieved an accuracy score of 0.4790 for a self-collected Personality Café forum dataset.

Table 4
F_1 scores compilation for the binary classification approach

Classification Models	Author	MBTI Binary Personality Traits			
		E/I	N/S	F/T	J/P
Extreme Gradient Boosting (XGBoost)	(Khan et al., 2020)	0.9856	0.9975	0.9472	0.9619
	(Maulidah & Pardede, 2021)	0.7182	0.7725	0.7051	0.6505
Logistic Regression (LR)	(Ryan et al., 2023)	0.8389	0.8821	0.8561	0.7578
	(Maulidah & Pardede, 2021)	0.6669	0.6939	0.7131	0.6314
Random Forest (RF)	(Maulidah & Pardede, 2021)	0.9450	0.9886	0.7306	0.7348
	(Ashraf et al., 2024)	0.9898	0.9975	0.9818	0.9823
BERT	(Basto, 2021)	0.3300	0.0000	0.7200	0.7200
Long-Short Term Memory (LSTM)	(Alsini et al., 2024)	-	-	0.9020	-

Table 5
Accuracy scores compilation for the multiclass classification approach

Author	Classification Models	Dataset	Accuracy
(Keh & Cheng, 2019)	BERT	Personality Café forum	0.4790
(Vaddem & Agarwal, 2020)	Extreme Gradient Boosting (XGBoost)	(Mitchell, 2017)	0.3117
(Shafi et al., 2021)	Ensemble Bagged Trees	Rev. Emmy Kegler's dataset	0.7075
(Cerkez et al., 2021)	Convolutional Neural Network (CNN)	(Mitchell, 2017), Reddit MBTI9k datasets & PANDORA dataset	0.6700
(Ontoum & Chan, 2022)	Naïve Bayes (NB)	(Mitchell, 2017)	0.4103
	Support Vector Machines (SVM)		0.4197
	Recurrent Neural Network (RNN)		0.4975
(Wang, Y., 2024)	Logistic Regression (LR)		0.6200

From the compilation of the multiclass MBTI personality type classification shown in Table 5, it was further observed that the highest accuracy was obtained by Shafi et al. (2021) Ensemble Bagged Trees model on Rev. Emmy Kegler's dataset with a score of 0.7075. In another comparison between the studies done on Mitchell (2017) dataset, the best performing model in terms of accuracy was the LR model in Wang, Y. (2024) with a score of 0.6200.

A study by Naz et al. (2025) offered an insight into the considerations on the variety of classification models chosen in the previous studies, where the researchers had discovered that the NB, LR, and SVM models had performed consistently in all four evaluation metrics, while the XGBoost model was noted to have contributed to better results through

overfitting reduction and generalisation enhancement. Besides that, the CNN and LSTM models obtained high scores in the precision and recall metrics, whereas another observation came from the BERT model, which showed to be able to offer variability in dependence on a given architecture and its implementation (Naz et al., 2025).

In gaining a deeper understanding of the MBTI personality types, the traits and psychological functions that make up each personality type were observed. The binary and multiclass MBTI personality type classification approaches were also discussed in depth in this section. As a conclusion, the best performing model in the binary and multiclass classification approaches through the utilisation of Mitchell (2017) dataset was Khan et al. (2020) XGBoost model and Wang, Y. (2024) LR model, respectively.

METHODOLOGY

The steps taken in this study's methodology were text mining, feature generation, machine learning, and model evaluation. The metrics involved in determining the best-performing classification model were accuracy, precision, recall, and F_1 score. Figure 4 shows the methodology processes involved in the study.

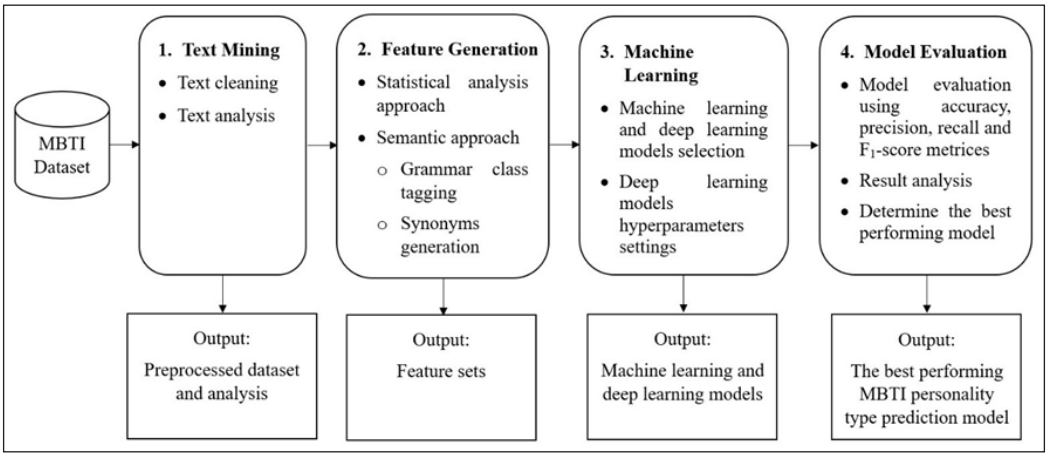


Figure 4. Methodology processes

MBTI Dataset

As done in most of the mentioned past research, this study similarly utilised the MBTI personality type dataset provided by Mitchell (2017) on a public data science platform named Kaggle. This dataset contained a total of 8 675 rows of data with two columns specifying the MBTI personality type of an individual, and their compilation of text posts on the Personality Café forum. The composition of each MBTI personality type's rows within Mitchell (2017) The dataset is shown in Table 6.

Text Mining

There were two steps included under the text mining process, namely, text cleaning and text analysis. In eliminating irrelevant elements from the raw dataset, text cleaning was done through the removal of delimiters, links or Universal Resource Locators (URLs), numbers, punctuations, stop words, and MBTI-related acronyms, lowercasing, shortened words or contractions expansion, lemmatisation, and WordNet library word existence checking. The removal of MBTI-related acronyms was made based on the consideration highlighted by Keh and Cheng (2019), where explicit mentions of the MBTI personality types may "distort the task or make the task too easy" for the classification model. The WordNet library word existence checking step, on the other hand, was taken to ensure that only English words, as referred from the WordNet library, were contained in the cleaned dataset.

Table 6
Composition of each MBTI personality type in Mitchell (2017) dataset

MBTI	Rows
ENFJ	190
ENFP	675
ENTJ	231
ENTP	685
ESFJ	42
ESFP	48
ESTJ	39
ESTP	89
INFJ	1 470
INFP	1 832
INTJ	1 091
INTP	1 304
ISFJ	166
ISFP	271
ISTJ	205
ISTP	337

For text analysis, the composition of total words and unique words in both the raw dataset and the pre-processed dataset will be compared. After going through the layers of preprocessing steps, the total number of words and unique words in the pre-processed dataset has drastically reduced in comparison to that of the raw dataset. The difference between the total words and unique words composition of the two versions of the dataset is shown in Table 7.

Table 7
Total words and total unique words in the raw dataset and pre-processed dataset

Dataset Version	Total Words	Total Unique Words
Raw Dataset	10 628 873	598 960
Pre-processed Dataset	4 391 333	30 625

Feature Generation

The feature generation approaches that were attempted under this process were a statistical analysis approach and a semantic approach, with the latter consisting of grammar class tagging and synonym generation. The details for each of the approaches were further discussed in the following subsections.

Statistical Analysis Approach

The specified statistical analysis approach was the Term Frequency-Inverse Document Frequency (TF-IDF), which was one of the commonly used approaches in the past research discussed. Due to this reason, the TF-IDF will be implemented as a baseline comparison to the other feature generation approaches experimented with in this study. The pseudocode for the steps taken in this approach is shown in Figure 5.

```
START

      IMPORT sklearn library's TfidfVectorizer

      Declare TfidfVectorizer as vectoriser

      READ preprocessed file

      Generate TF-IDF values for words in preprocessed file using vectoriser

      Filter TF-IDF top words

      WRITE TF-IDF top words into feature set file

END
```

Figure 5. Pseudocode for statistical analysis approach

Semantic Approach

There were two semantic approaches implemented under this section, namely, grammar class tagging and synonym generation.

Grammar Class Tagging

For grammar class tagging, only words with the same grammar class will be included within a feature set. With this being said, the grammar classes that will be focused on in this step would be the adjectives (a), verbs (v), and nouns (n). The grammar class filtering was done using WordNet library's Part of Speech (POS) tags, where each word's lemma form would be assigned a POS tag. Hence, relevant words will be accumulated based on their grammar class and will then be part of the feature set of that specific grammar class. The pseudocode for the steps taken in grammar class tagging is shown in Figure 6.

Synonyms Generation

The second approach will involve the generation of synonyms for keywords obtained in the grammar class tagging. By including synonyms as a possible part of the classification models' feature set, words that have a similar meaning to the keywords throughout the dataset would also be able to function as features. The synonyms generation step was also

```
START
      READ preprocessed file
      FOR word in file
            APPEND word into preprocessed word list
      END FOR
      SET preprocessed word list to only contain unique words
      FOR word in preprocessed word list
            Obtain POS tag for word from WordNet library's synsets
            IF POS tag equals to the sought grammar class' POS tag
                  APPEND word into finalised list
            END IF
      END FOR
      Remove empty strings from finalised list
      SET finalised list to only contain unique words
END
```

Figure 6. Pseudocode for grammar class tagging

```
START
      FOR word in finalised list
            APPEND word into synonyms list
            FOR synonyms in WordNet library's synsets
                  Obtain word's lemma
                  APPEND lemma into synonyms list
            END FOR
      END FOR
      Remove empty strings from synonyms list
      Lowercase all words in synonyms list
      SET synonyms list to only contain unique words
END
```

Figure 7. Pseudocode for synonyms generation

done through the utilisation of the WordNet library, and while the generated synonyms may overlap with either the keywords or the other keywords' synonyms, these repetitions will be removed in the finalised version of the feature set by only maintaining the unique words. The pseudocode for the steps taken in synonym generation is shown in Figure 7 below, as proceeded with the example of generated synonyms obtained from this approach in Figure 8.

Words	Synonyms
mythological	fabulous, mythic, mythologic, mythical
neurological	neurologic
prestigious	esteemed, honored
nice	gracious, squeamish, nice, prissy, skillful, overnice, courteous, decent, dainty
solvable	resolvable

Figure 8. Examples of generated synonyms

Combined Approaches

A combination of approaches mentioned in the previous sections into feature sets was also attempted. The variations of combined approaches include TF-IDF Top words with grammar class tagging filtering, as well as TF-IDF top words with grammar class tagging filtering + synonyms for both standard column representation and synset column representation.

TF-IDF Top Words with Grammar Class Tagging Filtering

For this approach, the words filtered out from the TF-IDF step will undergo another layer of filtering based on grammar class tagging, and thus, further reducing the feature set to only containing top words from TF-IDF that belonged to a certain grammar class. The pseudocode for this combination of approaches is shown in Figure 9.

```
START
      IMPORT sklearn library's TfidfVectorizer
      Declare TfidfVectorizer as vectoriser
      READ preprocessed file
      Generate TF-IDF values for words in preprocessed file using vectoriser
      Filter TF-IDF top words
      WRITE TF-IDF top words into TF-IDF feature set file
      READ TF-IDF feature set file
      FOR word in file
            APPEND word into TF-IDF word list
      END FOR
      FOR word in TF-IDF word list
            Obtain POS tag for word from WordNet library's synsets
            IF POS tag equals to the sought grammar class' POS tag
                  APPEND word into finalised list
            END IF
      END FOR
      Remove empty strings from finalised list
      SET finalised list to only contain unique words
END
```

Figure 9. Pseudocode for TF-IDF top words with grammar class tagging filtering

TF-IDF Top Words with Grammar Class Tagging Filtering + Synonyms

Extending the above combined approaches, words obtained from the synonyms generation were added into this variation, where words that were filtered by both the TF-IDF and grammar class tagging will undergo the synonyms generation step in expanding the words contained in the feature set. The pseudocode for this combined approach is shown in Figure 10.

```
START
      IMPORT sklearn library's TfidfVectorizer
      Declare TfidfVectorizer as vectoriser
      READ preprocessed file
      Generate TF-IDF values for words in preprocessed file using vectoriser
      Filter TF-IDF top words
      WRITE TF-IDF top words into TF-IDF feature set file
      READ TF-IDF feature set file
      FOR word in file
            APPEND word into TF-IDF word list
      END FOR
      FOR word in TF-IDF word list
            Obtain POS tag for word from WordNet library's synsets
            IF POS tag equals to the sought grammar class' POS tag
                  APPEND word into finalised list
            END IF
      END FOR
      Remove empty strings from finalised list
      SET finalised list to only contain unique words
      FOR word in finalised list
            APPEND word into synonyms list
            FOR synonyms in WordNet library's synsets
                  Obtain word's lemma
                  APPEND lemma into synonyms list
            END FOR
      END FOR
      Remove empty strings from synonyms list
      Lowercase all words in synonyms list
      SET synonyms list to only contain unique words
END
```

Figure 10. Pseudocode for TF-IDF top words with grammar class tagging filtering + synonyms

Machine Learning

Models Selection

Based on the MBTI personality type classifications observed in the previous research, there was a clear utilisation of machine learning and deep learning models. Taking the same

consideration into account, both machine learning and deep learning models were similarly utilised for the classification task. Table 8 shows the models that were implemented in this study.

Table 8
Machine learning and deep learning models utilised

Model Type	Utilised Models	Model / Layer Name
Machine Learning	Gaussian Naïve Bayes (GNB)	GaussianNB
	Multinomial Naïve Bayes (MNB)	MultinomialNB
	Support Vector Machine (SVM)	SVC
	Logistic Regression (LR)	LogisticRegression
Deep Learning	Convolutional Neural Network (CNN)	Conv1D
	Recurrent Neural Network (RNN), with Long-Short Term Memory (LSTM) layer	LSTM
	Distil Bidirectional Encoder Representations from Transformers (DistilBERT)	distilbert-base-uncased

Machine Learning Models Hyperparameters

The GNB and the MNB models had maintained the default hyperparameter settings, while some alterations were made on the SVM model's "gamma" hyperparameter and the LR models' "solver" and "random_state" hyperparameters. The hyperparameters for the GNB, MNB, SVM, and LR models utilised in this study are shown in the respective order of Tables 9, 10, 11, and 12.

Table 9
GNB model's hyperparameters details

Gaussian Naïve Bayes (GNB)	
Priors	None
Var_smoothing	1e-09

Table 10
MNB model's hyperparameters details

Multinomial Naïve Bayes (MNB)	
Alpha	1.0
Force_alpha	True
Fit_prior	True
Class_prior	None

Table 11
SVM model's hyperparameters details

Support Vector Machine (SVM)	
C	1.0
Kernel	rbf
Degree	3
Gamma	auto
Coef0	0.0
Shrinking	True
Probability	False
Tol	0.001
Cache_size	200
Class_weight	None
Verbose	False
Max_iter	-1
Decision_function_shape	ovr
Break_ties	False
Random_state	None

Deep Learning Models Hyperparameters

The hyperparameters for the LSTM and CNN models, as well as the DistilBERT model utilised in this study, are shown in Tables 13 and 14, respectively.

Training and Testing

The train_test_split function was implemented to set the size of the train and test sets from the dataset for the proceeding classification task. For this study, the train and test split proportion utilised was 80-20.

Model Evaluation

The performance of all implemented machine learning and deep learning models was evaluated using accuracy, precision, recall, and F_1 score metrics, as derived from the confusion matrix.

RESULTS AND DISCUSSION

Feature Sets

The feature sets obtained from the different feature generation approaches mentioned in the previous section had involved four feature sets from statistical analysis approach, three feature sets from semantic approach's grammar class tagging, three feature sets from semantic approach's synonyms generation with standard column representation, three feature sets from semantic approach's synonyms generation with synset column representation, with twelve feature sets for each of the combined approaches of statistical analysis approach with semantic approach's grammar class tagging filtering, statistical analysis

Table 12
LR model's hyperparameters details

Logistic Regression (LR)	
Penalty	l2
Dual	False
Tol	0.0001
C	1.0
Fit_intercept	True
Intercept_scaling	1
Class_weight	None
Random_state	1
Solver	liblinear
Max_iter	100
Multi_class	deprecated
Verbose	0
Warm_start	False
N_jobs	None
L1_ratio	None

Table 13
LSTM and CNN models' hyperparameter details

Hyperparameters	Long-Short Term Memory (LSTM)	Convolutional Neural Network (CNN)
Batch	448	448
Epoch	25	25
Layer name	LSTM	Conv1D
Layers	2 layers (384, 256)	3 layers (128, 64, 32)
Dropout layer	2 layers (0.25, 0.25)	-
Activation	Softmax	Sigmoid
Optimiser	Adam	Adam

Table 14
DistilBERT model's hyperparameters details

Distil Bidirectional Encoder Representations from Transformers (DistilBERT)	
Batch	32
Learning rate	10^{-5}
Early stopping epoch	5
Warm-up steps	100
Weight decay	0.01

approach with semantic approach's grammar class tagging filtering and synonyms generation with standard column representation, and statistical analysis approach with semantic approach's grammar class tagging filtering and synonyms generation with synset column representation. All of the stated feature sets are specified in Table 15.

Table 15
Feature sets and their count across different feature generation approaches

Feature Generation Approach	Feature Set
Statistical Analysis (TF-IDF)	Top 1 000 words (1 000 features)
	Top 5 000 words (5 000 features)
	Top 10 000 words (10 000 features)
	Top 25 000 words (25 000 features)
Semantic: Grammar Class Tagging	Adjectives (2 827 features)
	Verbs (3 199 features)
	Nouns (23 087 features)
Semantic: Synonyms Generation (Standard Representation)	Adjectives + Synonyms (6 934 features)
	Verbs + Synonyms (10 916 features)
	Nouns + Synonyms (35 736 features)
Semantic: Synonyms Generation (Synset Representation)	Adjectives + Synonyms (2 229 features)
	Verbs + Synonyms (2 354 features)
	Nouns + Synonyms (15 375 features)
Statistical Analysis + Semantic: Statistical Analysis (TF-IDF) Grammar Class Tagging Filtered	Top 1 000 Adjectives (122 features)
	Top 1 000 Verbs (117 features)
	Top 1 000 Nouns (629 features)
	Top 5 000 Adjectives (539 features)
	Top 5 000 Verbs (584 features)
	Top 5 000 Nouns (3 532 features)
	Top 10 000 Adjectives (977 features)
	Top 10 000 Verbs (1 170 features)
	Top 10 000 Nouns (7 268 features)
	Top 25 000 Adjectives (2 288 features)
	Top 25 000 Verbs (2 779 features)
	Top 25 000 Nouns (18 626 features)
Statistical Analysis + Semantic: Statistical Analysis (TF-IDF) Grammar Class Tagging Filtered + Synonyms Generation (Standard Representation)	Top 1 000 Adjectives Synonyms (1 007 features)
	Top 1 000 Verbs Synonyms (1 160 features)
	Top 1 000 Nouns Synonyms (4 222 features)
	Top 5 000 Adjectives Synonyms (2 611 features)
	Top 5 000 Verbs Synonyms (3 934 features)
	Top 5 000 Nouns Synonyms (12 449 features)
	Top 10 000 Adjectives Synonyms (3 666 features)

Table 15 *(continued)*

Feature Generation Approach	Feature Set
Statistical Analysis + Semantic: Statistical Analysis (TF-IDF) Grammar Class Tagging Filtered + Synonyms Generation (Synset Representation)	Top 10 000 Verbs Synonyms (6 202 features)
	Top 10 000 Nouns Synonyms (18 464 features)
	Top 25 000 Adjectives Synonyms (6 135 features)
	Top 25 000 Verbs Synonyms (10 177 features)
	Top 25 000 Nouns Synonyms (31 393 features)
	Top 1 000 Adjectives Synonyms (98 features)
	Top 1 000 Verbs Synonyms (94 features)
	Top 1 000 Nouns Synonyms (404 features)
	Top 5 000 Adjectives Synonyms (445 features)
	Top 5 000 Verbs Synonyms (483 features)
	Top 5 000 Nouns Synonyms (2 258 features)
	Top 10 000 Adjectives Synonyms (787 features)
	Top 10 000 Verbs Synonyms (950 features)
	Top 10 000 Nouns Synonyms (4 743 features)
	Top 25 000 Adjectives Synonyms (1 805 features)
	Top 25 000 Verbs Synonyms (2 092 features)
	Top 25 000 Nouns Synonyms (12 504 features)

Document-term Matrix Representation

Before proceeding with the classification step, a document-term matrix was generated for each of the feature sets involved, where the presence of a particular feature in each row of posts was marked with either "1" for a present word or "0" for an absent word. Since each column represented a single feature, the size of the document-term matrix will expand based on the size of the feature set itself. Thus, a feature set with a larger number of features will have a more extensive document-term matrix size, given that a single feature is represented in its respective column within the matrix.

To reduce the dimensionality of the document-term matrix, a word vector based on synset column representation for the semantic approach's synonyms generation feature sets was experimented with in this study. For this take, each column of the document-term matrix will represent the presence of a feature and its synonyms instead of having one column representing only one feature's presence throughout the dataset, as done in the standard representation previously explained. Hence, the classification runs for both the standard and synset column representations for the semantic approach's synonyms generation feature sets will be included in the following section as well.

Model Evaluation

The best classification results done on the statistical analysis approach, semantic approach's grammar class tagging, semantic approach's synonyms generation with standard representation, semantic approach's synonyms generation with synset column representation, alongside the attempts done on a combination of feature generation approaches, namely, statistical analysis with semantic approach's grammar class tagging filtering, statistical analysis with semantic approach's grammar class tagging filtering + synonyms generation with standard column representation, and statistical analysis with semantic approach's grammar class tagging filtering + synonyms generation with synset column representation, are as shown in Table 16**.** Due to the limitation presented by the extensive size of the document-term matrix involved, the classification attempt on the semantic approach's nouns + synonyms feature set with standard column representation, and the combined feature generation approaches of TF-IDF Top 25 000 nouns + synonyms feature set with standard column representation had not be carried out.

Table 16
The best machine learning models' classification results from all feature generation approaches

Feature Approach	Features	Model	Accuracy	Precision	Recall	F_1 Score
TF-IDF Top 10, 000	10 000	LR	0.80	0.78	0.80	0.78
TF-IDF Top 25, 000	25 000	LR	0.89	0.88	0.89	0.88
Nouns Only	23 087	LR	0.89	0.88	0.89	0.88
Verbs + Synonyms	10 916	LR	0.87	0.86	0.87	0.86
Nouns + Synonyms	15 375	LR	0.77	0.76	0.77	0.76
TF-IDF Top 10,000 Nouns	7 268	LR	0.89	0.88	0.89	0.88
TF-IDF Top 25, 000 Nouns	18 626	LR	0.89	0.88	0.89	0.88
TF-IDF Top 5, 000 Nouns + Synonyms	12 449	LR	0.89	0.88	0.89	0.88
TF-IDF Top 10, 000 Nouns + Synonyms	18 464	LR	0.89	0.88	0.89	0.88
TF-IDF Top 10, 000 Nouns + Synonyms	4 743	LR	0.75	0.74	0.75	0.74
TF-IDF Top 25, 000 Nouns + Synonyms	12 504	GNB	0.75	0.73	0.75	0.72

Proceeding from the machine learning models, Table 17 shows the classification results for the deep learning models' MBTI personality type classifications. The DistilBERT model had outperformed the CNN and RNN models in all the tested evaluation metrics.

From the machine learning classification runs done, the highest score across all evaluation metrices was obtained by the LR model through the utilisation of TF-IDF Top 10 000, TF-IDF Top 25 000, nouns only, TF-IDF Top 10 000 nouns, TF-IDF Top 25 000 nouns,

Table 17
Deep learning models' classification result

Model	Accuracy	Precision	Recall	F_1 Score
CNN	0.86	0.86	0.86	0.85
RNN (LSTM layer)	0.86	0.85	0.86	0.86
DistilBERT	0.88	0.87	0.88	0.87

TF-IDF Top 5 000 nouns + synonyms with standard column representation, and TF-IDF Top 10 000 nouns + synonyms with standard column representation feature sets, where all had a shared score of 0.89 for accuracy, 0.88 for precision, 0.89 for recall, and 0.88 for F_1 score. Though in terms of feature set size, the TF-IDF Top 10 000 nouns from the statistical analysis with the semantic approach's grammar class tagging filtering had the smallest feature set size. Thus, the best feature set and model combination from the machine learning classification experimented in this study would be the TF-IDF Top 10 000 nouns feature set through the utilisation of the LR model. Whereas for deep learning, DistilBERT achieved the best classification result as compared to the other tested models with a score of 0.88 for accuracy, 0.87 for precision, 0.88 for recall, and 0.87 for F_1 score.

DISCUSSION

Feature Sets Interpretability Comparison

In comparing the feature sets that contribute to the classification model's high performance in the context of human interpretability and understanding, part of the statistical analysis, the semantic analysis' grammar class tagging only, and the combined approaches of statistical analysis with grammar class tagging filtering feature sets' content are accumulated in Table 18.

Since TF-IDF had relied on the statistical calculations regarding the presence of words within the dataset, the words accumulated through this approach may contain random words that had high TF-IDF values, and thus, the top words from the statistical approach

Table 18
Part of the content from TF-IDF top 1000, adjectives only, and TF-IDF top 1000 adjectives' feature sets

Feature Set	Content
TF-IDF Top 1 000	like, think, people, know, one, get, really, time, feel, thing, make, say, much, well, want, love, type, good, way, see, go, also, lot, even, always, …
Adjectives Only	ingenuous, evangelical, neutral, graphical, rendezvous, unsharpened, fallacious, unsympathetic, theocratic, untreated, antic, supranational, …
TF-IDF Top 1000 Adjectives	related, particular, green, little, normal, huge, strong, black, open, romantic, impossible, live, special, entire, intuitive, new, several, …

had been based on the ranking for each word's calculated value. In a way, the computed TF-IDF values may help the classification model to capture patterns statistically and thus utilise the words included in the feature set to perform classification.

Whereas for the semantic analysis approach, the feature sets under this approach had focused on the classification of words based on the English grammar classes. Taking the adjectives-only feature set as an example, the words accumulated in this feature set all fall under the category of the same grammar class, namely "adjectives", with the characteristic of the words contained being "a word that describes a person or thing" (Oxford University Press, n.d.). Since this relates to the basic human understanding of the English language, the grouping of words through the semantic analysis' grammar class tagging would be clearer to the human understanding as compared to the calculations of values implemented in the statistical analysis approach. Despite this, the TF-IDF feature sets had obtained a much better classification result as compared to those of the grammar class tagging's feature sets, and thus, this led to the preferred implementation of the statistical approach in classification tasks to achieve a higher classification performance over better human interpretability.

Since the domain of this research focuses on the MBTI personality types of individuals, which involves the patterns discerned from human traits, relying solely on the statistical approach may not help in gaining a better understanding of the connection that may exist between the words used by individuals and their respective MBTI personality type. Seizing the chance of compromising between the statistical approach and the semantic approach's grammar class tagging, feature sets composed of these two approaches were tested in this study. The accuracy result achieved between the utilisation of the TF-IDF top words and the TF-IDF top words with grammar class tagging filtering feature sets were the same, which was 0.89. Further comparison between the feature set sizes of the two approaches had proved that the latter had achieved the same high performance with a much smaller feature set size, where the TF-IDF Top 10 000 nouns filtered feature set contained 7 268 features, while the TF-IDF Top 10 000 feature set contained 10 000 features. Hence, instead of solely relying on the calculated values from the statistical analysis approach, combining it with the semantic approach's grammar class tagging filtering has helped to reduce the size of the document-term matrix in achieving a good performance in the MBTI personality type classification.

Linear Regression and DistilBERT Models Comparison

The comparisons on the confusion matrices and the Receiver Operating Characteristic (ROC) curve graphs for the LR model with TF-IDF Top 10 000 nouns feature set and DistilBERT model, and the top 4 wrong MBTI personality types of predictions across the LR and DistilBERT models are shown in Figure 11, Figure 12, Figure 13, Figure 14, and Table 15, respectively. As observed from the confusion matrices comparison in Figures 11 and 12, both models seemed to similarly encounter low correct predictions on the INFJ, INFP, INTJ, and INTP personality types. Similarly, the ROC curve graphs for the LR model and the DistilBERT model in Figure 13 and Figure 14, respectively, showed that

the four stated MBTI personality types plotted lines had steered the closest to the diagonal line that indicated the neutrality between a true and a false prediction. From here, it can be concluded that both models encountered confusion when predicting the INFJ, INFP, INTJ, and INTP personality types. Since the four stated MBTI personality types were ranked with the highest number of posts contained in the dataset, this may be a contributing factor to the confusion in the models' predictions. While that, the predictions done on the MBTI personality types with a lesser number of posts were solidified through the utilisation of the oversampling technique.

From the wrong predictions for both models, as shown in Table 19, it can be concluded that most prediction confusion by the models had mostly centred on the MBTI personality types with the Intuition (N) trait and thus caused the mislabels to exist between personality types with the N trait.

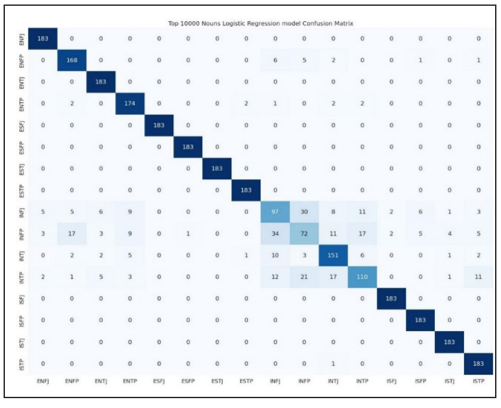


Figure 11. Logistic Regression model confusion matrix

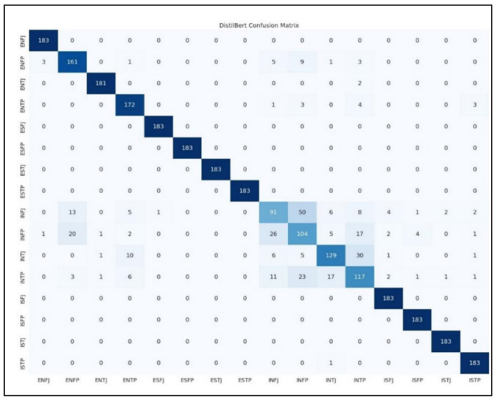


Figure 12. DistilBert model confusion matrix

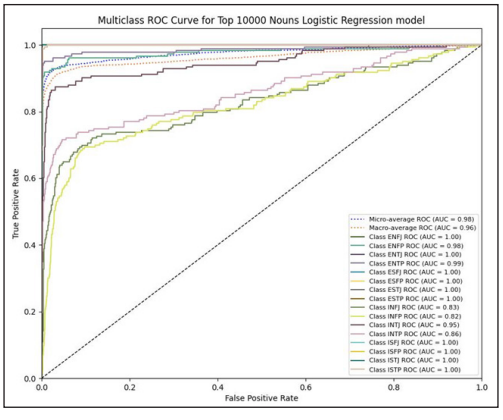


Figure 13. Logistic Regression model ROC curve

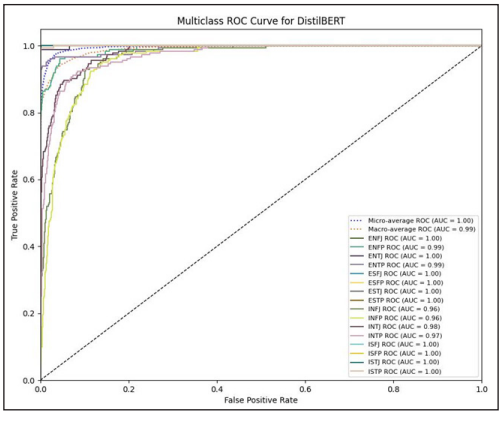


Figure 14. DistilBERT model ROC curve

Table 19
Comparison of the top 4 wrong predicted MBTI personality types across LR and DistilBERT models

Classification Model	Actual Label	Wrong Prediction
Linear Regression Model (LR)	INFJ	INFP, INTP, ENTP
	INFP	INFJ, ENFP, INTP
	INTJ	INFJ, INTP, ENTP
	INTP	INFP, INTJ, INFJ
DistilBERT Model	INFJ	INFP, ENFP, INTP
	INFP	INFJ, ENFP, INTP
	INTJ	INTP, ENTP, INFJ
	INTP	INFP, INTJ, INFJ

CONTRIBUTIONS

The contributions for this study include:

1. Features the synonyms generation

   As part of the feature generation step, this study has introduced an extension of features in the synonym generation approach through the inclusion of words' synonyms, as obtained from the WordNet library, alongside the features obtained from the grammar class tagging approach, to capture the variation of words that share the same meaning.

2. Document-term matrix size reduction

   An attempt to reduce the size of the document-term matrix was also experimented with in the synset column representation for the semantic approach's synonyms generation, where each column of the document-term matrix was represented by the presence of a feature and its synonyms, as opposed to having one column representing only a single feature's presence as done in the standard column representation of the matrix.

CONCLUSION

Based on the comparison of different feature sets used for the MBTI personality type classification done in this study, it can be concluded that the best performing classification model in terms of accuracy, precision, recall, and F_1 score metrices was the Logistic Regression (LR) model through the utilisation of the statistical analysis approach's TF-IDF Top 10 000 and Top 25 000 words, grammar class tagging's nouns only, statistical analysis with grammar class tagging filtering's TF-IDF Top 10 000 nouns and TF-IDF Top 25 000 nouns, statistical analysis with grammar class tagging filtering + synonyms generation with standard column representation's TF-IDF Top 5 000 nouns + synonyms and TF-IDF Top 10 000 nouns + synonyms as features.

The best classification result achieved from the deep learning model's utilisation had come from the DistilBERT model, with slightly lower scores across the evaluation metrics in comparison to the LR model. In an overall view, both the LR model and the DistilBERT model obtained a high score of above 80% for accuracy, precision, recall, and F_1 score metrics for the MBTI personality type classification task.

Due to the limitations encountered in this study, a possible direction to be pursued for future works related to the MBTI personality type classification would be to explore the utilisation of all nouns and their synonyms as features. Apart from that, focusing on the possible alternatives of reducing the document-term matrix size that contains many features without negatively affecting the performance of the classification model would benefit in improving the time taken for the model to perform the classification task.

ACKNOWLEDGEMENT

We extend our sincere gratitude to the Faculty of Computer Science and Information Technology (FCSIT) at Universiti Malaysia Sarawak (UNIMAS) for the facilities provided throughout this research. We would also like to thank Universiti Malaysia Sarawak (UNIMAS) for supporting the fees for this article. We confirm that this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

LIST OF ABBREVIATIONS

ENFJ : Extraverted-Intuition-Feeling-Judging Type
ENFP : Extraverted-Intuition-Feeling-Perceiving Type
ENTJ : Extraverted-Intuition-Thinking-Judging Type
ENTP : Extraverted-Intuition-Thinking-Perceiving Type
ESFJ : Extraverted-Sensing-Feeling-Judging Type
ESFP : Extraverted-Sensing-Feeling-Perceiving Type
ESTJ : Extraverted-Sensing-Thinking-Judging Type
ESTP : Extraverted-Sensing-Thinking-Perceiving Type
INFJ : Introverted-Intuition-Feeling-Judging Type
INFP : Introverted-Intuition-Feeling-Perceiving Type
INTJ : Introverted-Intuition-Thinking-Judging Type
INTP : Introverted-Intuition-Thinking-Perceiving Type
ISFJ : Introverted-Sensing-Feeling-Judging Type
ISFP : Introverted-Sensing-Feeling-Perceiving Type
ISTJ : Introverted-Sensing-Thinking-Judging Type
ISTP : Introverted-Sensing-Thinking-Perceiving Type
MBTI : Myers-Briggs Type Indicator

REFERENCES

Adawadkar, K., & Gandhi, V. (2023). Comparative analysis of classification algorithms for classifying psychotypes. *Proceedings of the 2023 3rd International Conference on Smart Data Intelligence (ICSMDI 2023)*, 483-488. https://doi.org/10.1109/ICSMDI57622.2023.00091

Akber, M. A., Ferdousi, T., Ahmed, R., Asfara, R., Rab, R., & Zakia, U. (2024). Personality and emotion: A comprehensive analysis using contextual text embeddings. *Natural Language Processing Journal, 9*, Article 100105.https://doi.org/10.1016/j.nlp.2024.100105

Alsini, R., Naz, A., Khan, H. U., Bukhari, A., Daud, A., & Ramzan, M. (2024). Using deep learning and word embeddings for predicting human agreeableness behaviour. *Scientific Reports, 14*(1), Article 29875. https://doi.org/10.1038/s41598-024-81506-8

Amirhosseini, M. H., & Kazemian, H. (2020). Machine learning approach to personality type prediction based on the Myers–Briggs Type Indicator®. *Multimodal Technologies and Interaction, 4*(1). https://doi.org/10.3390/mti4010009

Ashraf, N., Ahmad, R. S., Bano, S., Azeem, H. M., & Naz, S. (2024). Enhancing MBTI personality prediction from text data with an advanced word embedding technique. *VFAST Transactions on Software Engineering, 12*(3), 35-43. https://doi.org/10.21015/vtse.v12i3.1864

Basto, C. (2021). Extending the abstraction of personality types based on MBTI with machine learning and natural language processing (NLP). *arXiv*. https://doi.org/10.48550/arXiv.2105.11798

Briggs Myers, I., & Kirby, L. K. (2000). *Introduction to type: A guide to understanding your results on the Myers-Briggs Type Indicator*. Consulting Psychologists Press.

Cerkez, N., Vrdoljak, B., & Skansi, S. (2021). A method for MBTI classification based on the impact of class components. *IEEE Access, 9*, 146550-146567. https://doi.org/10.1109/ACCESS.2021.3121137

Choong, E. J., & Varathan, K. D. (2021). Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum. *PeerJ, 9*, Article e11382. https://doi.org/10.7717/peerj.11382

Fan, X. (2024). Research on audience psychological communication in the era of new media-Taking the MBTI phenomenon as an example. *SHS Web of Conferences, 199*, Article 02018. https://doi.org/10.1051/shsconf/202419902018

Fatima, N., Gul, S., Ahmed, J., Khand, Z. H., & Mujtaba, G. (2022). A rule-based machine learning model for career selection through MBTI personality. *Mehran University Research Journal of Engineering and Technology, 41*(2), 185-196. https://doi.org/10.22581/muet1982.2202.18

Garg, S., & Garg, A. (2021). Comparison of machine learning algorithms for content-based personality resolution of tweets. *Social Sciences and Humanities Open, 4*(1), Article 100178. https://doi.org/10.1016/j.ssaho.2021.100178

Geyer, P. (2014). *Isabel Briggs Myers, psychological type, and the spirit of C.G. Jung*.

Gu, Y., & Sharil, W. N. E. H. (2023). Study on the effect of personality type on the language learning strategies of non-English major students through MBTI test. *Educational Administration Theory and Practice Journal, 29*(4), 1-19. https://doi.org/10.52152/kuey.v29i4.756

Hua, J., & Zhou, Y. X. (2023). Personality assessment usage and mental health among Chinese adolescents: A sequential mediation model of the Barnum effect and ego identity. *Frontiers in Psychology, 14*, Article 1097068. https://doi.org/10.3389/fpsyg.2023.1097068

Jiang, H. (2024). The impact of personality types on second language vocabulary acquisition of college students: Based on MBTI personality categorisation. *Journal of Education, Humanities and Social Sciences IMPES, 26*, 704-710. https://doi.org/10.54097/c5r4zm20

Kaushal, P., B. P., N. B., M. S., P., S., K., & Koundinya, A. K. (2021). Myers–Briggs personality prediction and sentiment analysis of Twitter using machine learning classifiers and BERT. *International Journal of Information Technology and Computer Science, 13*(6), 48-60. https://doi.org/10.5815/ijitcs.2021.06.04

Ke, M. (2024). The influence of MBTI personality types on college students' academic performance: The mediating role of learning motivation. *Journal of Education, Humanities and Social Sciences PSHE, 29*, 449-454.https://doi.org/10.54097/vysk5519

Keh, S. S., & Cheng, I.-T. (2019). Myers-Briggs personality classification and personality-specific language generation using pre-trained language models. *arXiv*. https://doi.org/10.48550/arXiv.1907.06333

Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arif, A., & Khalid, H. A. (2020). Personality classification from online text using machine learning approach. *International Journal of Advanced Computer Science and Applications, 11*(3), 460-476. https://doi.org/10.14569/IJACSA.2020.0110358

Lee, H., & Shin, Y. (2024). A study on MBTI perceptions in South Korea: Big data analysis from the perspective of applying MBTI to contribute to the sustainable growth of communities. *Sustainability, 16*(10), Article 4152. https://doi.org/10.3390/su16104152

Li, W. (2021). *Predicting MBTI personality type of Twitter users* [Doctoral thesis, Rutgers University]. Rutgers University Libraries. https://doi.org/doi:10.7282/t3-75wc-2x18

Maulidah, M., & Pardede, H. F. (2021). Prediction of Myers–Briggs Type Indicator personality using long short-term memory. *Jurnal Elektronika dan Telekomunikasi, 21*(2), 104-111. https://doi.org/10.14203/jet.v21.104-111

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv*. https://doi.org/10.48550/arXiv.1310.4546

Mitchell, J. (2017). *(MBTI) Myers-Briggs personality type dataset* [Data set]. https://www.kaggle.com/datasets/datasnaek/mbti-type

Mushtaq, Z., Ashraf, S., & Sabahat, N. (2020). Predicting MBTI personality type with K-means clustering and gradient boosting. *Proceedings of the 2020 23rd IEEE International Multi-Topic Conference (INMIC 2020)*. https://doi.org/10.1109/INMIC50486.2020.9318078

Naz, A., Khan, H. U., Bukhari, A., Alshemaimri, B., Daud, A., & Ramzan, M. (2025). Machine and deep learning for personality traits detection: A comprehensive survey and open research challenges. *Artificial Intelligence Review, 58*(8), Article 11245. https://doi.org/10.1007/s10462-025-11245-3

NERIS Analytics Limited. (2013). *Our framework*.

Ontoum, S., & Chan, J. H. (2022). Personality type based on Myers-Briggs Type Indicator with text posting style using traditional and deep learning. *arXiv*. https://doi.org/10.48550/arXiv.2201.08717

Oxford University Press. (n.d.). *Adjective*. In *Oxford Advanced Learner's Dictionary*. Retrieved March 3, 2024, from https://www.oxfordlearnersdictionaries.com/definition/english/adjective

Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management, 58*(3), Article 102532. https://doi.org/10.1016/j.ipm.2021.102532

Riza, M. A., & Charibaldi, N. (2021). Emotion detection in Twitter social media using long short-term memory (LSTM) and fastText. *International Journal of Artificial Intelligence & Robotics, 3*(1), 15-26. https://doi.org/10.25139/ijair.v3i1.3827

Ryan, G., Katarina, P., & Suhartono, D. (2023). MBTI personality prediction using machine learning and SMOTE for balancing data based on statement sentences. *Information, 14*(4), Article 217.https://doi.org/10.3390/info14040217

Shafi, H., Sikender, A., Jamal, I. M., Ahmad, J., & Aboamer, M. A. (2021). A machine learning approach for personality type identification using MBTI framework. *Journal of Independent Studies and Research Computing, 19*(2). https://doi.org/10.31645/JISRC.43.19.2.2

Stajner, S., & Yenikent, S. (2021). Why is MBTI personality detection from texts a difficult task? In P. Merlo, J. Tiedemann, & R. Tsarfaty (Eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 3580-3589). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.312

Vaddem, N., & Agarwal, P. (2020). Myers-Briggs personality prediction using machine learning techniques. *International Journal of Computer Applications, 175*(23), 41-44. https://doi.org/10.5120/ijca2020920764

Wang, C., Gao, Y., & Xie, Y. (2024a). Analysing the dissemination of the Myers-Briggs Type Indicator (MBTI) in China: A case study of Weibo and Xiaohongshu texts. *EAI Endorsed Transactions on Smart Cities*. https://doi.org/10.4108/eai.15-3-2024.2346532

Wang, Y. (2024b). Logistic regression model for personality type prediction based on the Myers-Briggs Type Indicator. *Transactions on Computer Science and Intelligent Systems Research, 7*, 206-215.https://doi.org/10.62051/4d9gv137

Xue, D., Hong, Z., Guo, S., Gao, L., Wu, L., Zheng, J., & Zhao, N. (2017). Personality recognition on social media with label distribution learning. *IEEE Access, 5*, 13478-13488. https://doi.org/10.1109/ACCESS.2017.2719018

Zhang, H. (2023). MBTI personality prediction based on BERT classification. *Highlights in Science, Engineering and Technology CSIC, 34*, 368-374. https://doi.org/10.54097/hset.v34i.5497

Pertanika J. Sci. & Technol. 34 (2): 1107 - 1138 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Comparative Air Gasification of Agricultural Residues in an Indigenously Designed Fluidised Bed Gasifier

Shahbaz Hussain[1,3]*, Abdul Nasir[1], Ch. Arslan[1], Shakeel Ahmad Anjum[2], and Sibel Irmak[3]

[1]Department of Structures and Environmental Engineering, University of Agriculture, 38040 Faisalabad, Pakistan

[2]Department of Agronomy, University of Agriculture, 38040 Faisalabad, Pakistan

[3]Department of Agricultural and Biological Engineering, The Pennsylvania State University, 436 Shortlidge Rd, Agricultural Engineering Building, Room 227, 16802 University Park, PA, USA

ABSTRACT

This study evaluates the performance and operational feasibility of a low-cost, indigenously developed pilot-scale fluidised bed gasifier operated under fixed conditions using multiple biomass feedstocks. The adaptability of the reactor for multi-biomass gasification was assessed using sawdust (SD), wheat straw (WS), rice husk (RH), corn stalk (CS), and spent tea waste (STW). Gasification experiments were carried out in an externally heated fluidised bed reactor using air as the gasifying agent. The reactor was equipped with custom-designed lateral nozzles to ensure uniform air distribution and to minimise particle clogging during prolonged operation. All experiments were conducted at a constant fuel feed rate of 20 kgh^{-1}, operating temperatures between 750 and 950 °C, and equivalence ratios of 0.20, 0.25, and 0.30. Reactor performance was evaluated based on syngas composition (CO, H_2, CH_4), syngas yield, lower heating value (LHV), carbon conversion efficiency (CCE), and cold gas efficiency (CGE). The results showed that at an ER of 0.2, SD produced the highest CO (27.56%) concentration, LHV of 6.57 $MJNm^{-3}$, and maximum CGE of 76%. STW yielded the highest H_2 (21.45%) and CH_4 (3.39%) contents and syngas yield (2.31 Nm^3kg^{-1}) at an ER of 0.3, while CS exhibited the maximum CCE of 82% at an ER of 0.25. CGE values above 60% were obtained for all feedstocks. These findings demonstrate stable reactor operation, effective nozzle performance, and strong adaptability to diverse biomass types. The developed system shows potential for decentralised energy generation and sustainable utilisation of locally available biomass resources.

Keywords: Biomass, carbon conversion efficiency, cold gas efficiency, equivalence ratio, fluidised bed gasifier, lower and higher heating values, syngas yield

ARTICLE INFO

Article history:
Received: 28 August 2025
Accepted: 26 February 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.21

E-mail addresses:
shahbazalvi844@gmail.com (Shahbaz Hussain)
anawan@uaf.edu.pk (Abdul Nasir)
arslan_see@uaf.edu.pk (Ch. Arslan)
shakeel.anjum@uaf.edu.pk (Shakeel Ahmad Anjum)
sji5137@psu.edu (Sibel Irmak)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

In the last decade, the need for reliable and sustainable energy sources has become more important, particularly in Pakistan. With the growth of industries and populations, energy demand is rising. Previously, the world relied on fossil fuel reserves as the primary and most effective source of energy (Salim et al., 2014). High energy prices have added another layer of complexity for industries worldwide (Horobet et al., 2021). In addition, burning fossil fuels is a major cause of acid rain and greenhouse gas emissions due to the release of CO_2, NO_x, SO_x, and other pollutants, which pose a serious threat to the global climate (Najjar, 2011). A lot of research is being done to use renewable energy sources to partially replace fossil fuels. Renewable resources are the best way to reduce energy hazards and improve environmental and economic sustainability (Latif et al., 2020). Solid waste generation and management are other critical and emerging problems in Pakistan. Traditionally, solid wastes are either dumped in landfills or burned, particularly in the case of agricultural waste such as biomass that produces secondary pollutants which damage health and the environment. Pakistan generated approximately 49.6 million tons of solid waste annually, increasing at a rate of 2.4% per year (Sanjrani et al., 2023). Jabeen et al. (2021) estimated the greenhouse gas emissions generated from domestic solid waste in Faisalabad and suggested its potential use as an energy source for both domestic and industrial applications. Therefore, it is crucial to implement alternative waste management strategies, particularly for biomass, that minimise environmental harm and reduce secondary pollutants (Clauser et al., 2021).

Out of all renewable resources currently studied in Pakistan (solar, hydropower, biomass, and wind), biomass is regarded as the best because of its easy accessibility, constant supply, wider availability and unique environmentally friendly nature (Saghir et al., 2019). In Pakistan, total biomass potential is estimated as 50,000 GW hyr^{-1} (Farooqui, 2014), which accounts for up to 36% of the overall energy scenario of the country (Asif, 2009). According to the source, biomass can be classified into three different types: municipal solid waste (organic waste, garden waste, domestic garbage, etc.), industrial waste (pressing byproducts, etc.) and agricultural and forestry waste (crop residues, forestry processing waste, etc. (Sánchez et al., 2019). Biomass is an enormously available resource, geologically scattered across the globe (Devi et al., 2020).

Among all the potential biomass sources, as an agriculture-based economy, Pakistan produces a significant amount of crop residues, specifically rice husk (RH) and rice straw (RS), which are burned at a rate of 135.38 thousand tons and 307.7 thousand tons after crop harvesting to prepare the land for the next crop that release a harmful gas (Memon, 2022). Normally, rice husk is discarded or used in low-value purposes such as barn mating. Five primary crop residues (wheat straw, rice straw, rice husk, cotton straw, corn stover, and bagasse) were estimated and found that, in 2018, approximately 40 million tons of

crop residues were available in Pakistan for power generation, assuming a 50% residue removal factor (Kashif et al., 2020). Safdar et al. (2020) discussed the potential of crop residues as a primary renewable energy source, emphasising their abundant availability in Pakistan. Wheat straw is an abundant agricultural residue that contains high levels of lignocellulose, a key energy source (Nguyen et al., 2013). As a renewable feedstock, it is widely available after wheat harvesting, making it a sustainable biomass resource for energy conversion. Other biomass waste, like sawdust, which is an abundant waste resource obtained on a large scale during industrial processing, requires careful disposal is required for such type of waste. Sawdust is commonly used for various purposes; however, a portion is either openly dumped or burned for heating, leading to environmental issues such as pollution. This poses significant health risks, particularly in developing countries like Pakistan (Batool et al., 2021). Sawdust has been approved for its potential on the syngas composition and gasification performance (Baskara Sethupathy et al., 2012; Chen et al., 2015). Spent Tea Waste (STW) is a high-carbonaceous powdery biomass, which is produced after brewing the tea (Augustine et al., 2021). The global consumption of tea has grown significantly, reaching 6.6 million tons in 2021 (Ben Abdallah et al., 2023). Around 90% of tea consumption results in the generation of tea waste (Tian et al., 2016). In Pakistan, approximately 200,000 metric tons of spent tea waste are produced annually from the consumption of 100 billion cups of tea (Rameeza & Eun, 2022). Tea waste is generally disposed of or used as fertiliser (Raza et al., 2022). However, previous research on the treatment of tea waste, such as pyrolysis, gasification, and carbonisation, remains limited (Xu et al., 2009). Waste-to-Energy (WtE) is one of the strategies to manage and utilise municipal solid waste management (MSW) for energy generation, helping to decrease the dependency on fossil fuels and natural gas. From the previous studies, it has been found that biomass derived from crops, wood, and tea has a great potential for gasification to generate heat and electricity. To the author's knowledge, there is still a lack of studies on the gasification performance of spent tea waste.

Advanced conversion technologies are urgently needed in Pakistan. Biomass energy conversion processes include physical, thermochemical, and biological processes. Thermochemical conversion methods (combustion, gasification, and pyrolysis) (Brown, 2019), have been widely and commercially employed all over the world due to their less wasteful pretreatment process by reducing operational complexity and cost. Among all thermochemical conversion methods, gasification is considered the best due to its high efficiency in converting a wide range of feedstocks (Basu, 2010) and is one of the conventional thermochemical conversion methods (Akyüz et al., 2020). Gasification is a thermochemical process that converts biomass or waste into combustible gases at high temperatures (>650 °C) under controlled oxidising atmospheres. Producer gas primarily contains H_2, CO, CH_4, CO_2, and N_2 (Win et al., 2019; Xue et al., 2014). Biomass gasification

is the process of converting carbonaceous material like biomass in a partial oxidation environment in a series of chemical reactions, such as dehydration, devolatilization/pyrolysis, to produce syngas (a combustible fuel) which contains carbon monoxide, hydrogen, methane and hydrocarbons, etc. at a high temperature between 800-1000 °C (Ai et al., 2022). Gasification can be autothermal (direct) or allothermal (indirect). Autothermal gasifiers use part of the fuel for internal heat, while allothermal systems require an external heat source (Heidenreich & Foscolo, 2015).

Currently, various types of gasification reactors are available: fixed bed, entrained bed, and fluidised bed (Kordi & Seyyedi, 2021). Several factors influence the selection, including technology, heat load, types of materials, energy consumption, the environment, and economics. Although fixed beds, which include up-, down-, and cross-drafts, are the classical type of the gasifier, they are limited in their ability to use non-uniform fuel sizes and thermal load variations. Fluidised bed reactors, which can burn fine solid fuels with non-uniform sizes more rapidly, have been used recently in thermal industries, particularly in medium-sized power plants. Its fuel flexibility enables isothermal operating conditions and ensures highly efficient energy conversion with low tar (Buragohain et al., 2010; Sidek et al., 2020). A fluidised bed gasification reactor comprises a cylinder-like column filled with a bed material, typically inert sand, supported by a perforated plate with evenly distributed nozzles. Primarily, the sand bed is initially heated using an external heat source to facilitate the gasification process. Air, oxygen, and steam are utilised as oxidising agents, facilitating the fluidisation of bed materials and enabling the partial oxidation of biomass feedstock within a temperature range of 700-900 °C (Buragohain et al., 2010). The choice of gasifying agent affects syngas composition and heating value. Air is cost-effective and operationally simple, whereas oxygen–steam mixtures yield hydrogen-rich gas with higher LHV (Shayan et al., 2018).

Mahapatro and Mahanta (2020) examined low-grade coal, sawdust, and rice husk in a pressurised circulating fluidised bed gasifier. Increasing pressure improved CCE (54-97%), CGE, and gas yield. Ali et al. (2024) examined air gasification of municipal solid waste components (wood, paper, and cloth) in an atmospheric fluidised bed gasifier by varying equivalence ratio (0.2-0.5), temperature (800-950 °C), and $CaCO_3$ addition. Higher temperatures enhanced CO and H_2 production, LHV, and CCE, whereas increasing ER increased CO_2 yield and CCE while reducing syngas yield and LHV; $CaCO_3$ addition at 900 °C promoted syngas formation, suppressed CO_2, and achieved a maximum LHV of ~8000 kJ Nm^{-3} for wood gasification. Mohanty et al. (2025) studied pinewood air gasification in a lab-scale fluidised bed with dolomite, limestone, and activated charcoal. CCE and CGE increased with temperature (700-850 °C) and ER (0.20-0.35), reaching complete carbon conversion around ER ≈ 0.30–0.35. Tar content was reduced via activated carbon. Mallick et al. (2020a, 2020b) reported air-blown co-gasification of rice husk, sawdust, and bamboo

dust in a 50-kW circulating fluidised bed. ER (0.19–0.35) and temperature (800-900 °C) effects were evaluated. Higher ER increased gas yield, CCE, and CGE, while reducing CO and H_2 concentrations and LHV. Biomass blending enhanced CCE up to 98% and CGE to 62%. Cvetinović et al. (2024) reported that gasification is more favourable at low air-to-fuel and steam-to-fuel ratios, consistent with previous studies. Badu et al. (2025) studied a 200 kW semi-industrial air-blown fluidised bed. Increasing ER (0.22-0.45) improved bed temperature, gas yield, CCE (78.6-98.5%), CGE (51.8-71.4%), and overall efficiency (79-93%), while reducing syngas LHV (3.9-7.1 MJ/Nm^3) due to nitrogen dilution. Other works have demonstrated the importance of fluidised bed hydrodynamics, including air distributor and nozzle geometry, in improving particle motion, bed mixing, and pressure stability (Alashmawy et al., 2024; Raza & Ahsan, 2024; Tian et al., 2024). Yet, validation under actual gasification conditions remains limited.

Although substantial research has been conducted on fluidised bed gasification, several practical and operational challenges remain insufficiently addressed. One of the most critical issues is nozzle clogging and biomass agglomeration, which disrupt uniform air-biomass mixing, reduces effective mass and heat transfer, and ultimately degrades gasifier performance. These operational limitations are particularly pronounced in long-duration and pilot-scale operations. In Pakistan, very few studies have been reported on indigenously developed pilot-scale fluidised bed gasifiers. Most experimental investigations rely on imported or commercially purchased reactors, which significantly increase research costs and limit design adaptability to local biomass characteristics and operating conditions. Consequently, limited attention has been given to developing gasifier designs tailored to local resources, feedstock variability, and economic constraints. Furthermore, most of the existing literature focuses on laboratory-scale gasifiers, often investigating a single biomass type or a narrow operating window. Comparative gasification studies involving multiple biomass feedstocks under identical operating conditions remain scarce, particularly at the pilot scale.

These gaps highlight the need for low-cost, indigenously designed pilot-scale fluidised bed gasifiers that can ensure reliable operation, improved air distribution, and uniform heat transfer while accommodating a wide range of biomass feedstocks. Addressing these challenges provides the motivation for the present study, which aims to enhance reactor design and operational feasibility under realistic and locally relevant conditions.

This study investigates the air-gasification performance of five different biomass feedstocks (agricultural residues), such as sawdust (SD), wheat straw (WS), rice husk (RH), corn stalk (CS), and spent tea waste (STW), in an allothermal, indigenously designed, low-cost, and efficient pilot-scale fluidised bed gasifier. The laterally arranged nozzles are radially distributed over a perforated plate to prevent blockage by inert materials. Experiments were conducted over an equivalence ratio (ER) range of 0.20-0.30 within a

temperature range of 750-950°C, focussing on syngas composition, yield, lower heating value (LHV), higher heating value (HHV) of syngas, cold gas efficiency (CGE), carbon conversion efficiency (CCE), and overall reactor performance.

MATERIALS AND METHODS

Feedstocks Selection and Collection

The feedstocks utilised in this study included sawdust, wheat straw, rice husk, corn stalk and spent tea waste based on availability. Sawdust was sourced from various sawmills and carpenter shops, while crop residues, including wheat straw, rice husk, and corn stalk, were obtained from fields after harvesting. Spent tea waste was obtained from tea hotels and small tea shops, where it is generated in substantial quantities. All feedstocks were collected from Faisalabad due to its abundance of biomass potential resources, and many industrial activities in this region generate a ton of waste every year while processing the crops.

Pretreatment of Feedstocks

Pretreatment was done at the Agricultural Engineering Workshop, University of Agriculture, Faisalabad (UAF). Each ungrounded biomass feedstock was stored in an individual bag. Each biomass was then spread on an open field for two days under an open-air sun drying process. Spent tea waste was soaked to remove externally introduced beverage additives (milk and sugar) before drying, ensuring consistent and representative biomass behaviour during gasification while leaving the primary lignocellulosic structure largely unchanged and subsequently dried in a solar drier. After that, the wheat straw and corn stalks were chopped due to their larger particle size to achieve the required particle size suitable for grinding. All feedstocks were then ground using a grinder and then sieved to obtain a particle size between 0.2 mm and 1.5 mm, as shown in Figure 1. All biomass feedstocks were then stored at the Material Testing Laboratory storage facility under standard conditions, such as a dry and cool environment with controlled moisture levels and under well-ventilation to prevent degradation or contamination from microbial activities until laboratory-scale investigation and experimentation.

Properties of Feedstocks

Silica sand was utilised as the bed material in this study because it ensures uniform fluidisation, temperature, and heat-transfer effects in a gasifier, and it is very inexpensive. The mean particle sizes, apparent densities, bulk densities, and porosities of the silica sand and feedstocks are shown in Table 1.

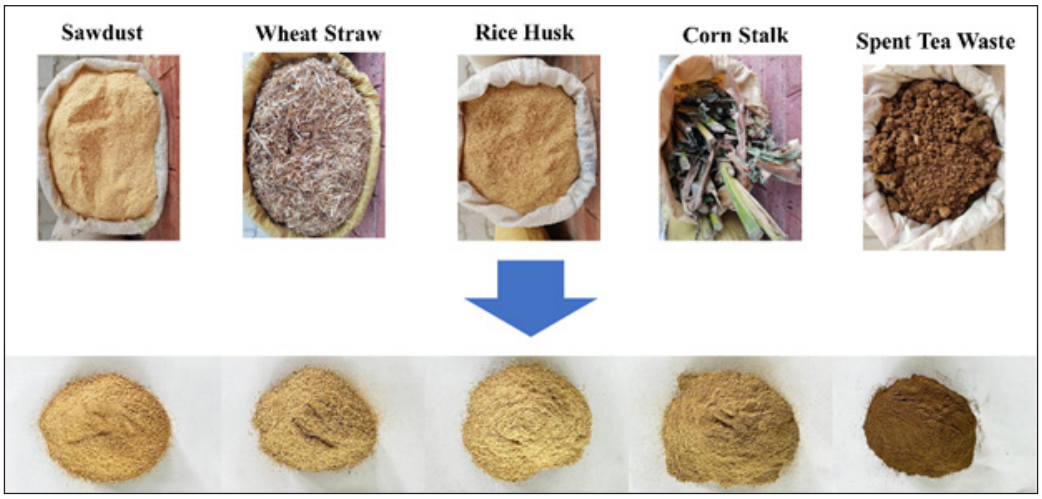


Figure 1. Pre-treatment of biomass feedstocks

Table 1
Properties of biomass feedstocks and bed material

Properties	Silica sand	Sawdust	Wheat Straw	Rice Husk	Corn Stalk	Spent Tea Waste
Mean particle size (μm)	384	467	434	576	678	386
Apparent density (kgm^{-3})	2630	1300	1350	1510	1330	1233
Bulk density (Kgm^{-3})	1602	259	188	264	233	273
Porosity	0.4	0.8	0.86	0.83	0.82	0.78

Characteristics of Feedstocks

Chemical compositions of these feedstocks were determined by the proximate and elemental analysis as presented in Table 2. Proximate analyses were performed in the Densification Laboratory, Department of Agricultural and Biological Engineering (ABE), the Pennsylvania State University, PA 16801, United States of America. Elemental analysis was done at the Iowa State University Chemical Instrumentation Facility (1240 Hach Hall), in the United States of America. All analyses were conducted before the combustion of the biomass feedstocks.

Experimental Setup

All experiments were conducted using an indigenised fluidised bed gasification reactor, which is situated at the solar park in front of the Material Testing Laboratory, Structures and Environmental Engineering, Agricultural Engineering Workshop, University of

Table 2
Proximate and elemental analysis of biomass feedstocks

Parameters	Sawdust	Wheat Straw	Rice Husk	Corn Stalk	Spent Tea Waste	Standards/References
Proximate						
Moisture contents (wt.%, adb)	9.65	6.08	6.85	6.97	6.77	(ASTM)-E-871-82
Volatile matter (wt.%, adb)	67.51	58.12	62.33	66.5	71.25	
Ash (wt.%, adb)	19.88	15.44	14.83	18.78	17.57	LAP
Fixed carbon (wt.%, adb)	2.96	20.36	15.99	7.75	4.41	(Alakangas, 2016)
Elemental						
C (wt.%, adb)	46.113	41.397	41.082	42.109	47.105	CHNS/0 Analyser
H (wt.%, adb)	5.695	5.512	5.674	5.387	6.51158	
N (wt.%, adb)	0.7125	0.65823	1.72104	0.972	3.94698	
S (wt.%, adb)	0	0	0	0	0.54299	
O (wt.%, adb)	47.480	52.433	51.526	51.533	41.894	(Bridgeman et al., 2010)
Higher heating value ($MJkg^{-1}$)	17.662	16.662	16.015	18.282	21.981	(Manual, 2008)
Lower heating value ($MJkg^{-1}$)	16.29	15.4	14.71	17.03	20.5	(Basu, 2010)

Agriculture, Faisalabad. A process diagram of an indigenised fluidised bed gasification reactor is shown in Figure 2.

A reactor was designed with a height of 1.524m and a diameter of 0.1524m, equipped with components like a blower, heater, plenum, distributor plate, nozzles, hopper, screw feeder, cyclone, and gas filter. All materials were locally available, and no complex components or high-cost systems were required, ensuring low capital expenditure. The distributor plate was equipped with specially designed 19 vertical nozzles to ensure effective air-fuel interaction within the reactor. Each nozzle contains two lateral openings positioned symmetrically on the left and right sides, along with a central bottom opening through which preheated air is introduced into the bed. Before each experiment, the inlet and outlet conditions of individual nozzles were carefully adjusted to verify even air distribution throughout the system. In addition, the syngas composition remained stable during prolonged steady-state operation, providing further evidence that air was evenly supplied to all nozzles throughout the gasification process. This configuration promotes uniform distribution of the gasifying agent, leading to improved mixing between the biomass fuel and air. As a result, stable reaction conditions are achieved while minimising the risks of bed clogging and particle agglomeration.

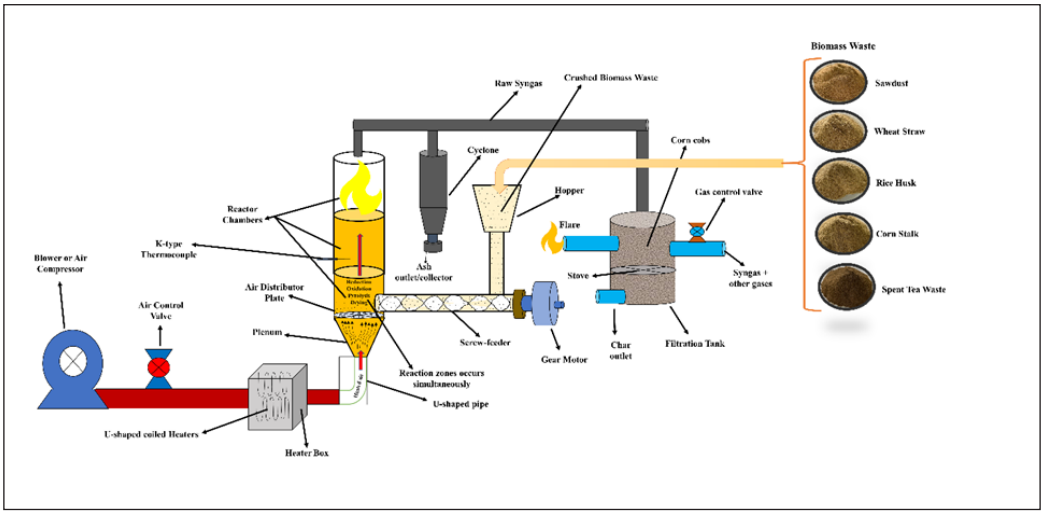


Figure 2. Process diagram of an indigenised fluidised bed gasification reactor

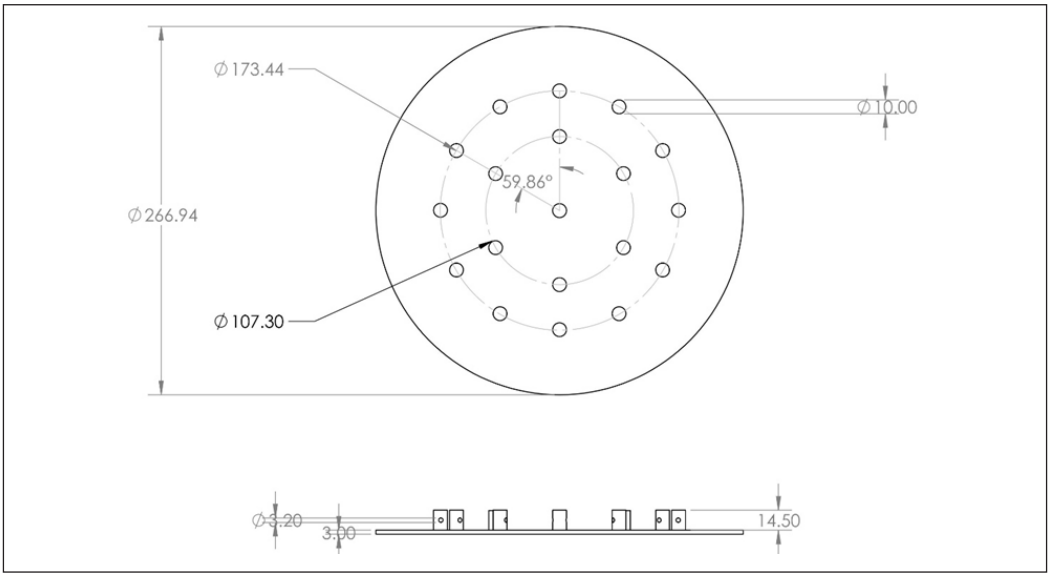


Figure 3. Distributor plate with vertical nozzles and dimensions

Furthermore, the nozzle geometry enhances the local air velocity, which supports homogeneous fluidisation and consistent reaction kinetics. The detailed arrangement and dimensions of the nozzles mounted on the distributor plate are illustrated in Figure 3.

The reactor was fabricated from mild steel. A complete photograph of the experimental setup, together with an AutoCAD drawing with labelled components, is shown in Figure 4. The dimensions of the setup are given in Table 3.

The experiments were conducted at three different ERs, 0.20, 0.25 and 0.30, with a constant feed rate of 20 kgh^{-1} to evaluate how different biomass types affect gasification performance. Based on operational stability considerations, the equivalence ratio (ER) was selected in the range of 0.2-0.3. Previous studies have shown that ER values below 0.2



Figure 4. Complete experimental setup of the fluidised bed gasifier with labelled AutoCAD drawing

Table 3
Design specifications and dimensions of the FBG Reactor

Description	Dimensions
Reactor Cylinder	Dia. 152.4 mm, Height 1524 mm
Cyclone	Length 609 mm, Dia. 101 mm
Gear Motor	90 hp, 70 rpm, 220 V
Hopper	Square in shape, H 254 mm, top L or W 279.4 mm, bottom L 127 mm W 114.3 mm Capacity 2 kg, 45°
Conveyor	Dia. 12.7 mm, Length 609 mm
Blower	3 hp, 2800 rpm
Air Control Valve	Manually operated, Opening Dia. 50.8 mm
Heater	4 heaters, 1500 Watt each, 220 V
Plenum	Upper Dia. 203 mm, Bottom Dia. 50.8 mm
Distributor Plate	Dia. 203 mm, Thickness 6.35 mm
Nozzles	19 Nozzles, Dia. 10 mm, Lateral holes Dia. 1.2 mm Bottom hole Dia. 3.2 mm
Frame	H 1824 mm, L 1676 mm
Filtration Cylinder	H 635 mm, Dia. 228 mm
Ash Collector	Dia. 50.8 mm

result in an insufficient oxygen supply, which can lead to unstable ignition and incomplete gasification. Conversely, increasing the ER beyond 0.3 introduces excess oxygen, promoting excessive oxidation and shifting the process toward combustion, which may deteriorate syngas quality, as reported in (Jamin et al., 2020a). Although some studies have employed ER values up to 0.35, such investigations typically utilise catalytic bed materials that enhance tar reforming and water-gas shift reactions, allowing stable operation at higher ER. Therefore, an ER range of 0.2-0.3 was chosen in the present study. At the start-up of the experiment, the gasifier was charged with 5 kg of silica sand used as bed material to provide a good heat transfer and uniform fluidisation with a velocity of fluidization at 0.176 ms^{-1}. A hopper consists of Sawdust, Wheat straw, Rice husk, Corn stalk and Spent tea waste, which were fed into the fluidised bed gasifier through a screw feeder, respectively, as a feedstock with the feeding rate of 20 kgh^{-1}. The reactor had a maximum rated thermal capacity of 114 kWh. The thermal capacity of the gasifier for each biomass was calculated individually using its lower heating value and a fixed feed rate, as summarised in Table 4, demonstrating that the system functions at a pilot scale.

Air being a gasifying agent coming from the blower or air compressor (2800 rpm) with air speed from 8-15 ms^{-1} measured from AM-4812 anemometer device for each biomass feedstocks was first heated by the four U-shaped coiled electric heaters enclosed in the box which is fitted on its way to increase the temperature of the air up to 400^{o}C in 20-30 minutes depends upon the speed of ambient air surrounding it and passes it through the nozzles holes from the plenum distributed on the distributor plate and it starts burning of the biomass at the temperature of 750-950^{o}C to carry out the gasification process. No nozzle blockage or airflow restriction was observed during 3–4 hours of uninterrupted reactor operation for any of the tested biomass feedstocks. Maintain the gasification process by monitoring the temperature. During steady-state operation, the reactor bed temperature was maintained within ±15^{o}C, while variations in hydrogen and carbon monoxide concentrations remained within ±5 vol%. Proper fluidisation was confirmed through theoretical analysis and experimental observation. In all experiments, the superficial gas velocity exceeded the minimum fluidisation velocity (U_{mf} = 0.089–0.152 ms^{-1}), resulting in stable bed operation.

Table 4
Thermal capacity of the gasifier for various biomass feedstocks

Biomass	Feed Rate (kg/h)	LHV (kJ/kg)	Thermal Capacity (kW)
Sawdust	20	16290	91
Wheat Straw	20	15400	86
Rice Husk	20	14710	82
Corn Stalk	20	17030	95
Spent Tea Waste	20	20500	114

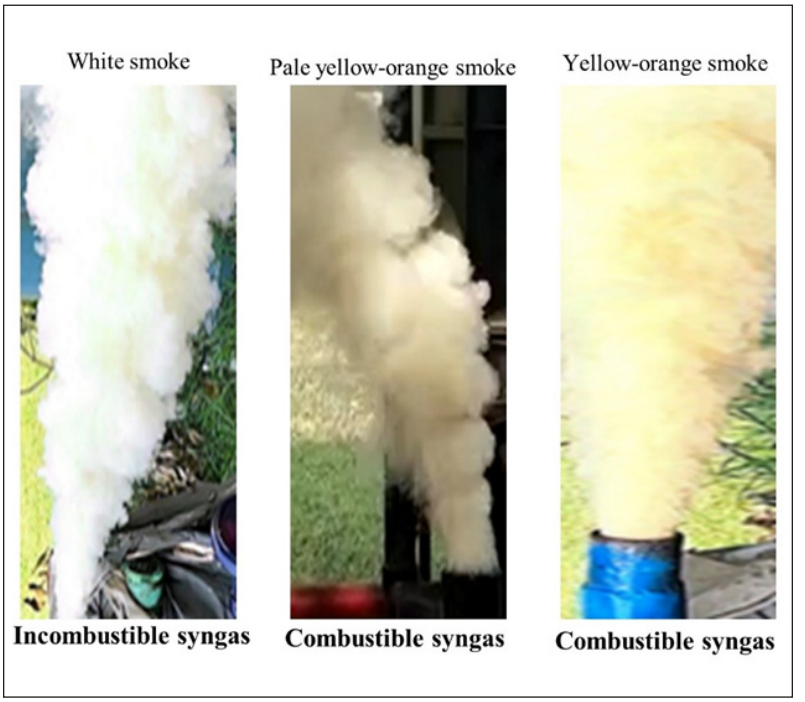


Figure 5. Comparison of syngas colour and combustibility

Visual observations showed uniform bed expansion and stable bubbling without channelling or stagnant zones. Consistent operating conditions ensured reliable fluidisation for all biomass feedstocks. A flame emerges from the flare stack stove, indicating the reactor is producing syngas. Ensure the reactor reaches targeted temperatures of (850-950°C, 750-850°C) and make sure syngas is neither too "tarry" nor too "sooty". Figure 5 shows a comparison of syngas combustibility based on the observed smoke colours.

Syngas quality was qualitatively evaluated based on flame stability, flame colour, and smoke emission characteristics during gasification experiments. Gasification (incomplete combustion of carbonaceous fuels) can be expressed using the sub-stoichiometric shows in Equation 1:

$$\text{Biomass} + \text{Limited Air} \rightarrow CO + H_2 + CH_4 + CO_2 + H_2O + N_2 \quad [1]$$

To collect the syngas, air sampling bags were used and flushed before opening the valve at the gas outlet of the system. Attach the air sampling bag to the outlet port of the hand pump using a coupling, pump the required amount of syngas into the bag, and seal it for gas chromatography analysis. Gas chromatography measurements had an estimated uncertainty of ±2–5%, with ±3% used for error propagation. Uncertainty in LHV, HHV,

Table 5
Chain of chemical reactions involved in the biomass gasification process

Reaction Type	Reaction Equation	Enthalpy	Reactions
Endothermic		ΔH (kJ/mol)	
Pyrolysis	Dry feedstock + heat → char + volatiles	--------	R1
Water-gas/ Steam Gasification	C [volatiles/char] + H_2O → CO + H_2	+131.4	R2
Boudouard/Reduction Reaction	C [volatiles/char] + CO_2 → 2CO	+172	R3
Water-Gas Shift (WGS) reaction	$CH_4 + H_2O \leftrightarrow 4CO + 3H_2$	+206	R4
Exothermic			
Oxidation Reaction	$C + O_2 \rightarrow CO_2$	-406	R5
	$2H_2 + O \rightarrow 2H_2O$	-242	R6
Methanation Gasification	$C + 2H_2 \leftrightarrow CH_4$	-75	R7
Water-Gas Shift (WGS) reaction	$CO + H_2O \leftrightarrow CO_2 + H_2$	-42.3	R8

CGE, and CCE was calculated using the root-sum-square method. Shut down the system by closing the gas out and air inlet valves. Finally, allow the system to cool down completely.

All the chemical reactions that occur inside the gasifier are discussed in Table 5. In the first zone, drying removes moisture from the biomass at 150-200 °C. This is followed by pyrolysis (R1) (200-500 °C), where biomass decomposes in an oxygen-deficient environment to produce char, tars, and volatile gases. In the oxidation zone (R5-R6) (800-1200 °C), partial combustion of char and volatiles with limited oxygen generates CO_2 and H_2O, releasing heat to sustain the endothermic gasification reactions. Finally, in the reduction zone (700-1000 °C), char reacts with H_2O and CO_2 through the Boudouard (R3) and water–gas reactions (R2), while the water-gas shift (R8) and methanation (R7) reactions further influence the syngas composition. As a result, the producer gas primarily consists of CO, H_2, CO_2, and CH_4.

Gasification Performance Metrics

Equivalence Ratio (ER)

The equivalence ratio (ER) is defined as the ratio of the actual oxygen-to-biomass ratio to the stoichiometric oxygen-to-biomass ratio required for complete combustion. This relationship is expressed in Equation 2, (Lv et al., 2003).

$$ER = \frac{\text{Weight of oxygen (air)/Weight of dry biomass}}{\text{Stoichiometric oxygen (air)/Biomass ratio}} \quad [2]$$

Lower Heating Value of Syngas

(LHV_{syngas}) refers to the chemical energy in MJ contained per unit of volume of the syngas, which is calculated according to the combustible components present in the syngas by using Equation 3, (Guo et al., 2014; Wan et al., 2013).

$$LHV_{syngas} = \frac{[12.622(\%H_2) + 10.788(\%CO) + 35.814(\%CH_4)]}{100} (MJNm^{-3}) \quad [3]$$

Higher Heating Value of Syngas

High heating value of syngas (HHV_{syngas}) is the total amount of heat energy released when a unit mass of syngas is completely burned. HHV_{syngas} can be calculated by the below-mentioned Equation 4, (Wan et al., 2013).

$$HHV_{syngas} = \frac{[(\varkappa_1 \cdot HHV)_{CO} + (\varkappa_3 . HHV)_{H_2} + (\varkappa_4 \cdot HHV)_{CH_4}]}{100}$$

OR [4]

$$HHV_{syngas} = (H_2\% \times 30.52 + CO\% \times 30.18 + CH_4\% \times 95) \times 4.19$$

where, '$\varkappa_1, \varkappa_3, \varkappa_4$' represent the volume fractions or percentages of CO, H_2, and CH_4, respectively. The higher heating values (HHV) of CO, H_2 and CH_4 are 12.71 $MJNm^{-3}$, 12.78 $MJNm^{-3}$, and 39.76 $MJNm^{-3}$, respectively.

Syngas Yield (Y_{syngas})

Syngas yield quantifies the volume or mass of produced syngas relative to the biomass feedstock input. It provided insight into the effectiveness of gasification in generating usable gas. The overall dry gas volume generated can be determined using Equation 5 (Guo et al., 2014; Wang et al., 2015), based on the nitrogen concentration in the product gas and the overall nitrogen introduced into the reactor through the air supply during gasification.

$$Y_{syngas} = \frac{Q_a \times \chi_{N_2}}{\chi'_{N_2}} \quad [5]$$

χ_{N_2} = represents the percentage fraction of N_2 in air under standard conditions, 78.12%

χ'_{N_2} = the volume percentage fraction of N_2 in the product gas at standard conditions, vol.%

Q_a = refers to the volume of air supplied per kilogramme of biomass, Nm^3kg^{-1}

[Volumetric air flow rate (m^3h^{-1}) ÷ Feed Consumption rate (kgh^{-1})

Carbon Conversion Efficiency (CCE)

The carbon conversion efficiency (CCE) refers to the proportion of carbon from the biomass that is transformed into carbon-based gaseous compounds relative to the total carbon content in the biomass. It was calculated as the ratio of the carbon contents in the syngas to the initial amount of carbon contents in the biomass feedstock, expressed as a percentage. Consequently, the carbon conversion efficiency can be determined using Equation 6 (Makwana et al., 2015; Sharma et al., 2011; Wang et al., 2015).

$$CCE\ (\%) = \frac{12(CO\% + CH_4\% + CO_2\% + nC_nH_m\%)}{22.4 \times \%C_{feedstock}} \times Y_{syngas} \times 100\% \quad [6]$$

in which,

CCE = Carbon Conversion Efficiency (%)

Y_{syngas} = Yield of dry syngas per kilogramme of dry biomass feedstock (Nm^3kg^{-1})

$\%C_{feedstock}$ = Carbon percentage in the dry biomass feedstock, determined through elemental analysis and CO_2, CO, CH_4, C_nH_m are the volumetric percentages present in the syngas

Cold Gas Efficiency (CGE)

Cold Gas Efficiency (CGE), expressed as a percentage, represents the amount of chemical energy present in the product gas relative to the original energy stored in the solid fuel. This efficiency metric is calculated based on the Lower Heating Value (LHV_f) of the feedstock, which accounts for the energy released during the combustion process, excluding the latent heat of water vapour produced in the reaction. Cold gas efficiency was calculated by Equation 7 (Serrano et al., 2016).

$$CGE\ (\%) = \frac{100 \times V_{syngas} \times LHV_{syngas}}{LHV_f} \quad [7]$$

where 'V_{syngas}' represents the specific gas yield (Nm^3 gas per kg of feedstock) and is determined using Equation 8:

$$V_{syngas} = \frac{m_{syngas}}{\rho_{syngas}} \quad [8]$$

where m_{syngas}, represents the specific mass of the syngas (kg gas kg^{-1} feedstock) and ρ_{syngas} is the density of the syngas (kg gas Nm^{-3} gas). The values of the m_{syngas} and ρ_{syngas} was

obtained at Normal Temperature and Pressure (NTP) using Equation 9 and 10, as follows (Susastriawan et al., 2019b).

$$m_{syngas} = \frac{MW_{syngas}}{n} \quad [9]$$

$$\rho_{syngas} = \frac{[(\varkappa_1 \cdot \rho)_{CO} + (\varkappa_2 \cdot \rho)_{CO_2} + (\varkappa_3 \cdot \rho)_{H_2} + (\varkappa_4 \cdot \rho)_{CH_4} + (\varkappa_5 \cdot \rho)_{N_2}]}{100} \quad [10]$$

At Normal Temperature and Pressure (NTP), the densities of the compounds are; ρ_{CO} = 1.165 kgNm^{-3}, ρ_{H_2}= 0.089 kgNm^{-3}, ρ_{CH_4}= 0.668 kgNm^{-3}, ρ_{CO_2}= 1.842 kgNm^{-3}, and ρ_{N_2}= 1.165 kgNm^{-3}

Where, '$\varkappa_1$ to $\varkappa_5$' are volume fractions or percentage of CO, CO_2, H_2, CH_4, and N_2, respectively.

Meanwhile, the volume fractions of CO, CO_2, H_2, CH_4, N_2 in the syngas, along with the global biomass gasification reaction and mass balance, the molar quantity of the feedstock (n) and the molecular weight of syngas (MW_{syngas}) can be determined through Equation 11 and 12.

$$n(CH_xO_yN_z) + m\varphi(O_2 + 3.76N_2) \rightarrow \varkappa_1 CO + \varkappa_2 CO_2 + \varkappa_3 H_2 + \varkappa_4 CH_4 + \varkappa_5 N_2 \quad [11]$$

$$MW_{syngas} = \frac{[(\varkappa_1 \cdot MW)_{CO} + (\varkappa_2 \cdot MW)_{CO_2} + (\varkappa_3 . MW)_{H_2} + (\varkappa_4 \cdot MW)_{CH_4} + (\varkappa_5 \cdot MW)_{N_2}]}{100} \quad [12]$$

where 'n' is molar number (moles) of the feedstock, which is expressed as the mass of the feedstock in kg to the molar mass of the feedstock in kgmol^{-1}, 'ϕ' is the equivalence ratio.

RESULTS AND DISCUSSION

Effect of Volumetric Air Flow Rate on the Equivalence Ratio

The equivalence ratio (ER) is defined as the mass ratio of the gasifying agent (air) to the biomass feedstock in a combustion system. In gasification, both the volumetric and mass air flow rates are crucial in determining the ER, as they directly influence the air-to-fuel mixture and combustion characteristics across various biomass types.

Figure 6 illustrates the relationship between the equivalence ratio (ER) and both volumetric air flow rate (m^3h^{-1}) and mass air flow rate (kgh^{-1}) in a gasification system. Both mass and volumetric air flow rate show a consistent increasing trend, ranging from 18-28 kgh^{-1} and 18-24 m^3h^{-1}, respectively, as ER increases from 0.20-0.30. This shows that a higher ER results in greater air supply. These ERs were calculated based on the biomass chemical composition in the methodology section.

Equivalence ratio directly depends on the air flow rate. In most of the studies, ER is maintained by adjusting the air and fuel supply. In this study, variation in air supply while holding other operating variables constant and adjusting it according to ER (0.2,0.25 and 0.3). This approach is widely accepted for fluidised bed gasifiers and medium-to-large-scale fixed bed gasifiers. ER is regulated by adjusting the air flow rate, where a higher ER requires an increased air supply, leading to an enhanced oxidation rate in the gasifier, as observed in previous studies (Upadhyay et al., 2019). At higher air volume per unit time by offering more oxygen into the reactor, thereby promoting complete combustion, which may result in an increase in syngas production. Lower ER makes less oxygen available for the gasification reaction to complete, thus affecting the gas composition with either value of ER (Esfahani et al., 2012). On the other hand, a lower volumetric air flow rate will result in lower ER. There may be incomplete gasification leading to lower syngas yield.

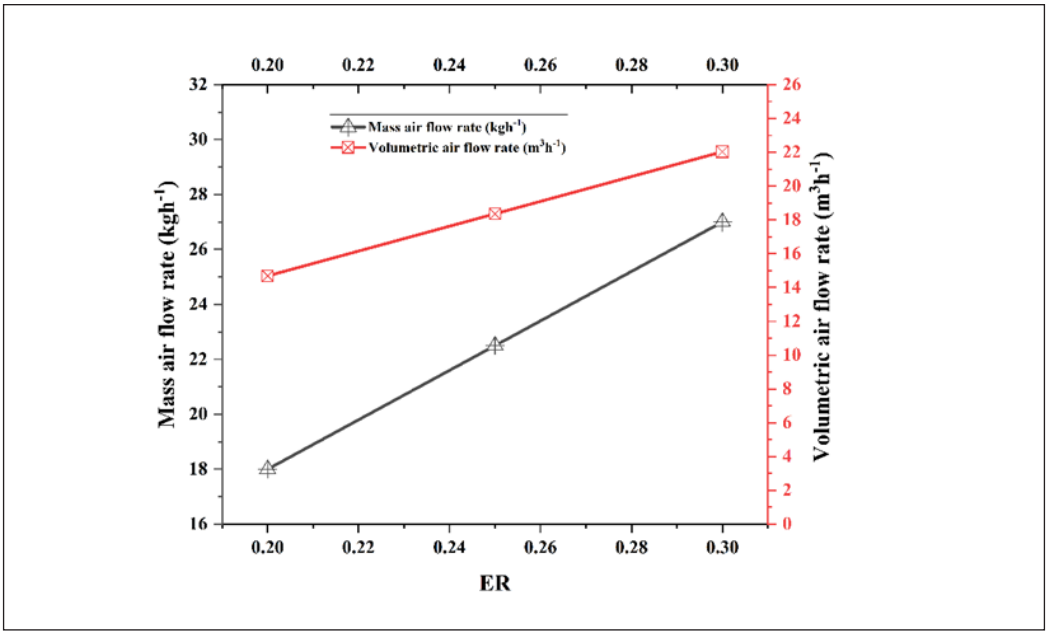


Figure 6. The relation between the mass air flow rate and the volumetric air flow rate with ER

Effect of Equivalence Ratio on the Syngas Compositions (A Combustible Gases)

The effect of the equivalent ratio on syngas composition was studied on various biomass feedstocks. Figure 7 (a-e) present the results of syngas compositions from gasification of sawdust, wheat straw, rice husk, corn stalk and spent tea waste by varying ER (0.20,0.25, and 0.30) at the gasification temperature ranging between 750 °C and 950 °C. Observed trends showed that the contents of CO_2, H_2, and CH_4 in the syngas composition decrease by increasing the equivalence ratio from 0.20 to 0.30. It is seen that the CO contents in the syngas in the gasification of sawdust were higher, of 27.56% at an equivalence ratio of 0.2, than those from the gasification of corn stalk, wheat straw, spent tea waste and rice husk, as 23.89, 23.45, 23.12 and 21.45%, respectively. For H_2 contents, it is higher in the gasification of spent tea waste as 21.45% at an equivalence ratio of 0.2 in comparison to 21.22, 20.48, 20.47 and 18.76% for sawdust, wheat straw, corn stalk and rice husk, respectively. The contents of CH_4 from gasification of the spent tea waste and wheat straw were reported as the highest at an equivalence ratio of 0.2, which is 3.39 and 3.16 %, respectively, as compared to the gasification of other biomass feedstocks such as corn stalk, sawdust and rice husk, at 2.91, 2.56 and 2.42 %, respectively.

Syngas compositions highly depend on the equivalence ratio. The percentage composition of syngas (CO, H_2, CH_4) depends on thermochemical reactions that take place inside the reactor. These chemical reactions are highly affected by the variation in design and operational parameters such as reactor design and size, feedstock type, moisture and ash contents present in the feedstock, equivalence ratio, etc. Present study revealed decreasing trends of carbon monoxide and hydrogen with the increasing ER means more oxygen fed into the fluidised bed gasification reactor leading to enhanced oxidation of hydrogen to form water resulting lower hydrogen production and carbon monoxide to form carbon dioxide thereby reducing level of carbon dioxide in the syngas composition i.e., ($H_2 + ½O_2 \leftrightarrow H_2O$) and ($CO + ½O_2 \leftrightarrow CO_2$) (Basu, 2010). Similar trends have been seen in the previous study (Untoria et al., 2024). The rate of carbon monoxide production in the water-gas reaction is higher than that of hydrogen and methane. Reaction rate sequences are Water Gas Shift > Bouduard > > Methanation reaction (Sikarwar et al., 2016). Therefore, it is reasonable that the CO has the highest percentage fraction in the syngas for all five biomass feedstocks at the observed equivalence ratio. In the case of methane, its composition increases at 0.2 equivalence ratio, but methane gas composition decreases when the equivalence ratio is increased beyond 0.24. This happens due to the oxidation of methane, in which methane is reacted with oxygen coming from the air through a blower, converting into carbon dioxide and steam ($CH_4 + 2O_2 \leftrightarrow CO_2 + 2H_2O$) (Basu, 2010). Methanation reaction is the slowest, and methane production is the lowest in the syngas from the gasification of all biomass feedstocks, i.e., less than 4%, similar as reported in earlier studies (Susastriawan et al., 2019a). A higher value of equivalence ratio represents more methane burns, thus reducing

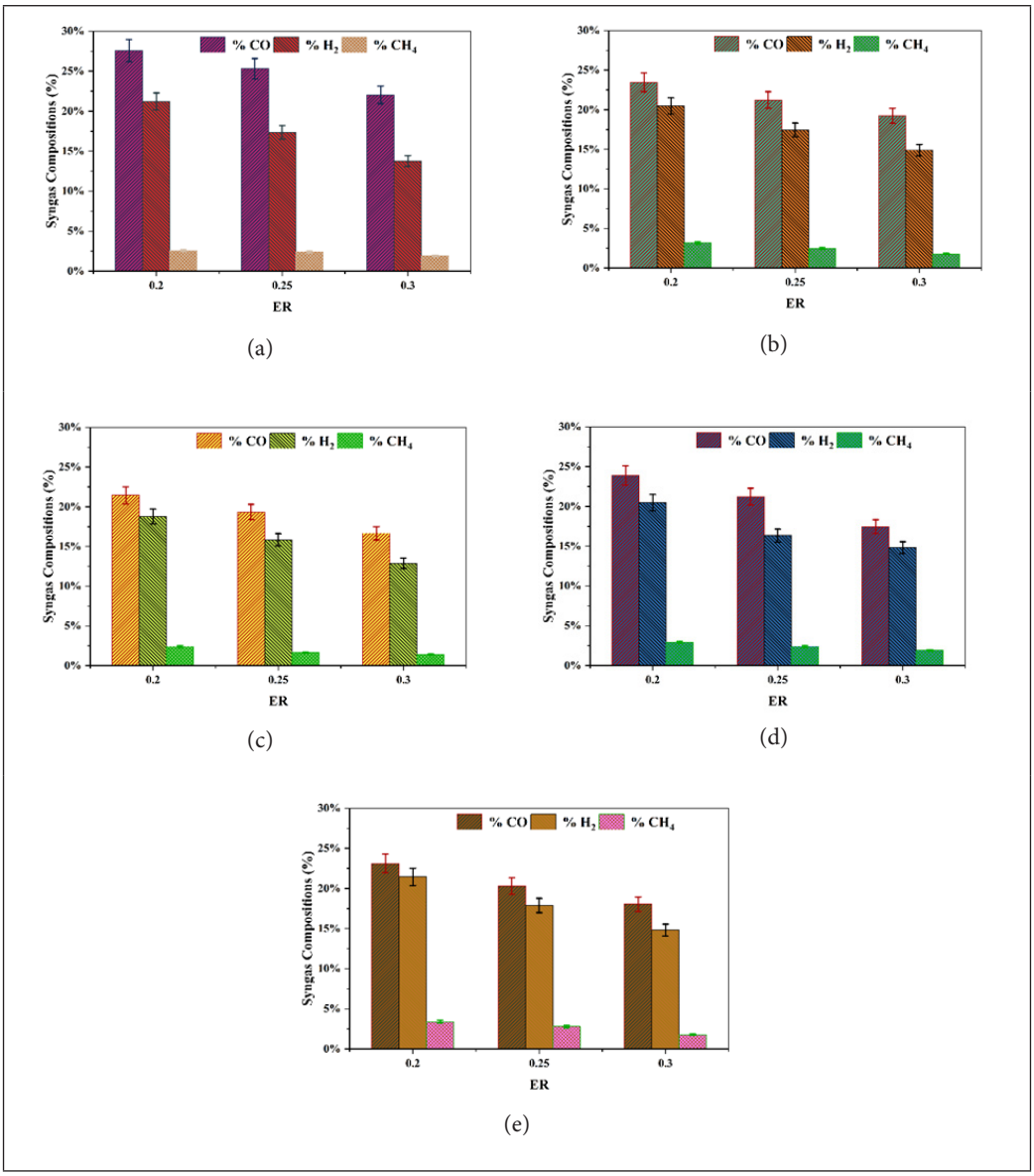


Figure 7. Syngas compositions comparison among different biomass feedstocks at three different ER; (a) Sawdust; (b) Wheat Straw; (c) Rice Husk; (d) Corn Stalk; (e) Spent Tea Waste

methane production (Guo et al., 2014). Jamin et al. (2020a) investigated the influence of equivalence ratio between 0.21 and 0.37 on raw and torrefied wood waste in a fluidised bed gasifier and concluded that with the increasing equivalence ratio, the contents of carbon monoxide, hydrogen and methane decreases with the increasing air supply, it matches with the present trends on CO, H_2 and CH_4 compositional analysis in the syngas. In the case of

biomass feedstocks, spent tea waste produces a higher amount of hydrogen gas as compared to other feedstocks due to more hydrogen being present in it. Higher carbon monoxide is produced in the case of sawdust due to the higher carbon content in it. The absence of a peak in syngas composition indicates stable gasification behaviour and efficient biomass conversion throughout the investigated ER range.

Effect of ER on HHV, LHV of Syngas, and Syngas Yield

Syngas yield mainly depends on the equivalence ratio, i.e., fuel and air feeding rate. Syngas yield was determined using Equation 5. It was observed from Figure 8 (a-e) that with the increase in ER, syngas yield increases. Syngas yield was noticed at three equivalence ratios (0.20, 0.25, 0.30). The highest syngas yields were observed of 2.31, 2.23, and 2.23 Nm^3kg^{-1} at ER of 0.3 in spent tea waste, sawdust and corn stalk, separately, in comparison with other feedstocks, i.e., rice husk and wheat straw, which were calculated as 2.16 and 2.10 Nm^3kg^{-1}, respectively. Conversely, lower and higher heating values of syngas primarily depend on the concentration of combustible gases, including carbon monoxide, hydrogen and methane present in it. These heating values were calculated by using Equation 3 and 4. It can be seen from Figure 8 (a-e) that with the increase in the ER from 0.2 to 0.3, the LHV and HHV of syngas decreased in all five biomass feedstocks. Highest LHV and HHV's of syngas were noticed in sawdust feedstock as 6.57 and 7.23 $MJNm^{-3}$, respectively at ER of 0.2 followed by spent tea waste, wheat straw, corn stalk, and then rice husk with the order of 6.42>6.25>6.20>5.55 $MJNm^{-3}$ for lower heating of syngas and 7.03>6.85>6.81>6.09 $MJNm^{-3}$ for higher heating value of syngas.

Syngas yield trends were discussed in the results section, which demonstrated that by increasing ER results, more air is injected into the reactor, resulting in an increase in the syngas production. The equivalence ratio plays a key role in biomass gasification (Sikarwar et al., 2017). The oxidation rates are enhanced by optimising the equivalence ratio. Increase in ER favouring high syngas yield due to increased combustion by producing more heat, resulting in more combustible gases involved in combustion reaction (Liu et al., 2018). As the gas heating values were evaluated based on the concentration of gases, first LHV increases at a higher rate due to stronger combustion reactions in the gasifier, which leads to a higher temperature that results in more generation of H_2 and CO by pyrolysis of biomass and tar decomposition. However, high ER results in the decrement of combustible gases and increment of non-combustible gases such as CO_2 and N_2, which decreases the lower heating value of the syngas as discussed by (Guo et al., 2014; Hendriyana, 2020; Mohammed et al., 2011; Upadhyay et al., 2019). Similar trends were observed in the case of higher heating value of syngas, which shows that at an ER of 0.2, sawdust found highest lower heating value of syngas, which is consistent with the findings of (Susastriawan et al., 2019a).

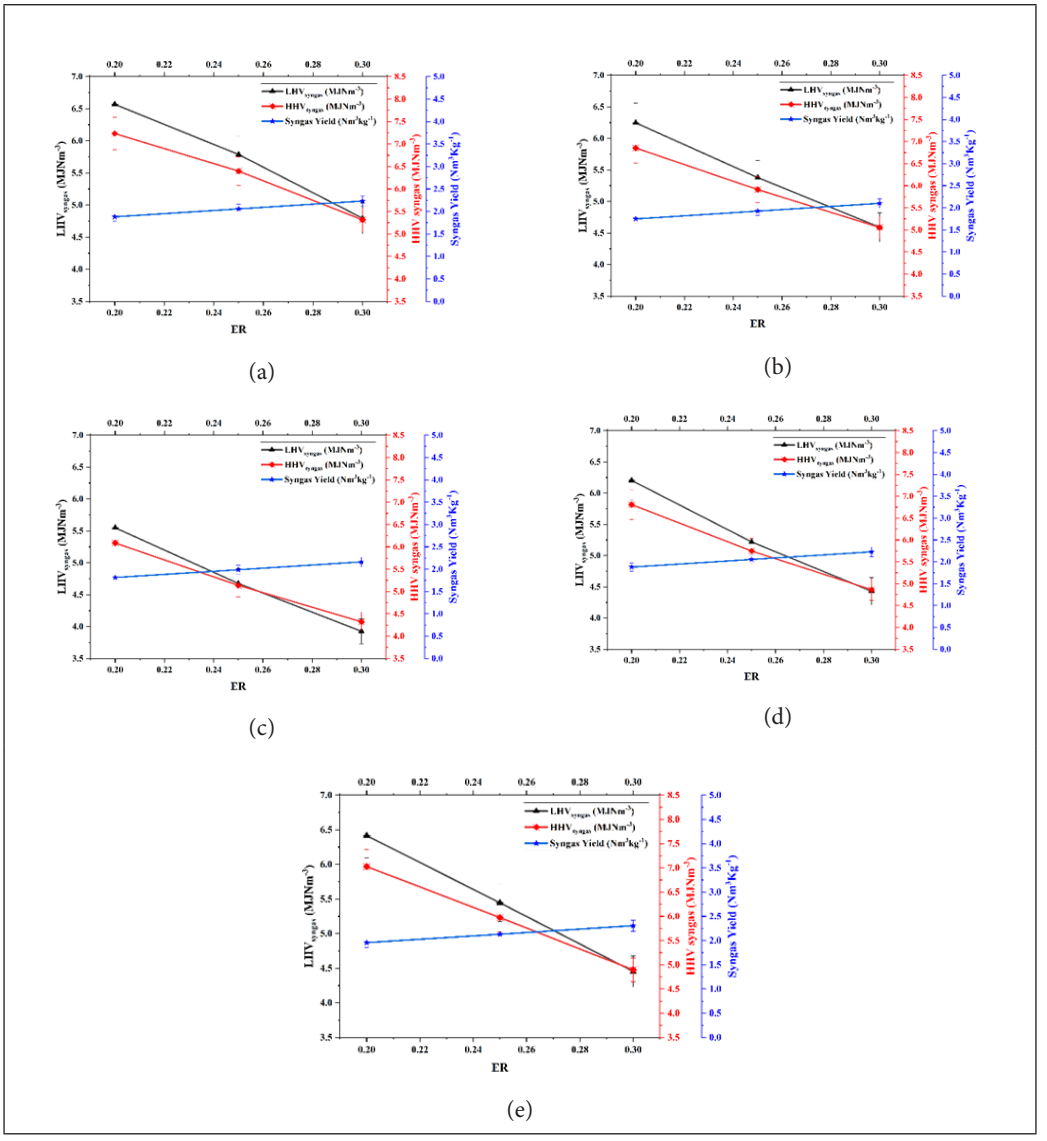


Figure 8. Comparison of LHV and HHV of syngas and Syngas yield among different biomass feedstocks at three different ER; (a) Sawdust; (b) Wheat Straw; (c) Rice Husk; (d) Corn Stalk; (e) Spent Tea Waste

Maximum HHV of syngas obtained at ER 0.2 indicates that the highest carbon monoxide and hydrogen production occurs at lower ER. By increasing the value of ER beyond 0.2, the HHV of syngas declines due to the higher concentration of nitrogen gas from the air flow rate. Similar trends were observed in previous studies (Jamin et al., 2020a). This contributes to lowering the energy content of the syngas by supplying more nitrogen gas, which dilutes the combustible gases (Lahijani & Zainal, 2011).

Effect of ER on Cold Gas Efficiency (CGE) and Carbon Conversion Efficiency (CCE)

It can be seen from several studies that ER plays a vital role in the determination of gasification performance. Figure 9 (a-e) illustrate the cold gas efficiency and carbon conversion efficiency values at various ER levels. These efficiencies were calculated using Equation 6 and 7, respectively. At ER 0.2, cold gas efficiency increased rapidly and approached a high level of about 75.84% in the case of sawdust in comparison with other biomass feedstocks, which is noted as maximum at 70.96%, 68.45%, 68.39% and 61.22% in the case of wheat straw, corn stalk, rice husk and spent tea waste. And began to decrease with further increase in ER. Meanwhile, with the increase in ER from 0.2 to 0.25,

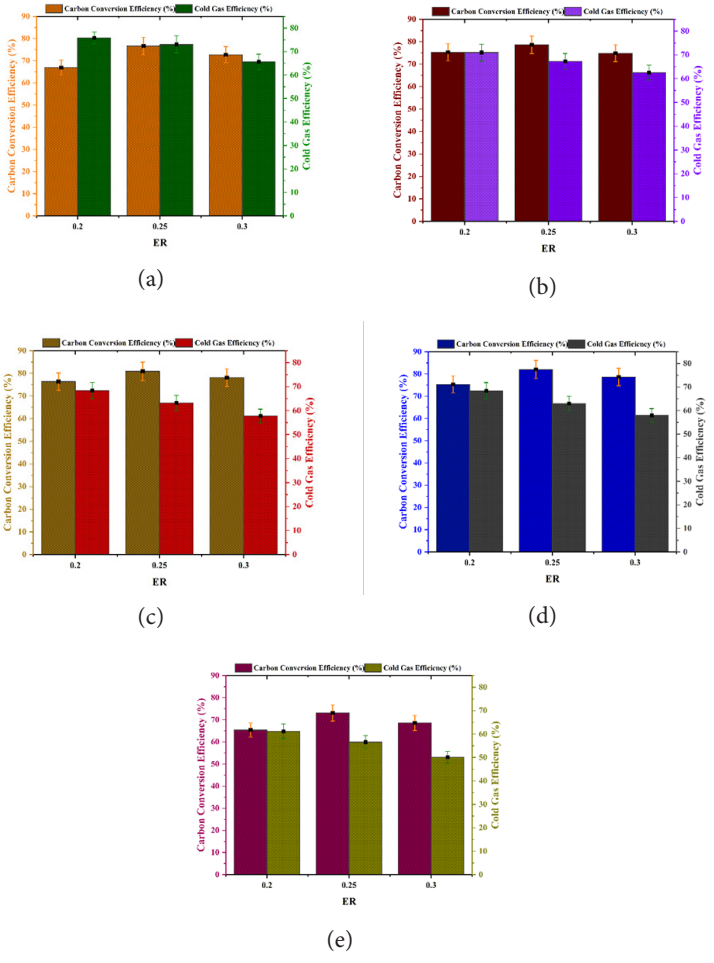


Figure 9. Comparison of Cold gas efficiency and Carbon conversion efficiency among various biomass feedstocks at three different ER; (a) Sawdust; (b) Wheat Straw; (c) Rice Husk; (d) Corn Stalk; (e) Spent Tea Waste

carbon conversion efficiency increases in each type of biomass feedstock, and the highest value is measured at ER of 0.25 in the case of corn stalk, which is 82.0%. All other biomass feedstocks followed the same trend as 81.0, 78.6, 76.6 and 73.1 % in rice husk, wheat straw, sawdust and spent tea waste, respectively. As ER continues to increase, carbon conversion efficiency declines.

To examine the effect of ER on energy and mass conversion, variations in cold gas efficiency and carbon conversion efficiency were analysed. An increase in the equivalence ratio (ER) promotes oxidation reactions, resulting in reduced CO and H_2 concentrations and, consequently, a decrease in the heating value of the syngas. As a result, the cold gas efficiency started to decline when ER exceeded 0.2, indicating a reduction in the overall energy conversion of biomass (Hendriyana, 2020; Karmakar et al., 2013; Zheng et al., 2016). Since the CGE of the fluidised bed gasification reactor fed by sawdust, wheat straw, rice husk, corn stalk and spent tea waste is still higher than 60 % at ER 0.2, it can be concluded that the FBG reactor is ideal for sawdust, wheat straw, rice husk, corn stalk and spent tea waste. In the earlier study, Yoon et al. (2012) reported a CGE ranging from 50% to 70% for rice husk gasification, while Wander et al. (2004) attained a CGE of 67.67% for sawdust gasification, which falls within the range of the present study. In contrast, carbon conversion efficiency exhibits a slight decrease beyond a certain equivalence ratio. At ER = 0.2, the limited oxygen supply favours reactions such as the Boudouard reaction (R3) and the water–gas reaction (R2), resulting in higher CO and H_2 concentrations and a correspondingly higher heating value of the syngas. However, the limited char oxidation at this ER leads to slightly lower carbon conversion efficiency, which may cause incomplete carbon conversion. At ER = 0.25, the oxygen supply is balanced, providing stable ignition and sufficient char oxidation, thereby increasing the carbon conversion efficiency and improving overall gasification performance. At ER = 0.3, a slight decrease in carbon conversion efficiency is observed, which can be attributed to the increased air flow rate. The higher flow raises the superficial gas velocity and reduces the residence time available for char–gas reactions, resulting in a marginal decline in carbon conversion efficiency at higher oxygen levels, as reported in the literature (Gurusamy, 2016; Khezri et al., 2016; Maitlo et al., 2022).

Wang et al. (2025) reported cold gas efficiency values exceeding 60% under optimised gasification conditions. Similarly, Susastriawan and Saptoadi (2023) observed peak cold gas efficiency in a downdraft gasifier within an equivalence ratio range of 0.20-0.25, with the maximum value occurring at an equivalence ratio of 0.20. In addition, system-level reviews by (Rahmah & Nurhilal, 2025; Wang et al., 2025) indicate that optimal gasification performance, characterised by high cold gas efficiency and carbon conversion efficiency, is generally achieved at equivalence ratios in the low-to-mid 0.2 range for biomass gasifiers.

These findings confirm that the performance metrics obtained in the present study fall within the typical range reported for optimised biomass gasification systems in recent literature.

Temperature Profile of the Gasification Process

Figure 10 illustrates the average temperature profile with respect to time at different ER values of 0.2, 0.25, and 0.3. For all biomass wastes, the reactor temperature rises rapidly during the initial heating stage and reaches a quasi-steady state after 40-50 minutes under external heating conditions, confirming stable operation. An increase in ER leads to higher peak temperatures, with average steady-state values of approximately 850 °C, 940 °C, and 1000 °C for ER values of 0.2, 0.25, and 0.3, respectively. The observed syngas composition trends reflect the effect of ER, with temperature variations being a naturally coupled outcome in air-blown fluidised bed gasification. This behaviour is attributed to the increased availability of oxygen at higher ER, which enhances exothermic oxidation reactions and internal heat generation, a trend generally consistent with previous studies (Jamin et al., 2020b; Sasujit et al., 2022; Valin et al., 2020). At longer operating times, a slight temperature decline is observed, likely due to heat losses and gradual changes in

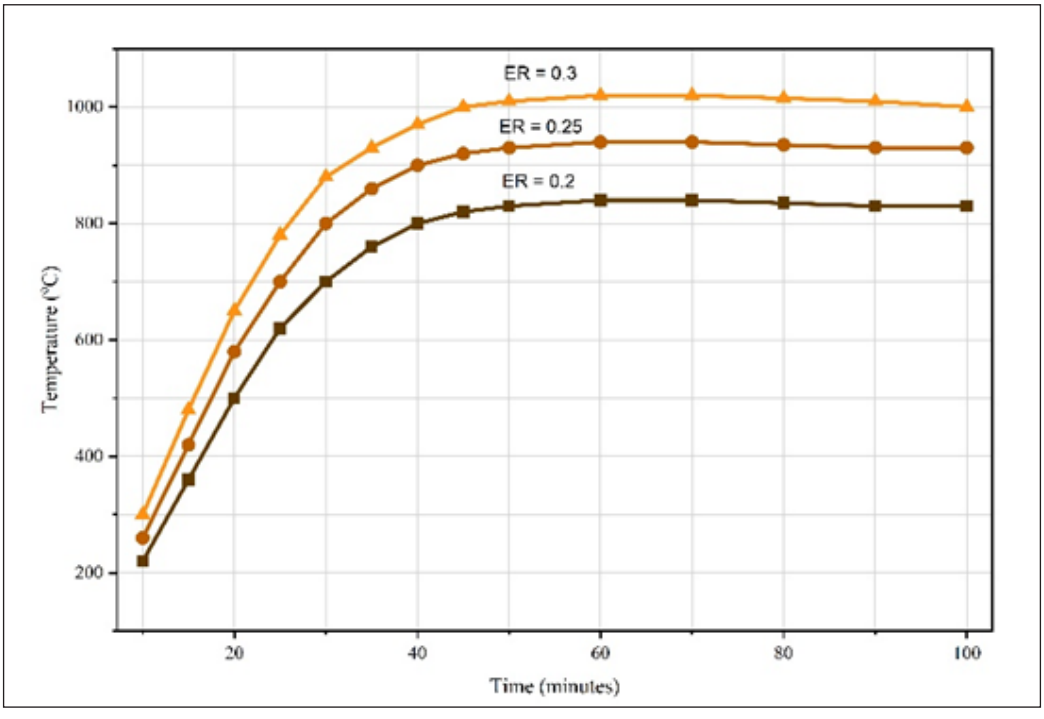


Figure 10. Average temperature profile inside the gasifier at three different ER

char reactivity. Overall, the findings of this study demonstrate good thermal behaviour of the system, consistent with the expected temperature profile during gasification.

CONCLUSION

The gasification performance of five biomass feedstocks was systematically investigated under varying equivalence ratios (ERs) with emphasis on syngas composition, heating value, syngas yield, cold gas efficiency (CGE), and carbon conversion efficiency (CCE). Among the tested feedstocks, sawdust produced the highest CO concentration of 27.56% at an optimum ER of 0.2, while the maximum H_2 content of 21.45% was obtained from spent tea waste. The highest lower and higher heating values of syngas, 6.57 and 7.23 MJ Nm^{-3}, respectively, were also observed for sawdust at an ER of 0.2. Spent tea waste yielded the maximum syngas production of 2.31 Nm^3kg^{-1} at an ER of 0.3. The highest CGE of 75.84% was achieved with sawdust at an ER of 0.2, and CGE values above 60% for all feedstocks at this ER indicate the operational feasibility of the reactor. Among the evaluated biomasses, corn stalk exhibited the highest carbon conversion efficiency of 82.0% at an ER of 0.25, demonstrating its strong gasification potential.

In conclusion, the indigenised pilot-scale fluidised bed gasifier demonstrated stable operation and efficient gasification across various biomass feedstocks. These findings highlight the potential of locally designed fluidised bed gasifiers for sustainable heat and power generation and provide a foundation for further scale-up and optimisation studies.

Future studies should focus on the effects of particle size distribution and blended solid fuels on gasification behaviour. Furthermore, detailed techno-economic analysis (TEA) and mass-energy balance modelling are essential for multiscale system optimisation and for facilitating demand-driven implementation of the proposed technology.

ACKNOWLEDGEMENT

The authors sincerely acknowledged the Department of Agricultural and Biological Engineering at Penn State University, 16802 Pennsylvania, USA, and Higher Education Commission (HEC) Pakistan for providing scholarship and fellowship support through the International Research Support Initiative Programme (IRSIP), which facilitated lab-scale analysis (Grant No. 1-8/HEC/HRD/2022/11973). They also extend their gratitude to the National Cleaner Production Center (NCPC), Rawalpindi, Pakistan, for funding the development of the reactor (Grant No. NCPC/Cb/F-39-2/17122020).

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTION

Shahbaz Hussain: formulation, investigation, methodology, visualisation, writing, original draft preparation. Prof. Dr Abdul Nasir: formal analysis, methodology, validation, and reviewing. Dr Ch. Arslan: formal analysis, reviewing. Dr. Shakeel Ahmad Anjum: formal analysis. Dr. Sibel Irmak: formal analysis-reviewing.

ABBREVIATIONS

FBG	:	Fluidised bed sasifier
SD	:	Sawdust
WS	:	Wheat straw
RH	:	Rice husk
CS	:	Corn stalk
STW	:	Spent tea waste
RS	:	Rice straw
CCE	:	Carbon conversion efficiency
CGE	:	Cold gas efficiency
LHV	:	Lower heating value
NO_x	:	Nitrogen oxides
SO_x	:	Sulfur oxides
CO_2	:	Carbon dioxide
CH_4	:	Methane
CO	:	Carbon monoxide
N_2	:	Nitrogen
H_2	:	Hydrogen
WtE	:	Waste-to-Energy
MSW	:	Municipal solid waste
ER	:	Equivalence ratio
UAF	:	University of agriculture, faisalabad
ABE	:	Agricultural and biological engineering
PA	:	Pennsylvania
ASTM	:	American society for testing and materials
LAP	:	Laboratory analytical procedure
WGS	:	Water-gas shift
NTP	:	Normal temperature and pressure

TEA : Techno-economic analysis
HEC : Higher education commission
NCPC : National cleaner production centre

REFERENCES

Ai, N., Chen, L., & Fu, Y. (2022). A novel analysis on pyrolysis and gasification process of rice straw feedstock. *Sustainable Energy Technologies and Assessments, 51*, Article 101866. https://doi.org/10.1016/j.seta.2021.101866

Akyüz, A., Akyurek, Z., Naz, M., Sulaiman, S., & Gungor, A. (2020). Hydrogen conversion using gasification of tea factory wastes. *Journal of the Serbian Chemical Society, 85*(7), 967-977. https://doi.org/10.2298/JSC190215013A

Alakangas, E. (2016). Biomass and agricultural residues for energy generation. In *Fuel flexible energy generation* (pp. 59-96). Elsevier. https://doi.org/10.1016/B978-1-78242-378-2.00003-1

Alashmawy, M. M., Hassan, H. S., Ookawara, S., & Hassan, H. (2024). Fluidised bed hydrodynamics under variable air inlet configurations to enhance gasification reactor efficiency: Computational study. *Powder Technology, 439*, Article 119740. https://doi.org/10.1016/j.powtec.2024.119740

Ali, M., Mahmood, F., Magoua Mbeugang, C. F., Kumar, S., Tang, J., & Li, B. (2024). Synthesis of syngas from municipal solid waste in a fluidised bed gasifier. *Thermal Science, 28*(5), 3647-3655. https://doi.org/10.2298/TSCI231109083A

Asif, M. (2009). Sustainable energy options for Pakistan. *Renewable and Sustainable Energy Reviews, 13*(4), 903-909. https://doi.org/10.1016/j.rser.2008.04.001

Badu, P. O., Debal, M., Girods, P., Aubert, S., & Rogaume, Y. (2025). Biomass gasification in an autothermal semi-industrial fluidised bed gasifier: Syngas characterisation, and energy balance. *Biomass and Bioenergy, 197*, Article 107831.https://doi.org/10.1016/j.biombioe.2025.107831

Baskara Sethupathy, S., Chokkalingam, V., & Natarajan, E. (2012). An experimental exploration of the biomass gasification in a fluidised bed. *Journal of Renewable and Sustainable Energy, 4*(4), Article 043110. https://doi.org/10.1063/1.4742334

Basu, P. (2010). *Biomass gasification and pyrolysis: Practical design and theory*. Academic Press.

Batool, F., Islam, K., Cakiroglu, C., & Shahriar, A. (2021). Effectiveness of wood waste sawdust to produce medium- to low-strength concrete materials. *Journal of Building Engineering, 44*, Article 103237.https://doi.org/10.1016/j.jobe.2021.103237

Ben Abdallah, A., Ben Hassen Trabelsi, A., Veses, A., García, T., López, J. M., Navarro, M. V., & Mihoubi, D. (2023). Enhancing the production of syngas from spent green tea waste through dual-stage pyrolysis and catalytic cracking. *Catalysts, 13*(10), Article 1334. https://doi.org/10.3390/catal13101334

Bridgeman, T. G., Jones, J. M., Williams, A., & Waldron, D. J. (2010). An investigation of the grindability of two torrefied energy crops. *Fuel, 89*(12), 3911-3918. https://doi.org/10.1016/j.fuel.2010.06.043

Brown, R. (Ed.). (2019). *Thermochemical processing of biomass*. Wiley. https://doi.org/10.1002/9781119417637

Buragohain, B., Mahanta, P., & Moholkar, V. S. (2010). Biomass gasification for decentralised power generation: The Indian perspective. *Renewable and Sustainable Energy Reviews, 14*(1), 73-92. https://doi.org/10.1016/j.rser.2009.07.034

Chen, Z., Hu, M., Zhu, X., Guo, D., Liu, S., Hu, Z., Xiao, B., Wang, J., & Laghari, M. (2015). Characteristics and kinetic study on pyrolysis of five lignocellulosic biomass via thermogravimetric analysis. *Bioresource Technology, 192*, 441-450. https://doi.org/10.1016/j.biortech.2015.05.062

Clauser, N. M., González, G., Mendieta, C. M., Kruyeniski, J., Area, M. C., & Vallejos, M. E. (2021). Biomass waste as sustainable raw material for energy and fuels. *Sustainability, 13*(2), Article 794.https://doi.org/10.3390/su13020794

Cvetinović, D., Erić, A., Milutinović, N., Petrov, N., Anđelković, J., & Bakić, V. (2024). Thermodynamic equilibrium modelling of biomass gasification: Effects of operating conditions on gasifier performance. *Journal of King Saud University-Science, 36*(9), Article 103370. https://doi.org/10.1016/j.jksus.2024.103370

Esfahani, R. M., Wan Ab Karim Ghani, W. A., Mohd Salleh, M. A., & Ali, S. (2012). Hydrogen-rich gas production from palm kernel shell by applying air gasification in fluidised bed reactor. *Energy & Fuels, 26*(2), 1185-1191.https://doi.org/10.1021/ef2010892

Farooqui, S. Z. (2014). Prospects of renewables penetration in the energy mix of Pakistan. *Renewable and Sustainable Energy Reviews, 29*, 693-700. https://doi.org/10.1016/j.rser.2013.08.083

Guo, F., Dong, Y., Dong, L., & Guo, C. (2014). Effect of design and operating parameters on the gasification process of biomass in a downdraft fixed bed: An experimental study. *International Journal of Hydrogen Energy, 39*(11), 5625-5633.https://doi.org/10.1016/j.ijhydene.2014.01.130

Gurusamy, M. M. K. P. P. (2016). An experimental study on producer gas generation from bagasse fuel in a circulating fluidised bed gasifier. *IJSTE-International Journal of Science Technology & Engineering, 3*. http://www.ijste.org

Heidenreich, S., & Foscolo, P. U. (2015). New concepts in biomass gasification. *Progress in Energy and Combustion Science, 46*, 72-95. https://doi.org/10.1016/j.pecs.2014.06.002

Hendriyana, H. (2020). Effect of equivalence ratio on the rice husk gasification performance using updraft gasifier with air suction mode. *Jurnal Bahan Alam Terbarukan, 9*(1), 30-35. https://doi.org/10.15294/jbat.v9i1.23527

Horobet, A., Vrinceanu, G., Popescu, C., & Belascu, L. (2021). Nonrenewable energy prices and stock prices of EU financial companies: A short versus long-term analysis. In *[Book title unavailable]* (pp. 55-70). https://doi.org/10.1007/978-3-030-59972-0_5

Jabeen, F., Adrees, M., Ibrahim, M., Yamin, M., Batool, S. A., Mahmood, A., & Khalid, S. (2021). Estimation of greenhouse gases emission from domestic solid waste of Faisalabad city and schemed formula thereof. *Pakistan Journal of Agricultural Sciences, 58*(1), 69-79. https://doi.org/10.21162/PAKJAS/21.9924

Jamin, N. A., Saleh, S., & Abdul Samad, N. A. F. (2020a). Influences of gasification temperature and equivalence ratio on fluidised bed gasification of raw and torrefied wood wastes. *Chemical Engineering Transactions, 80*, 127-132. https://doi.org/10.3303/CET2080022

Jamin, N. A., Saleh, S., & Abdul Samad, N. A. F. (2020b). Influences of gasification temperature and equivalence ratio on fluidised bed gasification of raw and torrefied wood wastes. *Chemical Engineering Transactions, 80*, 127-132. https://doi.org/10.3303/CET2080022

Karmakar, M. K., Mandal, J., Haldar, S., & Chatterjee, P. K. (2013). Investigation of fuel gas generation in a pilot scale fluidised bed autothermal gasifier using rice husk. *Fuel, 111*, 584-591. https://doi.org/10.1016/j.fuel.2013.03.045

Kashif, M., Awan, M. B., Nawaz, S., Amjad, M., Talib, B., Farooq, M., Nizami, A. S., & Rehan, M. (2020). Untapped renewable energy potential of crop residues in Pakistan: Challenges and future directions. *Journal of Environmental Management, 256*, Article 109924. https://doi.org/10.1016/j.jenvman.2019.109924

Khezri, R., Azlina, W., & Tan, H. B. (2016). An experimental investigation of syngas composition from small-scale biomass gasification. *International Journal of Biomass & Renewables, 5*(1). https://doi.org/10.61762/ijbrvol5iss1art13915

Liu, L., Huang, Y., Cao, J., Liu, C., Dong, L., Xu, L., & Zha, J. (2018). Experimental study of biomass gasification with oxygen-enriched air in fluidised bed gasifier. *Science of the Total Environment, 626*, 423-433.https://doi.org/10.1016/j.scitotenv.2018.01.016

Lv, P., Chang, J., Xiong, Z., Huang, H., Wu, C., Chen, Y., & Zhu, J. (2003). Biomass air-steam gasification in a fluidised bed to produce hydrogen-rich gas. *Energy & Fuels, 17*(3), 677-682. https://doi.org/10.1021/ef0201811

Mahapatro, A., & Mahanta, P. (2020). Gasification studies of low-grade Indian coal and biomass in a lab-scale pressurised circulating fluidised bed. *Renewable Energy, 150*, 1151-1159. https://doi.org/10.1016/j.renene.2019.10.038

Maitlo, G., Ali, I., Mangi, K. H., Ali, S., Maitlo, H. A., Unar, I. N., & Pirzada, A. M. (2022). Thermochemical conversion of biomass for syngas production: Current status and future trends. *Sustainability, 14*(5), Article 2596.https://doi.org/10.3390/su14052596

Makwana, J. P., Joshi, A. K., Athawale, G., Singh, D., & Mohanty, P. (2015). Air gasification of rice husk in bubbling fluidised bed reactor with bed heating by conventional charcoal. *Bioresource Technology, 178*, 45-52. https://doi.org/10.1016/j.biortech.2014.09.111

Mallick, D., Mahanta, P., & Moholkar, V. S. (2020a). Co-gasification of biomass blends: Performance evaluation in circulating fluidised bed gasifier. *Energy, 192*, Article 116682. https://doi.org/10.1016/j.energy.2019.116682

Mallick, D., Mahanta, P., & Moholkar, V. S. (2020b). Co-gasification of coal/biomass blends in 50 kWe circulating fluidised bed gasifier. *Journal of the Energy Institute, 93*(1), 99-111. https://doi.org/10.1016/j.joei.2019.04.005

Manual, P. O. I. (2008). *Model 1341 oxygen bomb calorimeter no. 204M*.

Memon, T. A. (2022). Assessment of rice residues as potential energy source in Pakistan. *Sukkur IBA Journal of Emerging Technologies, 5*(1), 41-53. https://doi.org/10.30537/sjet.v5i1.982

Mohammed, M. A. A., Salmiaton, A., Wan Azlina, W. A. K. G., Mohammad Amran, M. S., & Fakhru'l-Razi, A. (2011). Air gasification of empty fruit bunch for hydrogen-rich gas production in a fluidised-bed reactor. *Energy Conversion and Management, 52*(2), 1555-1561. https://doi.org/10.1016/j.enconman.2010.10.023

Mohanty, R., Mahanta, P., Mahapatro, A., & Sharma, R. P. (2025). Catalytic gasification of pinewood biomass in a fluidised bed reactor with dolomite, limestone, and activated carbon: An experimental study. *Energy, 325*, Article 136131. https://doi.org/10.1016/j.energy.2025.136131

Najjar, Y. S. H. (2011). Gaseous pollutants formation and their harmful effects on health and environment. *Innovative Energy Policies, 1*, 1-9. https://doi.org/10.4303/iep/E101203

Nguyen, T. L. T., Hermansen, J. E., & Nielsen, R. G. (2013). Environmental assessment of gasification technology for biomass conversion to energy in comparison with other alternatives: The case of wheat straw. *Journal of Cleaner Production, 53*, 138-148. https://doi.org/10.1016/j.jclepro.2013.04.004

Rahmah, D. A., & Nurhilal, O. (2025). Analysis of the influence of temperature, equivalence ratio (ER) and gasifying agent against cold gas efficiency (CGE) and tar: Systematic literature review. *Jurnal Material dan Energi Indonesia, 15*(2), 82-98.

Raza, N., & Ahsan, M. (2024). Influence of distributor plate design on mixing characteristics of rice husk biomass in a bubbling fluidised bed gasifier: An experimental and CFD study. *Fuel, 358*, Article 129893. https://doi.org/10.1016/j.fuel.2023.129893

Raza, R., Ajaz, A., Rashid, M., & Nasir, A. (2022). Utilisation of spent tea waste for the development of new milk chocolate product. *Journal of Research (Science), 33*(2), 64-72.

Safdar, H. M., Nasir, A., Khan, F. U. H., & Ahmad, R. (2020). Enhancing the quality of maize, wheat, rice and cotton residue briquettes by optimising the operational parameters. *Pakistan Journal of Agricultural Sciences, 57*(1), 189-198. DOI: 10.21162/PAKJAS/20.7990

Saghir, M., Zafar, S., Tahir, A., Ouadi, M., Siddique, B., & Hornung, A. (2019). Unlocking the potential of biomass energy in Pakistan. *Frontiers in Energy Research, 7*, Article 24. https://doi.org/10.3389/fenrg.2019.00024

Salim, R. A., Hassan, K., & Shafiei, S. (2014). Renewable and non-renewable energy consumption and economic activities: Further evidence from OECD countries. *Energy Economics, 44*, 350-360. https://doi.org/10.1016/j.eneco.2014.05.001

Sánchez, J., Curt, M. D., Robert, N., & Fernández, J. (2019). Biomass resources. In *The role of bioenergy in the bioeconomy* (pp. 25-111). Elsevier. https://doi.org/10.1016/B978-0-12-813056-8.00002-9

Sanjrani, A., Sanjrani, M. A., Rind, S., & Shaikh, K. H. (2023). Solid waste management: Environmental impact and solutions, specially focussed on Pakistan. *Journal of Entrepreneurship Management and Innovation, 5*(1). https://doi.org/10.52633/jemi.v5i1

Sasujit, K., Homdoung, N., & Tippayawong, N. (2022). Production of producer gas from densified agricultural biomass in downdraft gasifier and its application to small diesel engines. *Energy Engineering: Journal of the Association of Energy Engineering, 119*(5), 2149-2167. https://doi.org/10.32604/ee.2022.022069

Serrano, D., Kwapinska, M., Horvat, A., Sánchez-Delgado, S., & Leahy, J. J. (2016). *Cynara cardunculus* L. gasification in a bubbling fluidised bed: The effect of magnesite and olivine on product gas, tar, and gasification performance. *Fuel, 173*, 247-259. https://doi.org/10.1016/j.fuel.2016.01.051

Sharma, A. M., Kumar, A., Patil, K. N., & Huhnke, R. L. (2011). Performance evaluation of a lab-scale fluidised bed gasifier using switchgrass as feedstock. *Transactions of the ASABE, 54*(6), 2259-2266. https://doi.org/10.13031/2013.40639

Shayan, E., Zare, V., & Mirzaee, I. (2018). Hydrogen production from biomass gasification: A theoretical comparison of using different gasification agents. *Energy Conversion and Management, 159*, 30-41. https://doi.org/10.1016/j.enconman.2017.12.096

Sidek, F. N., Abdul Samad, N. A. F., & Saleh, S. (2020). Review on effects of gasifying agents, temperature and equivalence ratio in biomass gasification process. *IOP Conference Series: Materials Science and Engineering, 863*(1), Article 012028. https://doi.org/10.1088/1757-899X/863/1/012028

Sikarwar, V. S., Zhao, M., Clough, P., Yao, J., Zhong, X., Memon, M. Z., Shah, N., Anthony, E. J., & Fennell, P. S. (2016). An overview of advances in biomass gasification. *Energy & Environmental Science, 9*(10), 2939-2977.https://doi.org/10.1039/C6EE00935B

Sikarwar, V. S., Zhao, M., Fennell, P. S., Shah, N., & Anthony, E. J. (2017). Progress in biofuel production from gasification. *Progress in Energy and Combustion Science, 61*, 189-248. https://doi.org/10.1016/j.pecs.2017.04.001

Susastriawan, A. A. P., & Saptoadi, H. (2023). *Journal of Sustainable Development of Energy, Water and Environment Systems, 11*(1), Article 1080435. https://doi.org/10.13044/j.sdewes.d10.0435

Susastriawan, A. A. P., Saptoadi, H., & Purnomo. (2019a). Comparison of the gasification performance in the downdraft fixed-bed gasifier fed by different feedstocks: Rice husk, sawdust, and their mixture. *Sustainable Energy Technologies and Assessments, 34*, 27-34. https://doi.org/10.1016/j.seta.2019.04.008

Susastriawan, A. A. P., Saptoadi, H., & Purnomo. (2019b). Effect of tuyer distance above grate on propagation front and performance of downdraft gasifier with the feedstock of rice husk. *Renewable Energy, 134*, 1034-1041.https://doi.org/10.1016/j.renene.2018.11.110

Tian, L., Shen, B., Xu, H., Li, F., Wang, Y., & Singh, S. (2016). Thermal behaviour of waste tea pyrolysis by TG-FTIR analysis. *Energy, 103*, 533-542. https://doi.org/10.1016/j.energy.2016.03.022

Tian, M., Li, J., Mo, W., Jiao, K., Peng, W., Yang, X., & Zhang, S. (2024). Simulation and analysis of hydrodynamic behaviour in different nozzles and its corresponding fluidised beds. *Processes, 12*(8), Article 1656.https://doi.org/10.3390/pr12081656

Untoria, S., Rouboa, A., & Monteiro, E. (2024). Hydrogen-rich syngas production from gasification of sewage sludge: Catalonia case. *Energies, 17*(6), Article 1492. https://doi.org/10.3390/en17061492

Upadhyay, D. S., Sakhiya, A. K., Panchal, K., Patel, A. H., & Patel, R. N. (2019). Effect of equivalence ratio on the performance of the downdraft gasifier: An experimental and modelling approach. *Energy, 168*, 833-846.https://doi.org/10.1016/j.energy.2018.11.133

Valin, S., Ravel, S., de Vincent, P. P., Thiery, S., Miller, H., Defoort, F., & Grateau, M. (2020). Fluidised bed gasification of diverse biomass feedstocks and blends: An overall performance study. *Energies, 13*(14), Article 3706.https://doi.org/10.3390/en13143706

Wan, C., Yu, F., Zhang, Y., Li, Q., & Wooten, J. (2013). Material balance and energy balance analysis for syngas generation by a pilot-plant scale downdraft gasifier. *Journal of Biobased Materials and Bioenergy, 7*(6), 690-695. https://doi.org/10.1166/jbmb.2013.1374

Wander, P. R., Altafini, C. R., & Barreto, R. M. (2004). Assessment of a small sawdust gasification unit. *Biomass and Bioenergy, 27*(5), 467-476. https://doi.org/10.1016/j.biombioe.2004.04.003

Wang, L., Zhou, T., Hou, B., Yang, H., Hu, N., & Zhang, M. (2025). A comprehensive review of biomass gasification characteristics in fluidised bed reactors: Progress, challenges, and future directions. *Fluids, 10*(6), Article 147. https://doi.org/10.3390/fluids10060147

Wang, Z., He, T., Qin, J., Wu, J., Li, J., Zi, Z., Liu, G., Wu, J., & Sun, L. (2015). Gasification of biomass with oxygen-enriched air in a pilot scale two-stage gasifier. *Fuel, 150*, 386-393. https://doi.org/10.1016/j.fuel.2015.02.056

Win, M. M., Asari, M., Hayakawa, R., Hosoda, H., Yano, J., & Sakai, S. (2019). Characteristics of gas from the fluidised bed gasification of refuse paper and plastic fuel (RPF) and wood biomass. *Waste Management, 87*, 173-182. https://doi.org/10.1016/j.wasman.2019.02.002

Xu, G., Murakami, T., Suda, T., Matsuzawa, Y., & Tani, H. (2009). Two-stage dual fluidised bed gasification: Its conception and application to biomass. *Fuel Processing Technology, 90*(1), 137-144. https://doi.org/10.1016/j.fuproc.2008.08.007

Xue, G., Kwapinska, M., Horvat, A., Li, Z., Dooley, S., Kwapinski, W., & Leahy, J. J. (2014). Gasification of *Miscanthus* × *giganteus* in an air-blown bubbling fluidised bed: A preliminary study of performance and agglomeration. *Energy & Fuels, 28*(2), 1121-1131. https://doi.org/10.1021/ef4022152

Yoon, S. J., Son, Y.-I., Kim, Y.-K., & Lee, J.-G. (2012). Gasification and power generation characteristics of rice husk and rice husk pellet using a downdraft fixed-bed gasifier. *Renewable Energy, 42*, 163-167. https://doi.org/10.1016/j.renene.2011.08.028

Zheng, J.-L., Zhu, M.-Q., Wen, J.-L., & Sun, R. (2016). Gasification of bio-oil: Effects of equivalence ratio and gasifying agents on product distribution and gasification efficiency. *Bioresource Technology, 211*, 164-172.https://doi.org/10.1016/j.biortech.2016.03.088

Pertanika J. Sci. & Technol. 34 (2): 1139 - 1161 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Spectral Acoustic Characterisation for Bearing Fault Detection in Marine Propulsion Motors

Iradiratu Diah Prahmana Karyatanti[1]*, Istiyo Winarno[1], Muhammad Taufiqurrohman[1], Dwisetiono[2], and Ardik Wijayanto[3]

[1]Department of Electrical Engineering, Faculty of Engineering and Marine Science, Hang Tuah University, 60111 East Java, Indonesia

[2]Department of Marine Engineering, Faculty of Engineering and Marine Science, Hang Tuah University, 60111 East Java, Indonesia

[3]Department of Electronic Engineering, Electronic Engineering Polytechnic Institute of Surabaya, 60111 East Java, Indonesia

ABSTRACT

Damage to the bearings of a ship's propulsion motor represents a significant reason why ships experience operational disruptions. Therefore, it is important to detect and diagnose early on problems related to the bearings of a propulsion motor early on to reduce maintenance costs and prevent potential maritime accidents. Additionally, since bearings subjected to variable loading and speed in dynamic environments have the possibility for various types of damage to occur, an accurate, non-destructive detection method is required to allow for early indication of damage to bearings. A condition monitoring system based on acoustic signals was developed within this study to be able to monitor the health of the bearings of a propulsion motor without damaging them. A method called Short Time Fourier Transform (STFT) was used to analyse the time vs frequency characteristics of the acoustic signal generated by the motor during various operating conditions. The STFT analysis revealed unique spectral patterns that identified different forms of bearing damage, i.e., failure of the inner-race, outer-race, ball of the bearing, and cage. Testing of the system was conducted with a variety of combinations of speed and load to identify the reaction of the system to variations in operational conditions. Monitoring ship propulsion motors using the

ARTICLE INFO

Article history:
Received: 28 October 2025
Accepted: 26 December 2025
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.22

E-mail addresses:
iradiratu@hangtuah.ac.id (Iradiratu Diah Prahmana Karyatanti)
istiyo.winarno@hangtuah.ac.id (Istiyo Winarno)
taufiqurrohman@hangtuah.ac.id (Muhammad Taufiqurrohman)
dwisetiono@hangtuah.ac.id (Dwisetiono)
ardik@pens.ac.id (Ardik Wijayanto)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

STFT algorithm has been proven to be suitable for processing non-linear data with a high accuracy of 97.5%. Development of this system will provide a safe and environmentally friendly way to monitor the operation of equipment at sea.

Keywords: Acoustic signals, bearing fault detection, ship propulsion motor, spectral analysis, STFT

INTRODUCTION

The motor bearing is a critical part of the propulsion system on a ship. It performs a key function in providing navigational safety and operational efficiency. Therefore, it is extremely important that the motor bearing is reliable as it affects continuity of operations, fuel consumption, and the overall safety of the vessel. An integral part of the motor bearing is the bearing itself, which assists in reducing friction between moving parts and provides rotational load support (Wu et al., 2025). Damage to the bearing may cause increased friction, higher than normal energy use, total system failure, and large amounts of repairs, which can create significant unplanned downtime for the vessel (Abbasi et al., 2025). Historically, ship propulsion motor bearings were maintained using two types of corrective maintenance. The first type was a reactive maintenance method. Maintenance was performed after the component failed. The primary drawback of this type of maintenance was the costly and unplanned vessel downtime. The second type of maintenance historically used was predictive or time-based maintenance. Maintenance was performed at a set time based on when the maintenance interval was reached, regardless of the actual condition of the bearing being assessed. The main drawback of this type of maintenance was unnecessary maintenance on components that were operating satisfactorily and thus producing high cost, and additional waste from the industry. To improve on these previous types of maintenance, designers and engineers developed and continued to implement intelligent and real-time monitoring systems to provide the capability to continuously monitor the condition of the bearings during operation (Karatuğ et al., 2023).

The utilisation of vibration data, current, and flux to predict the failure of bearings is an established means of predictive maintenance because it does not require disassembly of the machine and can be performed while the machine is running with no interruption in production. All three methods have their own unique disadvantages. Some of the most obvious disadvantages of using a vibration-based method of failure prediction include the high susceptibility of this type of method to external influences (i.e., adjacent equipment), which can create noise that will reduce the reliability of the analysis. A disadvantage of this type of method is the necessity of installing accelerometer sensors directly onto the bearing housing. The installation of accelerometers onto the bearing housing is often difficult to accomplish in closed systems or when machines are operated in extremely hostile environments (Chu et al., 2024).

In terms of costs, sensors, and associated equipment for collecting sensor data are also somewhat expensive. Additionally, with slow-rotating machines, the amplitudes of vibration produced will be low; therefore, detecting defects during their initial development phases will be extremely difficult. On the other hand, the method of fault diagnosis via current signature analysis has several benefits. The first is that no additional sensors need to be attached to the machine, as the diagnosis process uses the existing stator current of the induction motor (Niu et al., 2023). However, the diagnostic technique suffers from a lack of specificity. Several possible causes exist which could produce fluctuations in the currents, including changes in loads, voltage imbalance, or other types of electrical system disturbance. Therefore, determining that the cause of these fluctuations was due to a defective bearing would be very difficult. Typically, an increase in current levels occurs when significant damage has occurred to the bearing. At the initial stages of damage, the bearing condition may not have resulted in a noticeable increase in the current pattern. As a result, the current signature method alone will typically not provide sufficient information to diagnose early-stage bearing damage. Advanced signal processing techniques will likely be required, such as Wavelet Transform, Short Time Fourier Transform (STFT), or Fast Fourier Transform (FFT) (Kumar & Anand, 2024).

The monitoring of bearing conditions based on magnetic flux (Flux Signal Analysis) is an infrequently used method and still appears to be in the research and development phase in several different branches of industry. The measured magnetic flux signals exhibit very weak amplitudes and are sensitive to electromagnetic noise from the environment. Earlier damage in the bearings will usually produce only very slight changes in the flux signals, making them difficult to correct from the normal signals. In addition, the flux depends strongly on operating conditions of the motor, e.g., voltage, current, temperature, etc., and is thus quite complicated to analyse and requires the use of mathematical models or intelligent algorithms for correct interpretation (Čulić et al., 2005). Disregarding the shortcomings of the other methods, however, the most effective method of bearing condition monitoring is by using the simultaneous integration of several data sources by sensor fusion techniques or multi-domain integration methods. Thus, the accuracy of the detection of bearing damage is improved, and the availability of timely maintenance decisions is assured.

Monitoring bearings using acoustic data is beneficial for several reasons. It is often possible to detect early signs of damage using acoustic data before detecting damage through vibration or electrical currents (Lu et al., 2025). Acoustic sensors can be mounted externally; therefore, they offer less complexity than vibration sensors. In addition, acoustic signals contain high-frequency information; thus, they enable the detection of subtle symptoms such as micro-friction or small crack formation (Shuai et al., 2023). Moreover, the digitalised signal processing enables

the detection and classification of damage in real-time through the acoustic signals. Characteristic sound patterns are emitted by anomalies in the operation of motors such as bearing wear, rotor imbalance, misalignment, and abnormal friction (Suman et al., 2022). The acoustic signals are measured with high-precision microphones placed at a certain distance to provide operator safety while enhancing the detection capability of the system (Karyatanti et al., 2025).

Damage to the bearings of a ship's propulsion motor is one of the most critical factors that can affect the reliability of a ship's propulsion system. Several condition monitoring techniques have been developed that include those based on vibration, current, and magnetic flux. Each of these techniques has some disadvantages, including the requirement for invasive sensors and issues with environmental noise levels, as well as lower accuracy during the early stages of damage or when there are changes in load and speed. To address these shortcomings, the study proposes a novel condition monitoring technique for bearings based on acoustic signals. The proposed condition monitoring technique is non-invasive, real-time, and adaptive to allow adjustments to variations in the operating conditions of the ship's motor regarding load and speed. The ability to utilise acoustic signals provides for high sensitivity in the early detection of damage, even when the disturbance level is at a micro-level and cannot be detected by traditional methods.

This condition monitoring system uses time-frequency analysis using the Short-Time Fourier Transform (STFT) method to extract the spectral characteristics associated with different types of bearing damage, i.e., the inner-race, outer-race, ball of the bearing, and cage. The primary innovation of this study is the development of an acoustic-based condition monitoring system for ship propulsion motor bearings that operates under dynamic conditions, particularly where the load and/or speed of the motor are changing. Unlike many previous studies, which evaluated systems at static loads or at constant speed, this study evaluates systems during variable loads and/or speeds, which is common in actual maritime environments. The system can identify and localise the type and location of bearing damage (i.e., inner-race, outer-race, ball of the bearing, and cage) based on spectral characteristics extracted from STFT analyses of acoustic signals. Additionally, the use of acoustic sensors as the primary data source provides a cost-effective and non-invasive condition monitoring system compared to vibration or electrical current monitoring systems, which require expensive contact sensors and complicated installation procedures. The system is also reliable in non-linear signal conditions and noisy maritime environments due to the transformation of the signal to the time-frequency domain, which permits precise identification of localised changes. It is expected that the effectiveness of predictive maintenance for ship propulsion systems will be improved, and the risk of unexpected failures that affect safety and operational performance will be reduced.

MATERIALS AND METHODS

An extensive experimental testing methodology is needed to create a dependable acoustic condition monitoring system that will work successfully in a dynamic operational environment. This testing methodology includes the effects of load, motor speed, and the non-linear acoustic signal characteristics from acoustic sources. The propulsion motor is being simulated with several forms of controlled bearing damage simulating real-world operational conditions in maritime operations. All the acoustic data collected was then analysed with the STFT algorithm to transform the signals from the time domain to the time-frequency domain, providing a better illustration of anomaly patterns produced by mechanical damage. The spectrum features developed from this were utilised to determine the type and location of damage. Each step of the testing process was developed to test the dependability of the acoustic-based monitoring system under the various load and speed conditions, as well as to test how well the method works in terms of environmental disturbances and dynamic variations within the ship's propulsion system. Induction motors were chosen as the type of motor used for the propulsion system, and the bearings are artificially damaged in specific locations to simulate real-world damage conditions.

The angular speed of the motor and amount of load applied to the shaft in the propulsion system of the ship could affect the characteristics of the signals produced by the mechanical components. Thus, the effects of load and speed must be examined in this study to test the reliability of the acoustic condition monitoring system. When the motor is running at lower speeds, the acoustic signals that are produced have lower energy and a smaller range of frequencies. Under these conditions, it may be more difficult to see the effects of bearing damage since the intensity of the acoustic signals is still within the range of what is considered normal. Conversely, when the motor is running at higher speeds, the effect of bearing damage produces a significant increase in the energy of the acoustic signals, which are spread out over a larger range of frequencies, thus making it easier to see the effects of bearing damage through spectral analysis.

The amount of load also affects the dynamic response of the system. Increased loads result in greater compressive forces between the bearing components, therefore producing louder and more consistent sounds because of mechanical defects. In addition, when the system is under increased load, the symptoms of damage will produce a distinct and repeated sound pattern that can be recognised through time-frequency analysis utilising STFT. Conversely, under decreased loads, minor damage may not produce enough acoustic disturbance to be recognised.

Therefore, the goal of this research is to incorporate various speeds and loads into the testing to evaluate the sensitivity and robustness of the acoustic monitoring system to changes in operational conditions. It is essential to establish that the system can function

in an adaptive manner in dynamic maritime environments, while continuously producing accurate diagnostic results across all modes of operation of the ship's engines.

Figure 1 is a block diagram of a bearing condition monitoring system within a ship's propulsion system, which uses acoustic signals to make non-invasive damage detections. The ship's propulsion system utilises a motor to drive the shaft and propeller. The bearing components support the shaft and facilitate the smooth rotation of the shaft and are the focal point of the monitoring because failure of the bearings can cause severe disruption to the propulsion performance. The system identifies signs of bearing damage by employing acoustic sensors in the form of high-precision microphones located near the bearing components. The microphones capture the sounds produced during the operation of the motor at various speeds and loads. The acoustic signals captured are non-linear and complex; therefore, they must be preprocessed through time-frequency conversion (i.e., STFT) to create a visual display of the sound spectrum in the form of a spectrogram. A diagram of the configuration of this block diagram is depicted in Figure 2.

Once the spectrum data has been generated, it is sent to a processing unit (based on a Raspberry Pi). Not only does this system detect the presence of damage, but it also identifies the position where damage to the bearing occurs (whether it occurs on the inner-race, outer-race, ball of the bearing, or cage). The analytical results are displayed on a monitor screen in the form of a colored spectrogram illustrating the energy intensity over time across various frequencies.

The test-bearing used in the study was a rolling-element bearing whose physical condition was illustrated in Figure 3. It was shown how the bearing exhibited different types of damage to its raceways (the circular surface where the balls run). There were three types of damage shown: in Figure 3(a), the bearing had cracks on its outer race (minor defects); in Figure 3(b), the bearing had similar minor defects on the inside raceway (inner race); and in Figure 3(c), the bearing had more serious damage to the outer race. These types of damage were artificially created to simulate typical failures experienced in marine propulsion systems (e.g., through wear, excessive loading, or imbalanced operation of the motor).

These artificial failures in the bearing were used as the test object for assessing the accuracy of an acoustic signal-based condition monitoring system to identify and classify both the location and type of damage occurring to the bearing races. The acoustic signal generated by the rotating motor with a damaged bearing has distinguishable spectral features relative to the signal generated by a normal condition. A machine learning-based classification technique and time-frequency transformation were used to predict the ability of the system to accurately determine whether the damage was in either the outer or the inner race. Testing with the perforated bearing object in the validation phase of the system will allow for comparison of the performance of the system in identifying damage (both in terms of magnitude and location) in both a healthy condition and a damaged condition.

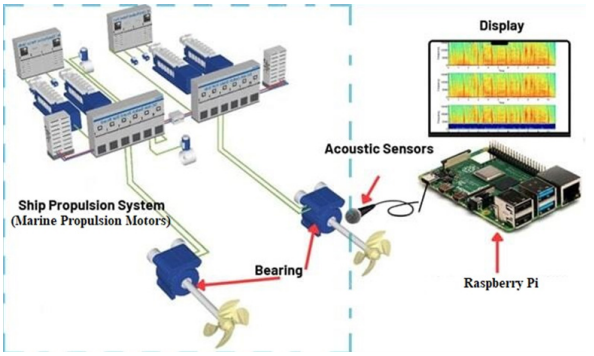


Figure 1. Scheme of the monitoring system for the condition of ship propulsion engines

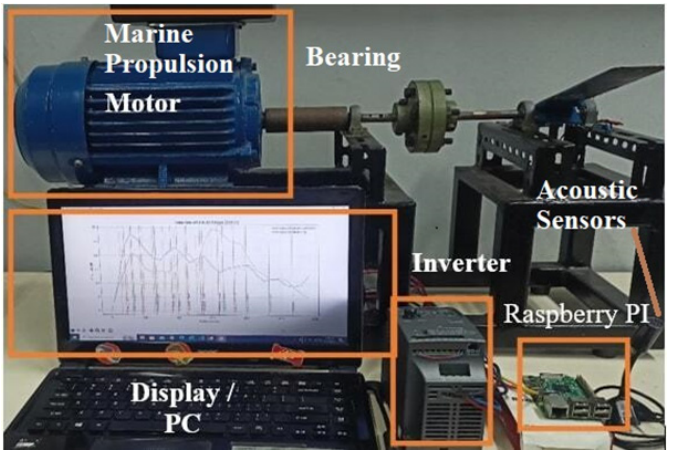


Figure 2. Implementation of a test system for monitoring the condition of ship propulsion engines



Figure 3. Test bearings for ship propulsion motors: (a) outer race minor defect; (b) inner race minor defect; (c) outer race serious damage

In addition, this will enable assessment of the sensitivity of the acoustic sensor and the associated signal processing algorithms to noise caused by damage located at different positions within the bearing.

Experimental tests are important to demonstrate the ability of the monitoring system to carry out real-time damage detection within the dynamic operating environment of a vessel. One of the more common methods of analysing the acoustic signals generated from a motor in terms of research work carried out regarding the use of acoustic signals for condition monitoring of propulsion motor bearings is the STFT method. This method is in some part popular because the acoustic signals generated from a motor are non-stationary (especially if there are present mechanical faults). A signal is classified as being non-stationary if the frequency contents are changing with time. Therefore, it will be appreciated that a Fourier Transformation will not give sufficient information about the local alterations in the frequency contents due to mechanical faults, giving only the global frequency contents of the signal. The STFT overcomes this disadvantage through time-frequency analysis of signals, allowing the study of spectral changes developed as the motor turns. The STFT enables this to be done because it divides the long length of the signal into small segments, which are considered as being locally stationary, and a window function is applied to each section to reduce the effect of spectral leakage. After the preceding procedures are put into practice, a Fast Fourier Transform (FFT) is employed. The result is a time-frequency spectrum or spectrogram, which shows signal energy distribution on both parameters, both in time and frequency. The formulation of STFT is provided below (Geraei et al., 2024).

$$STFT\{x(t)\}(t,\omega) = \int_{-\infty}^{\infty} x(\tau).\omega(\tau - t).e^{-j\omega\tau}\, d\tau$$

where:

$x(\tau)$ = original signal in the time domain (motor sound signal)

$\omega(\tau - t)$ = window function, which is shifted over time

$e^{-j\omega\tau}$ = exponential basis of the Fourier transform

ω = angular frequency (rad/s)

t = centre of the window

An expression for the Discrete STFT is described below:

$$STFT\left[\left[x[n]\right]\right](m,\omega) = \sum_{-\infty}^{\infty} x[n].\omega(n - m).e^{-j\omega n}\, d\tau$$

$x[n]$ = discrete signal (sampled from a microphone)

$\omega(n-m)$ = fixed-length window function shifted to index m

The results of the Short-Time Fourier Transform are usually displayed as a spectrogram, which is a two-dimensional graph of energy amplitude in time and frequency.

When a bearing is damaged, the spectral characteristics of the acoustic signal generated will display a particular frequency pattern that is a strong indicator of mechanical defects. When a bearing element is damaged (for example, cracked, worn or holed), the repeated collision of the bearing element with other parts of the bearing produces mechanical impulses that are recorded as non-stationary signals. As previously stated, these impulses can be identified via spectrum analysis as amplitude spikes at certain frequencies known as the bearing characteristic frequencies.

Damage to a bearing could cause an increase in the magnitude of the vibrations at the harmonic frequencies generated by an induction motor. Due to the asymmetry of the flux intensity in the air gap, the inductance of the stator changes as well. The acoustic signals contain information about the status of defects or anomalies found in the motor (Wu et al., 2022). Bearing frequency is determined from the calculation of kinematical parameters, geometrical parameters, and the rotational speed of the bearing. When the bearing rotates along a path, the parameters include V = linear velocity, ω = angular velocity, r = distance, d = diameter of the ball bearing, θ = contact angle, D_p = pitch diameter, dan N = number of ball bearings. The linear velocity of the bearing within the cage V_c is expressed in Equation 1:

$$V_c = \frac{V_i+V_o}{2} = \frac{\omega_i r_i+\omega_o r_o}{2} \quad [1]$$

with $V_c = \omega_c r_c = \omega_c \left(\frac{D_p}{2}\right)$ where $r_c = \frac{D_p}{2}$. Bysubstituting the relationship between angular and linear velocity, the cage angular velocity ω_c can be derived, as shown in Equation 2.

$$\omega_c = \left(\frac{2V_c}{D_p}\right) = \left(\frac{2}{D_p}\right).V_c = \left(\frac{2}{D_p}\right)\left(\frac{\omega_i r_i+\omega_o r_o}{2}\right) = \frac{\omega_i r_i+\omega_o r_o}{D_p} \quad [2]$$

Distances of *inner-race* (r_i) and *outer-race* (r_o) are:

$$r_i = \frac{D_p}{2} - \frac{d\cos\theta}{2} \text{ and } r_o \frac{D_p}{2} + \frac{d\cos\theta}{2}$$

Considering the geometrical relationships of the inner and outer race radii, the expression is further expanded into Equation 3.

$$\omega_c = \frac{\omega_i\left(\frac{D_p}{2}-\frac{d\cos\theta}{2}\right)+\omega_o\left(\frac{D_p}{2}+\frac{d\cos\theta}{2}\right)}{D_p} = \frac{1}{2}\left[\omega_i\left(1-\frac{d\cos\theta}{D_p}\right)+\omega_o\left(1+\frac{d\cos\theta}{D_p}\right)\right] \quad [3]$$

By converting angular velocity into frequency form using $\omega = 2\pi f$, then the cage frequency is obtained as given ini Equation 4:

$$f_c = \frac{1}{2}\left[f_i\left(1-\frac{d\cos\theta}{D_p}\right)+f_o\left(1+\frac{d\cos\theta}{D_p}\right)\right] \quad [4]$$

The frequency of the outer-race bearing rotation, also called the Ball Pass Frequency of Outer Race (BPFO), is defined in Equation 5:

$$\begin{aligned} f_{BPFO} &= N(\omega_c-\omega_o) \\ &= N\left\{\frac{1}{2}\left[f_i\left(1-\frac{d\cos\theta}{D_p}\right)+f_o\left(1+\frac{d\cos\theta}{D_p}\right)\right]-f_o\right\} \\ &= \frac{N}{2}(f_i-f_o).\left(1-\frac{d\cos\theta}{D_p}\right) \end{aligned} \quad [5]$$

The outer-race bearing frequency is assumed to be zero, since it is fixed in place outside of the motor case. Therefore, the inner-race frequency is equivalent to the frequency of the motor rotor shaft ($f_i = f_s$). It follows the simplified BPFO expression in Equation 6:

$$f_{BPFO} = \frac{N}{2}f_s\left(1-\frac{d\cos\theta}{D_p}\right) \quad [6]$$

Therefore, if calculated in the same way, the Ball Pass Frequency of the Inner Race (BPFI) could be expressed as in Equation 7:

$$\begin{aligned} f_{BPFI} &= N(\omega_i-\omega_c) \\ &= N\left\{f_i-\frac{1}{2}\left[f_i\left(1-\frac{d\cos\theta}{D_p}\right)+f_o\left(1+\frac{d\cos\theta}{D_p}\right)\right]\right\} \\ &= \frac{N}{2}(f_i-f_o)\left(1+\frac{d\cos\theta}{D_p}\right) \\ f_{BPFI} &= \frac{N}{2}f_s\left(1+\frac{d\cos\theta}{D_p}\right) \end{aligned} \quad [7]$$

The value of the ball bearing frequency, known as Ball Spin Frequency (BSF), is calculated as presented in Equation 8:

$$\omega_r = \frac{V_r}{r_r} = \frac{(\omega_i - \omega_c)r_i}{r_r} = \frac{\left(\omega_i - \left\{\frac{1}{2}\left[\omega_i\left(1 - \frac{d\cos\theta}{D_p}\right) + \omega_i\left(1 + \frac{d\cos\theta}{D_p}\right)\right]\right\}\right).r_i}{r_r}$$

$$\omega_r = \frac{D_p}{2d}(\omega_i - \omega_o)\left[1 - \left(\frac{d\cos\theta}{D_p}\right)^2\right] f_r = \frac{D_p}{2d}(f_i - f_o)\left[1 - \left(\frac{d\cos\theta}{D_p}\right)^2\right]$$

$$f_{BSF} = \frac{D_p}{2d} f_s \left[1 - \left(\frac{d\cos\theta}{D_p}\right)^2\right] \quad [8]$$

Acoustic characteristics of the acoustic signals created by a ship's propulsion motor will clearly indicate the state of the bearings by detecting certain frequencies, specifically BPFO, BPFI, BSF, and FTF. Each of the above-mentioned frequencies relates to the damage of bearings to individual bearing parts. Under normal bearing conditions, the amplitudes of the signals at these frequencies remain relatively small and steady. However, when there is some type of damage, such as spalling, wear or cracking, the amplitude of the signal at one or more of the characteristic frequencies will significantly rise. The rise in the amplitude of the signal represents a good indicator of the damage to the bearing part associated with the respective frequency.

Time-frequency transformation methods used in conducting spectral analysis enable the visualisation of patterns of acoustic energy over multiple frequencies. If the amplitude exceeds the threshold established under the bearing being free of defects, then the system will immediately report the existence of a defect in the bearing. Moreover, the sound signal related to the fault contains harmonic components formed based on the fundamental frequency of the bearing. These harmonic components develop from the nonlinear interaction between the mechanical defects and the dynamic forces existing within the rotating system. The relationship between the harmonic frequencies and the characteristic frequencies of the bearing may be represented as Equation 9:

$$f_p = |k.f_v| \quad [9]$$

where f_p is the component of the harmonic, f_v is the frequency of bearing, and k is a constant k=1,2, 3,....

When the harmonic peaks emerge in the spectrum, they enhance the diagnosis of damage and represent the degree of mechanical disturbances occurring.

Therefore, the observation of the amplitude and distribution of the harmonic components from the acoustic signals at the characteristic frequencies of bearings provides the fundamental basis for acoustic-based damage diagnostic systems.

RESULTS AND DISCUSSION

In this chapter, the results of an investigation regarding how effectively a spectral-based technique could detect bearing problems in ship propulsion motors are described and analysed using acoustic signals recorded under regulated running conditions (both from bearings that were healthy and those that had failed).

A harmonic frequency is a measure of the relationship between the RPM of a ship's propulsion motor and acoustic characteristics created by defects in the bearings. When the bearing elements pass over a fault, they produce impulse signals. The Bearing Pitch Frequency Outer (BPFO) and Bearing Pitch Frequency Inner (BPFI) are characteristic frequencies produced by impulses at the bearing pitch frequency when the elements pass over a fault. Equation 5 and 6 were used to calculate the harmonic frequencies of the outer and inner raceways of the test motor bearing with diameter d = 7.25mm, $Cos\ \alpha = 0° = 1$, D_p= 38.5mm, and N = 9 (Table 1). The presence of impulses at each frequency within the acoustic harmonic spectrum serves as a symptom that there may be some damage. The variations in amplitude values at harmonic frequencies, as well as the distribution of energy in the resulting spectrum and spectrogram, are used to determine the acoustic characteristics that signify what kind of damage is present. Loaded tests were performed on this investigation since it simulates real world conditions more accurately than unloaded tests; therefore, the data presented in this paper should provide a more realistic and more accurate diagnostic capability.

Table 1
Harmonic frequency bearing

	The Ship's Propulsion Motor Speed			
BPFO	**1495.3 RPM**	**1560.3 RPM**	**1618.7 RPM**	**1676.1 RPM**
fo	91.02 Hz	94.98 Hz	98.53 Hz	102.03 Hz
fox2	182.05 Hz	189.97 Hz	197.07 Hz	204.07 Hz
fox3	273.07 Hz	284.95 Hz	295.61 Hz	306.10 Hz
fox4	364.10 Hz	379.94 Hz	394.15 Hz	408.14 Hz
fox5	455.13 Hz	474.92 Hz	492.69 Hz	510.17 Hz
BPFI	**1495.3 RPM**	**1560.3 RPM**	**1618.7 RPM**	**1676.1 RPM**
fo	133.26 Hz	139.05 Hz	144.26 Hz	149.37 Hz
fox2	266.52 Hz	278.11 Hz	288.52 Hz	298.75 Hz
fox3	399.78 Hz	417.17 Hz	432.78 Hz	448.13 Hz
fox4	533.05 Hz	556.23 Hz	577.04 Hz	597.51 Hz
fox5	666.31 Hz	695.29 Hz	721.31 Hz	746.89 Hz

Two different conditions of a ship's propulsion motor are illustrated in Figure 4: a reference signal (healthy) and test signal (damaged). Comparing these two signals will help us determine whether the reference signal (representing the healthy motor) stayed within a steady state amplitude and was not subject to extreme fluctuations. The fluctuations in amplitude indicate that the acoustic vibration caused by all the motor components functioned properly, and there was no physical interference. Although it may appear to be so at first glance, the test signal (from the damaged motor) has many characteristics that are identical to the reference signal. However, as stated previously; to properly identify and diagnose any underlying damage, these signals cannot be solely evaluated in the time domain, as depicted in Figure 4. The next critical step is to convert the signals into the frequency domain using the STFT method. This conversion enables the separation and evaluation of each frequency component contained within the signals. As such, it will allow for the identification of the specific frequencies that may be associated with bearing damage.

A time-frequency map, also known as a spectrogram, is the graphical representation of the results obtained after converting a time-domain signal into a frequency-time image. A spectrogram displays the square magnitude $[X(\tau, f)]^2$ in a two-dimensional image format. The horizontal axis in the graph represents time (τ) while the vertical axis represents frequency (f) (Figure 5).

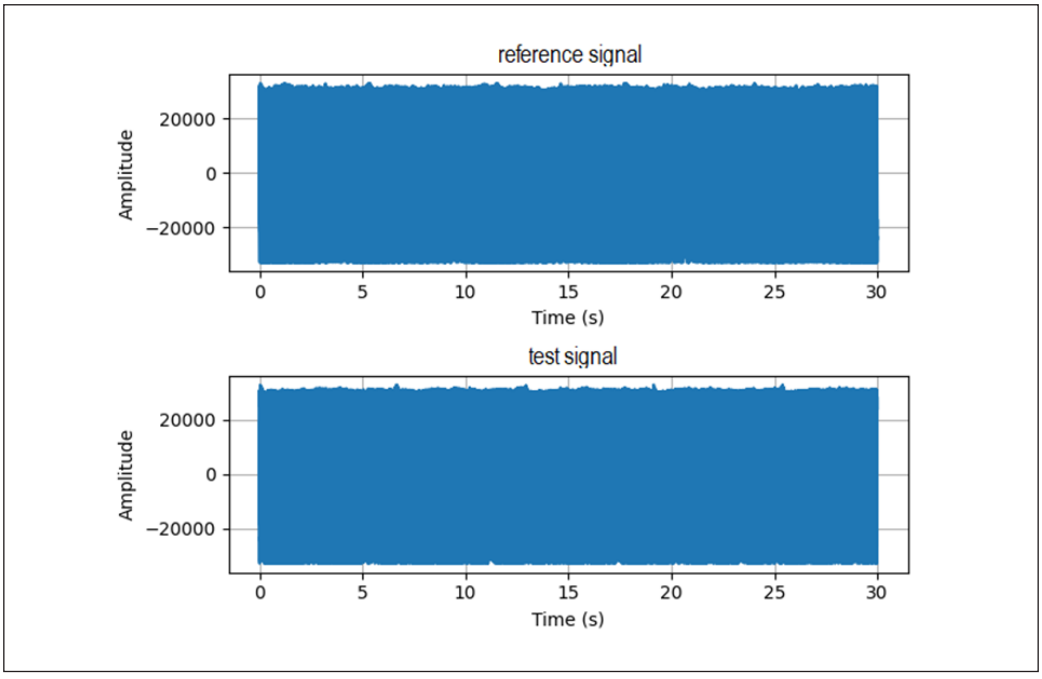


Figure 4. Reference sound signal and test sound signal under load conditions

The colours within the spectrogram illustrate the energy level of the signal, where red represents high energy levels, and blue represents low energy levels. Energy distribution is consistent and even in a healthy motor, which is representative of normal operation. Conversely, in a motor where the bearing has failed, there is a significant increase in energy level at frequencies, specifically in the low-to-mid frequency range, which is an indicator of abnormal conditions caused by bearing failure. The difference in the characteristics illustrated above serves as an important indicator that bearing damage can be identified by analysing the frequency content of acoustic signals.

Figure 6 illustrates the window amplitude of the STFT for both acoustic signals, namely the reference signal and test signal. In the context of the analysis of the condition of bearings, selecting a proper window amplitude minimises leakage; thus, it allows for the peaks of the BPFO and their harmonics to appear clearly in the spectrum. Therefore, if a window amplitude not suitable for the application is chosen, the energy of other frequencies will leak into the BPFO and BPFI frequency bands and lead to a false diagnosis. For the reference signal window, it is seen that the amplitude of the window is concentrated primarily in the low frequency range (approximately 0-1000 Hz), with the largest peak occurring approximately 14,000 magnitude units. Following this, the amplitude of the window decreased significantly, eventually reaching near-zero magnitudes at higher frequencies. This pattern illustrates that the acoustic signal from the motor, when it is in a healthy condition, is typically clean, stable, and devoid of harmonic disturbances or abnormal frequencies.

In addition to the characteristics mentioned above, the signal test window graph shows greater amplitude peaks (about 17,500 units of magnitude) and a less stable amplitude profile, characterised by the occurrence of several amplitude spikes in the mid-high frequency range (from 1000 to 10,000 Hz). The existence of these supplementary frequency components, therefore, constitutes a clear sign of anomaly or damage to a bearing that produces additional acoustic phenomena at specific frequencies. In other words, the test signal displays spectral anomalies that do not appear in the reference signal; therefore, the STFT method is demonstrated to be effective in identifying localised damage (for example, in the inner or outer races of a bearing) through the differences in the patterns of amplitude spectra in the frequency domain.

Figure 7 to 10 show the results of sound signal extraction using the STFT method to identify frequency characteristics related to the condition of the bearings in the ship's propulsion motor. These graphs show a comparison between two signal conditions, namely the average reference amplitude (blue line) as the normal bearing condition and the average test amplitude (orange line) as the tested bearing condition or one that is indicated to be damaged. The tests were conducted at several different speeds: 1495.3 rpm (Figure 7), 1560.3 rpm (Figure 8), 1618.7 rpm (Figure 9), and 1676.1 rpm (Figure 10).

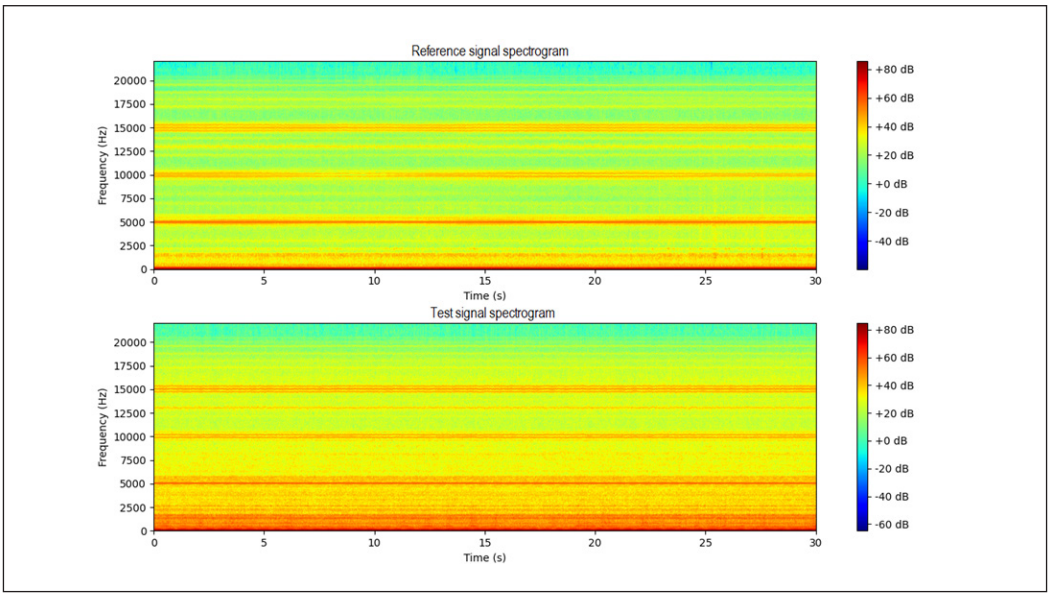


Figure 5. Spectrogram of sound signals in testing the outer race and inner race conditions of propulsion motors

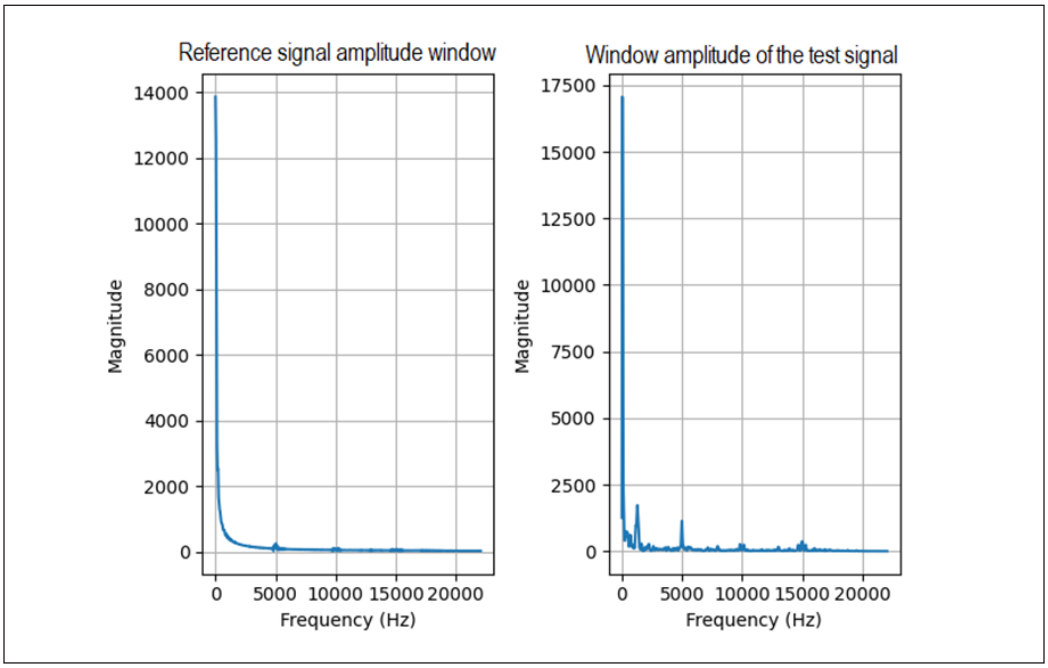


Figure 6. Acoustic signal window from the STFT process

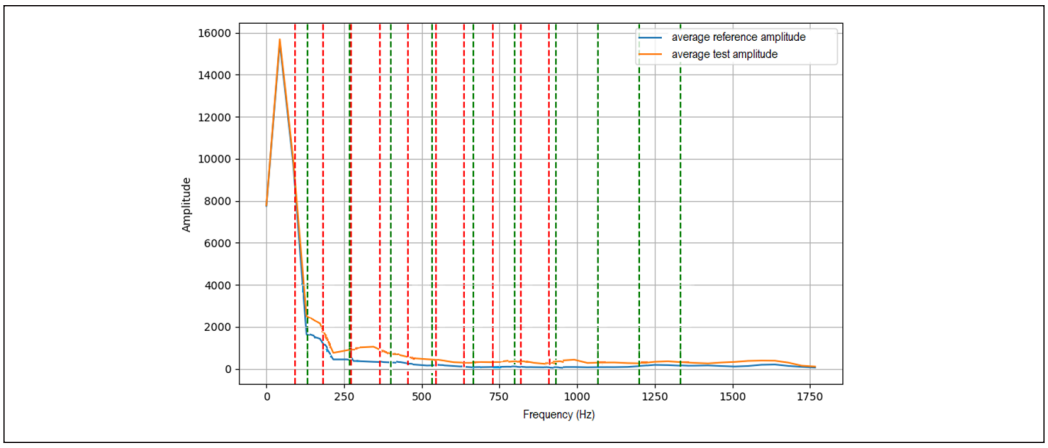


Figure 7. Average amplitude graph, motor with a speed of 1495.3 rpm

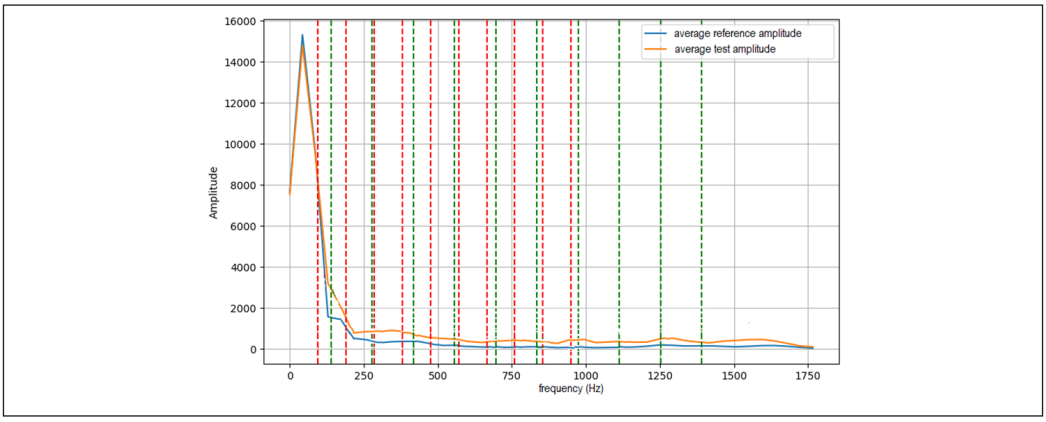


Figure 8. Average amplitude graph at a speed variation of 1560.3 rpm

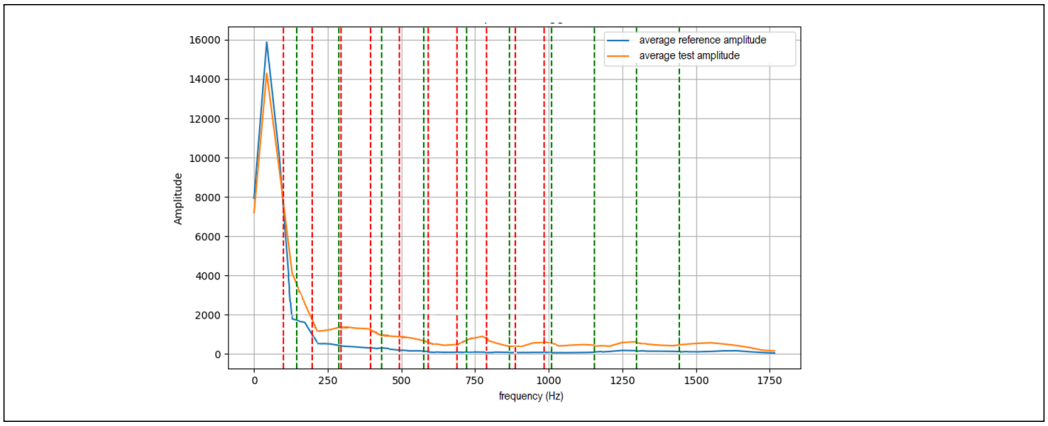


*Figure 9***.** Average amplitude graph of the motor at a speed of 1618.7 rpm

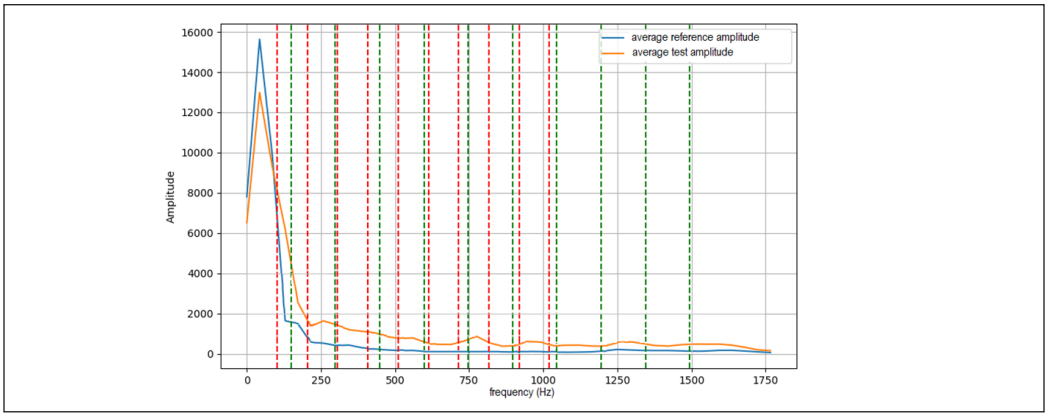


Figure 10. Average amplitude graph of the motor at a speed of 1676.1 rpm

The graph is the basis for spectrum analysis to determine whether the bearing is damaged or still in normal condition. Observations are focussed on the amplitude at the bearing harmonic frequency according to Table 1. When the test signal has an increase in amplitude at the inner-race or outer-race frequency, this indicates damage. Under load conditions, the sound signals produced by the motor contain more information due to the increased compressive force applied to the bearing components, which, in the event of defects or wear, will produce vibrations or sounds that are more pronounced and clearer.

Table 2 shows the results of the accuracy recap of the bearing condition monitoring performed on the propulsion motor of the vessel, about the detection of defects in the outer race (BPFO) and inner race (BPFI) bearings, at a range of motor speeds, from 1495.3 rpm to 1676.1 rpm. The reference value (Ref) corresponds to the acoustic amplitude measured under healthy bearing conditions, whereas the test value represents the amplitude recorded during the evaluation of the bearing (as presented in Figure7 to 10). According to the validation principle (Val) adopted, when the amplitude of the test value exceeds the reference value, the bearing is considered defective (Val = 1). This is consistent with the reality of the situation, and therefore receives a notation of 1, indicating correct monitoring (Val = 1). On the contrary, when the test value does not exceed the reference value, or does not match the actual situation, it is considered incorrect (Val = 0). As an example of a case from Table 2 regarding the BPFO at a speed of 1618.7 rpm at the third harmonic frequency (fox3) is studied. Under normal conditions (Ref), the amplitude of vibration for fox3 was recorded at 59.84. In turn, the test results of the analysed bearing (Uji) showed an amplitude of 61.79. Since the test value exceeded the reference, it met the criteria for declaring the outer race bearing as damaged. The findings are consistent with the actual field conditions and therefore resulted in a column (Val) equal to 1, which indicates a correct detection result. This example also demonstrates that the monitoring system retains

Table 2
Recapitulation of the accuracy of outer race and inner race bearing monitoring

BPFO	1495,3 Rpm			1560,3 Rpm			1618,7 Rpm			1676,1 Rpm		
	Ref	Test	Val	Ref	Test	Val	Ref	Test	Val	Ref	Test	Val
fo	79.78	79.92	1	79.35	79.67	1	79.99	79.05	0	79.30	78.58	1
fox2	62.55	65.72	1	62.49	65.3	1	53.32	59.85	1	53.55	61.07	1
fox3	52.13	57.58	1	48.44	57.16	1	50.84	61.79	1	50.48	61.89	1
fox4	49.11	59	1	50.26	57.35	1	48.64	60.98	1	47.79	59.82	1
fox5	44.83	52.49	1	47.15	53.46	1	45.56	57.54	1	42.22	56.1	1
FBFI												
fo	63.24	66.72	1	63.02	68.7	1	64.01	71.11	1	63.35	75.16	1
fox2	52.13	57.58	1	52.28	57.24	1	50.84	61.79	1	50.48	61.89	1
fox3	48.34	56.25	1	50.09	55.33	1	47.97	57.97	1	45.89	59.09	1
fox4	42.48	51.83	1	44.33	52.33	1	43.03	55.79	1	39.38	53.61	1
fox5	37.85	47.56	1	38.63	49.91	1	39.04	56.31	1	39.34	54.45	1

*Note.*Val = validation; 0 = inaccurate; 1 = accurate

its sensitivity to distinguish between healthy and damaged states, although the increase in amplitude is only 1.95 units, and particularly at specific harmonic frequencies. Such sensitivity is necessary for early detection, since slight variations in amplitude at the typical BPFO frequency can represent early warning signs of potential failure of the outer race of the bearing. Therefore, overall, the monitoring system used during the tests under different load conditions provided an accuracy of 97.5%.

Highly reliable fault detection was achieved through the high accuracy rate of the Bearing Monitoring System at 97.5%. It shows that almost all the amplitude measurement results at the characteristic frequencies of the bearings (both fundamental and harmonic frequencies) were verified against the true field conditions. Therefore, there will be fewer than 2.5% of cases where the detection results will not match, indicating an extremely low error rate for the system.

Achievement of this high accuracy rate, using the STFT signal analysis method employed, confirms that the method can distinguish between undamaged and damaged bearing conditions. The system has a high sensitivity, allowing the detection of small changes in amplitudes between reference conditions and test conditions, thus allowing the early diagnosis of damage. High accuracy indicates that the system produces consistent results under different operating conditions (i.e., varying motor speeds). In practice, an accuracy rate of 97.5% suggests that the system would provide sufficient reliability to be used as a real-time monitoring tool for propulsion motor systems on ships. Reliability is essential in the prevention of sudden failure and the reduction of risks associated with potential major damage and unanticipated maintenance costs. As such, the system can

become an important component in predictive maintenance strategies and therefore ensure that the operation of the vessels remains safe and efficient.

As demonstrated by the review of literature, regarding the variety of subject areas examined in research focused on monitoring of ship propulsion systems, the number of subject areas is very large and diverse. Subject areas include driving motors, both PMSM and IM, and the corresponding gearbox shafts, bearings, propulsion shafts, and vibrational-acoustic properties of parallel propulsion systems. Differences among subject areas directly affect the type of sensor(s) employed. For example, research on motors commonly uses measurements of currents, voltages, vibrations and temperatures, whereas research on bearings typically relies on measurements made with vibration or acoustic sensors. The signal processing methods applied in this area of research also differ because researchers attempt to tailor their analytical procedures to the nature of the measured data and the difficulties posed by the marine environment. For example, time-frequency domain analyses, such as STFT, Hilbert envelope analysis, chaotic analysis and phase space reconstruction are commonly used to capture the dynamic features of the signals caused by variations in speed and load. Also, signal decomposition methods, such as VMD, optimised using meta-heuristics such as GWO or WOA, are often applied to separate signal components from the noise.

Many studies have reported good accuracy rates, ranging from 97.5% to 100%, during laboratory experiments or simulations. Nevertheless, it is expected that the accuracy in real-world scenarios could be affected by the variability of the load, variation of the speed, and the complexity of environmental disturbances. In general, the trend in the research appears to be toward the development of intelligent diagnostic systems that are multi-sensorial, adaptable and able to perform real-time analysis. These systems combine the analysis of the domain in both time and frequency, signal decomposition, and machine learning to enable early detection and high-reliability prediction of potential failures. Accordingly, the future direction of the development will be the creation of integrated platforms that allow for monitoring the health of the entire propulsion system of a ship (engine, motors, gearboxes, shafts, and propeller) through non-invasive methods with high precision.

Compared to previous studies listed in Table 3, there are three major advantages of this method: The first advantage is that this proposed method utilises acoustic data to measure the condition of bearings of ship propulsion motors. Since this method uses a non-invasive technique to measure vibrations, it does not require direct attachment of vibration sensors to rotating components, which will make it easier to install these sensors, reduce the amount of money spent on vibration sensor maintenance and cause minimal disruptions to vessel operation. Most of the prior studies utilising vibration measurements, however, do require physical access to the rotating components and therefore, can be difficult to implement on ships at sea. The second major advantage of the proposed method is that it was designed

Table 3
Similar research on ship component monitoring

Works	Monitoring Objects	Highlight
(Anitha et al., 2024)	ship propulsion engine	Early detection of damage and real-time maintenance scheduling are achieved through data analysis involving temperature, pressure, vibration, current, or other operational parameters
(Park & Oh, 2023)	ship propulsion engine	Monitoring cylinder exhaust gas temperature (CET) and maximum cylinder combustion pressure (CCP)
(Marey & Ali, 2023)	bearings on a ship's gearbox shaft	This study monitors the condition of ship generator shaft gearbox bearings using an adaptive signal approach and deep learning models. The data used is vibration data
(Lv et al., 2025)	propulsion engines, gearboxes, and other ship propulsion components	The combination of simulation data and real data (vibration) within the Digital Twin framework effectively overcomes the limitations of damage data, while opening opportunities for the implementation of more practical and precise intelligent diagnosis systems in the maritime industry
(Lu et al., 2025)	ship propulsion shafting	This study shows that vibration signals—when analysed through chaos theory and phase space reconstruction—can be an effective method for detecting early changes in the alignment of a ship's propulsion shaft. This technique provides a strong theoretical basis for the development of real-time shaft alignment monitoring systems in the shipping industry
(Xu et al., 2024)	acoustic vibration characteristics of the ship's parallel propulsion system, namely vibration (lateral, torsional, axial)	Acoustic vibration data from the test platform is used in conjunction with an RBF neural network model to accurately predict and optimise system performance even with limited samples
(Ma et al., 2023)	Permanent Magnet Synchronous Machine (PMSM)	Specifically, the detection of inter-turn short circuits in ship propulsion motors (simulation using Simulink models)
(Kato & Otaka, 2025)	induction motor in ship propulsion	The ship propulsion motor system uses envelope processing based on the Hilbert transform to extract the characteristic fault frequency of the bearing (BPFO)
Proposed method	ship propulsion motor bearings	Monitoring ship propulsion motor bearings using acoustic data, which is non-linear data, STFT has proven to be suitable for processing non-linear data with a high sensitivity 97.5%

to work with nonlinear data; STFT was shown to provide excellent extraction of frequency information from complicated acoustic waveforms. This method also has a very high degree of detection sensitivity (97.5%), allowing it to detect small changes in the bearing condition when operates in an environment of environmental noise generated by the ship. A few alternative methods have demonstrated similar levels of accuracy using simulated or laboratory data, but none of these alternatives was tested using the nonlinear acoustic data commonly found in marine environments. The third major advantage of the proposed method is that it provides the opportunity to integrate the acoustic measurement method with existing real-time monitoring systems without requiring expensive additional sensor systems. Overall, the primary advantages of the proposed method include its ease of use, low cost, and ability to accurately perform under adverse signal conditions, all of which make it a better choice than many other methods that were developed and tested using simulated data and/or special-purpose sensors.

CONCLUSION

A monitoring system for the condition of ship propulsion motor bearings was successfully designed and tested using acoustic signal analysis based on the STFT. These test results clearly show that this method has great potential for detecting changes in bearing condition and shown a great level of sensitivity at an accuracy rate of 97.5%, even though the data from the bearing conditions were nonlinear and affected by both the ambient noise from the vessel and variability in the rotational speeds of the motor. An additional benefit of this acoustic methodology is that it is noninvasive, which will enable easy implementation into vessels' operating environments without disruption or requiring sensors directly installed onto rotating components. Therefore, given the low cost and ability to be used in real time, this method provides a viable solution for developing real-time monitoring systems for use onboard ships while improving operational reliability, preventing catastrophic failures, and optimising maintenance scheduling. Additionally, the success of this method opens the door to the integration of the Internet of Things (IoT) and digital twin technologies to enhance predictive analytics in the maritime industry.

ACKNOWLEDGEMENT

This research was funded by the Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, and Teknologi Republik Indonesia for the year 2025 contract number 128/C3/DT.05.00/PL/2025; 039/LL7/DT.05.00/PL/2025, B/15/DT.05.00/PL/UHT.C7/V/2025.

REFERENCES

Abbasi, M. A., Huang, S., & Khan, A. S. (2025). Fault detection and classification of motor bearings under multiple operating conditions. *ISA Transactions, 156*, 61-69. https://doi.org/10.1016/j.isatra.2024.11.008

Anitha, K., Vijayakumar, R., Jeslin, J. G., Elangovan, K., Jagadeeswaran, M., & Srinivasan, C. (2024). Marine propulsion health monitoring: Integrating neural networks and IoT sensor fusion in predictive maintenance. In *Proceedings of the 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)* (pp. 1-6). Dehradun, India. https://doi.org/10.1109/DICCT61038.2024.10533037

Chu, T., Nguyen, T., Yoo, H., & Wang, J. (2024). A review of vibration analysis and its applications. *Heliyon, 10*(5), Article e26282.https://doi.org/10.1016/j.heliyon.2024.e26282

Čulić, L., Kaštelan, N., Zubčić, M., & Pavić, I. (2025). Induction motor multiple fault analysis using stray flux signals: A review. *Transactions on Maritime Science, 14*(2), 5-5. https://doi.org/10.7225/toms.v14.n02.s04

Geraei, H., Rodriguez, E. A. V., Majma, E., Habibi, S., & Al-Ani, D. (2024). A noise invariant method for bearing fault detection and diagnosis using adapted local binary pattern (ALBP) and short-time Fourier transform (STFT). *IEEE Access, 12*, 107247-107260. https://doi.org/10.1109/ACCESS.2024.3438106

Karatuğ, Ç., Arslanoğlu, Y., & Soares, C. G. (2023). Design of a decision support system to achieve condition-based maintenance in ship machinery systems. *Ocean Engineering, 281*, Article 114611. https://doi.org/10.1016/j.oceaneng.2023.114611

Karyatanti, I. D. P., Winarno, I., Dewantara, B. Y., & Wibisono, A. S. (2025). The implementation of nonlinear signal techniques for enhanced monitoring of marine propulsion systems. *International Journal of Marine Engineering Innovation and Research, 10*(2), 359-367. https://doi.org/10.12962/j25481479.v10i2.6378

Kato, Y., & Otaka, M. (2025). Compressed sensing of vibration signal for fault diagnosis of bearings, gears, and propellers under speed variation conditions. *Sensors, 25*(10), Article 3167. https://doi.org/10.3390/s25103167

Kumar, R., & Anand, R. S. (2024). Health monitoring and fault analysis of induction motors: A review. *Engineering Research Express*, 1-18. https://doi.org/10.1088/2631-8695/ad8b10

Lu, L., Li, G., Xing, P., He, W., Feng, Z., & Zhang, H. (2025). Investigation on alignment state evolution of ship propulsion shafting based on chaotic characteristics of vibration signals. *Ocean Engineering, 322*, Article 120521. https://doi.org/10.1016/j.oceaneng.2025.120521

Lv, Y., Xiong, K., Yao, J., Zhao, S., & Li, Y. (2025). Virtual-physical collaborative intelligent fault diagnosis for marine rotating machinery under data scarcity towards digital twin. *Ocean Engineering, 340*, Article 122198. https://doi.org/10.1016/j.oceaneng.2025.122198

Ma, F., Qi, L., Ye, S., Chen, Y., Xiao, H., & Li, S. (2023). Research on fault diagnosis algorithm of ship electric propulsion motor. *Applied Sciences, 13*(6), Article 4064. https://doi.org/10.3390/app13064064

Marey, N. A., & Ali, A. A. (2023). Novel measurement and control system of universal journal bearing test rig for marine applications. *Alexandria Engineering Journal, 73*, 11-26. https://doi.org/10.1016/j.aej.2023.04.006

Niu, G., Dong, X., & Chen, Y. (2023). Motor fault diagnostics based on current signatures: A review. *IEEE Transactions on Instrumentation and Measurement, 72*, 1-19. https://doi.org/10.1109/TIM.2023.3285999

Park, J., & Oh, J. (2023). Anomalistic symptom judgment algorithm for predictive maintenance of ship propulsion engine using machine learning. *Applied Sciences, 13*(21), Article 11818. https://doi.org/10.3390/app132111818

Shuai, H., Junxia, L., Lei, W., & Wei, Z. (2023). Research on acoustic fault diagnosis of bearings based on spatial filtering and time-frequency domain filtering. *Measurement, 221*, Article 113533. https://doi.org/10.1016/j.measurement.2023.113533

Suman, A., Kumar, C., & Suman, P. (2022). Early detection of mechanical malfunctions in vehicles using sound signal processing. *Applied Acoustics, 188*, Article 108578. https://doi.org/10.1016/j.apacoust.2021.108578

Wu, C., Long, X., Shang, D., & Chen, F. (2025). Investigation on the influence of bearing certainty and random elevation on the friction self-excited vibration of ship propeller shaft system. *Journal of Sound and Vibration, 608*, Article 119034. https://doi.org/10.1016/j.jsv.2025.119034

Wu, G., Yan, T., Yang, G., Chai, H., & Cao, C. (2022). A review on rolling bearing fault signal detection methods based on different sensors. *Sensors, 22*(21), Article 8330. https://doi.org/10.3390/s22218330

Xu, J., Liu, X., Xie, Z., Jiao, C., Ta, N., & Rao, Z. (2024). Analysis and prediction of vibro-acoustic characteristics of parallel propulsion systems for large-scale marine ships. *Applied Ocean Research, 143*, Article 103863. https://doi.org/10.1016/j.apor.2023.103863

Pertanika J. Sci. & Technol. 34 (2): 1163 - 1180 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Quantification of Some Physical Properties of Matoa (*Pometia pinnata*) Fruit: Mass Modelling for Machine Designing and Postharvest Processing

Ubaidillah Ubaidillah*, Risty Dwi Wulandari Silalahi, Dimas Firmanda Al Riza, and Bambang Dwi Argo

Department of Biosystems Engineering, Faculty of Agricultural Technology, Universitas Brawijaya, 65145 East Java, Indonesia

ABSTRACT

Matoa (*Pometia pinnata*) is an underutilised seasonal fruit commonly consumed raw in Indonesia. Despite its nutritional, health, and economic benefits, it lacks postharvest processing and derivative products. This study quantified the physical and thermal properties of matoa fruit to aid in the design of processing equipment. The fruit's principal dimensions averaged 39.30±3.19 mm in length, 29.35±1.41 mm in width, and 28.72±1.38 mm in thickness. The mass ranged from 6.51 to 21.31 g, with volumes between 10.5 and 24.5 cm^3. The average geometric and equivalent diameters were 32.10 mm, and the fruit had an ellipsoid shape. The static coefficient of friction was highest on galvanised iron and lowest on stainless steel. Four empirical models, i.e. linear, quadratic, power, and S-curve, were evaluated to predict fruit mass using physical attributes. The multivariate quadratic model with principal dimensions (*LWT*) provided the highest R^2 value for mass prediction. These findings offer a scientific basis for designing grading, sorting, and processing equipment, and for developing postharvest and derivative products.

Keywords: Mathematical modelling, matoa fruit, postharvest handling, regression, underutilised fruit

ARTICLE INFO

Article history:
Received: 16 November 2025
Accepted: 24 December 2025
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.23

E-mail addresses:
ubaidillah88@ub.ac.id (Ubaidillah Ubaidillah)
ristysilalahi567@gmail.com (Risty Dwi Wulandari Silalahi)
dimasfirmanda@ub.ac.id (Dimas Firmanda Al Riza)
dwiargo@ub.ac.id (Bambang Dwi Argo)
* Corresponding author

INTRODUCTION

Pometia pinnata is an underutilised tropical fruit tree in the *Sapindaceae* family, and popular as Fijian longan in English or matoa in Indonesia (Matra et al., 2019). The study of the variability of the genus Pometia in Indonesia, including *Pometia pinnata*, has been reported as early as the 1960s (Jacobs, 1962). The species is distributed throughout

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

its origins across tropical Asia and the Asia Pacific region, including areas such as the Andaman Islands, Sri Lanka, Southeast Asia, and Papua New Guinea. In Java, matoa is rarely grown in dedicated orchards and typically maintained as a backyard crop. Matoa is considered a seasonal fruit, with a flowering season from around July to October and harvest season from around October to December. The fleshy fruit is edible, and it has been extensively reported for its health benefits (Suzuki et al., 2021). Its by-products also have potential bioactivities but are inedible, including seeds, rind, stem bark and leaves (Hanafi et al., 2020). The matoa tree is also widely used as timber for housing (Burley et al., 2011) and is also planted for aesthetic and functional purposes, such as fencing, marking boundaries, or providing shade.

Although the nutrient composition of the edible fruit has not been published previously (Lim, 2013), there has been research on the medicinal utilisation of matoa fruit. All the fruit parts, i.e., rind, edible arillode, and seed, have total phenolic content including n-hexane, ethyl acetate and methanol as reported previously by Irawan et al. (2017a). Irawan et al. (2017b) also evaluated the DPPH free radical scavenging activity of all fruit parts. The matoa fruit has a generally rounded to oval-elliptic form. It has a smooth rind that ranges from thin to moderately thick, displaying colour variations from yellow-green to deep red or nearly black. The rind encloses an edible gelatinous, juicy, sweet, semi-transparent white, pulpy arillode that envelops a single large, curved seed. The arillode is traditionally eaten raw, with no further processing or derivative products available on the market. The problem with seasonal fruits in Indonesia, especially matoa, is that the fruits are abundantly available in the market for a certain limited time window, with a low consumption rate due to being eaten raw. As a result, the fruit may end up as food waste during the harvest season. The fruit is considered an underutilised fruit attributed to its highly perishable character, lack of processing knowledge and derivative products, despite its nutritional and health benefits as well as economic potential. To increase its potential and benefits, further processing and derivative product development are mandatory.

Well-developed postharvest processing and its machines are urgently needed to enhance and commercialise the utilisation of matoa fruit. The physical and thermal properties and their interactions with the fruit materials must be well understood to enable the design and optimisation of processing machinery (Kamble et al., 2021). The absence of physical and thermal attributes may increase product losses, product failure, and mechanical breakdown. The appropriate processing and machinery design can result in increased work efficiency and adequate process applications (Vivek et al., 2018). The comprehensive quantifications and understanding of the engineering characteristics of matoa fruit give substantial input for researchers, machine designers, and food-product developers (Murakonda et al., 2022). Various basic processing and handling, including peeling, cutting, grading, separation, cleaning, conveying, packaging, transportation, and other postharvest tools,

are fundamentally influenced by fruit physical properties, e.g., size, mass, volume, density, shape, surface area, porosity, appearance, fraction of fruit parts, as well as fruit frictional properties (Barbhuiya et al., 2020). In addition, quantification of the thermal characteristics, namely heat conductivity, specific heat, diffusivity, and latent heat, can help in various thermal processing and preservation activities involving heat transfer, such as cooking, drying, refrigeration, thermal treatment, and freezing operations (Singh et al., 2019; Singh & Meghwal, 2019). The design of the effective storage and cooling systems for the fruit also needs to consider the usefulness of the thermal properties (Sonawane et al., 2020).

In recent years, the market acceptability of the fruit has been mainly influenced by specific traits, i.e., similar size and weight and uniform shape. Mass-based fruit grading is beneficial for packaging since it can improve packaging formations and configurations, reduce handling waste and transport resources, and develop the marketability of the fruit (Birania et al., 2022). Mass-based fruit grading is also simpler and cheaper to conduct than grading based on other attributes, i.e., by either using direct weighing or applying appropriate empirical models that can predict the fruit mass from measurable physical parameters (Mahawar et al., 2019). Since direct weighing is time-consuming and tedious, it seems plausible to develop an appropriate mathematical model to estimate the matoa fruit mass based on physical attributes. The developed fruit mass models could provide a rapid, precise, simple, and cheap fruit classification and grading process.

Numerous investigations have been conducted in the past few years for the development of an appropriate mathematical model for estimating the mass of agricultural commodities using their measurable physical parameters. These investigations encompass an array of agricultural products, spanning both underutilised and widely recognised commodities. Noteworthy examples include sohiong fruit (Vivek et al., 2018), *Terminalia chebula* and *Terminalia bellerica* fruit (Pathak et al., 2019; Pathak et al., 2020), kinnow mandarin (Mahawar et al., 2019), pepper berries (Azman et al., 2020), ash gourd seed (Gade et al., 2020), Indian coffee plum (Barbhuiya et al., 2020), kendu fruit (Panda et al., 2020), green plantain banana (Kamble et al., 2021), blood fruit (Sasikumar et al., 2021), hempseed (Saini et al., 2021), strawberry (Birania et al., 2022), wood apple (Murakonda et al., 2022), kadamb (Panda et al., 2022), amla (Tomar & Pradhan, 2022), guava (Bibwe et al., 2022), and potato (Altuntas & Mahawar, 2022).

Despite the phenotype variability of the fruit being previously reported (Zulfahmi & Rosmaina, 2024), to our knowledge, no prior publication has addressed the physical and thermal characteristics of matoa fruit. Hence, the objective of the present study was to quantify in detail various key physical and thermophysical parameters of matoa fruit that are needed for proper machine design and postharvest processing. In addition, to establish a mathematical model that appropriately correlates the primary physical variables and fruit mass of the matoa fruit. The commonly used four regression forms-linear, quadratic,

power, and S-curve-were applied for estimating the fruit mass. The findings of the present research are expected to promote enhanced utilisation and market development of matoa, thereby supporting the economic well-being of local farmers.

METHODS

Sample Preparation

Fresh and mature matoa fruits (Figure 1) were procured during the harvesting season (October 2023), sourced from smallholder orchards in Wates Village (7°55′ S, 112°7′ E), Kediri Regency, East Java, Indonesia. The collected fruits were transported promptly to the Department of Biosystems Engineering, Universitas Brawijaya, for further analysis. After manual removal of peduncles, leaves, and dust, the fruits were sorted by eliminating defective, broken, and bruised fruits and stored at ambient temperature until further experiments. The measurements were conducted in less than 24 h after harvest. The experiments on physical and thermal properties were conducted in the Laboratory of Agricultural Power and Machinery, Universitas Brawijaya. In total, 200 samples were randomly selected for physical and thermal properties quantification, i.e., 183 samples for physical properties and the remaining for thermal properties.

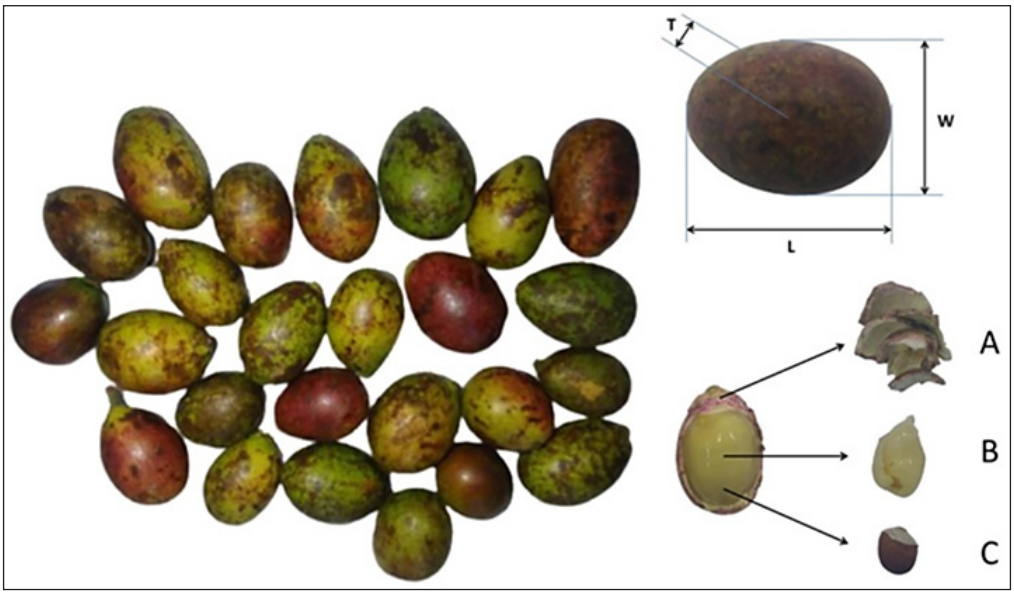


Figure 1. Representation of collection and measuring principal dimension of matoa (Pometia pinnata) fruit: (A) rind; (B) edible arillode; and (C) seed

Quantification of Physical Properties

Dimensional Properties of the Matoa Fruit

The 183 fruits were selected at random from the bulk sample to determine the physical characteristics of matoa. The dimensional properties, i.e., axial dimension, mass, and volume of each matoa fruit, were measured manually. The principal dimensions of the fruit, namely length (L), width (W), and thickness (T), as illustrated in Figure 1, are measured using a digital vernier calliper with ±0.01 mm accuracy (Mitutoyo, Japan). The mass of each fruit (M) was measured using a digital balance with ±0.001 g accuracy (FSR-A320, Fujitsu, Japan). The volume of each individual fruit was measured by employing the water displacement method (Mahawar et al., 2019; Vivek et al., 2018). Table 1 shows the equations used for the quantification of physical properties of matoa fruits.

Gravimetric Properties

The cardboard container with certain dimensions (235×175×70) mm^3 was utilised to determine the matoa fruit bulk density sample, and calculated using the following equation (Panda et al., 2020), where ρ_B, m_{c+f}, m_c, and V_c represent bulk density of fruit, mass of container with fruit, mass of container, and volume of container, respectively as shown in Equation 1.

$$\rho_B = \frac{m_{c+f} - m_c}{V_c} \quad [1]$$

The true density (ρ_T) was calculated using Equation 2, which represents the ratio of a unit mass of each fruit (M) to its true volume (V_t) (Pathak et al., 2019).

$$\rho_T = \frac{M}{V_t} \quad [2]$$

The porosity or void fraction is defined as the ratio of inter-granular voids to the total space occupied by the fruit, and calculated using Equation 3, where ε, ρ_B, and ρ_T represent porosity, bulk density, and true density, respectively (Barbhuiya et al., 2020).

$$\varepsilon = \frac{\rho_T - \rho_B}{\rho_T} \times 100 \quad [3]$$

Table 1
Equations for the physical properties of the matoa fruits

Attributes	Equation
Arithmetic mean diameter (D_a)	$D_a = \frac{(L+W+T)}{3}$
Geometric mean diameter (D_g)	$D_g = (L \times W \times T)^{\frac{1}{3}}$
Equivalent mean diameter (D_e)	$D_e = \left(\frac{\left\{ L \times (W+T)^2 \right\}}{4} \right)^{\frac{1}{3}}$
Surface area (SA)	$SA = \pi (D_g)^2$
Projected area perpendicular to length (PA_L)	$PA_L = \frac{\pi LW}{4}$
Projected area perpendicular to width (PA_W)	$PA_W = \frac{\pi WW}{4}$
Projected area perpendicular to thickness (PA_T)	$PA_T = \frac{\pi TW}{4}$
Criteria projected area (CPA)	$CPA = \frac{PA_L + PA_W + PA_T}{3}$
Sphericity (Ø)	$A_R = \frac{W}{L}$
Aspect ratio (A_R)	$A_R = \frac{W}{L}$
Flakiness ratio (F_R)	$E_R = \frac{L}{W}$
Elongation ratio (E_R)	$E_R = \frac{L}{W}$
Eccentricity (e)	$e = \left[1 - \left(\frac{W}{L} \right)^2 \right]^{\frac{1}{2}}$

Table 1 (*continued*)

Attributes	Equation
Minimum radius curvature (R_{min})	$R_{\min} = \frac{H}{2}$
Maximum radius curvature (R_{max})	$R_{\max} = \frac{H^2 + \frac{L^2}{4}}{2 \times H}$
Ellipsoid volume (V_{ellip})	$V_{ellip} = \frac{4\pi}{3} \times \left(\frac{L}{2}\right) \times \left(\frac{W}{2}\right) \times \left(\frac{T}{2}\right)$
Oblate spheroid volume (V_{osp})	$V_{osp} = \frac{4\pi}{3} \times \left(\frac{L}{2}\right) \times \left(\frac{W}{2}\right)^2$
Prolate spheroid volume (V_{pro})	$V_{pro} = \frac{4\pi}{3} \times \left(\frac{L}{2}\right)^2 \times \left(\frac{W}{2}\right)$

Note. Adapted from Sahin and Sumnu (2007) and Thomas et al. (2025)

Frictional Properties

The static friction coefficient of matoa fruit was measured using a friction testing apparatus employing five surface materials, i.e., plywood, galvanised iron sheet, stainless steel sheet, rubber sheet, and cardboard. The materials were chosen since they are the most common materials used for machinery, transportation, and packaging. The measurements of the static coefficient of friction were made in two orientations, i.e., along the fruit's longitudinal axis and across its radial directions from the calyx end. Each fruit was placed on the surface material, and the platform was slowly inclined until the fruit began to move and eventually slide or roll off. The static coefficient of friction for matoa fruit is then calculated using Equation 4 as follows, where μ and α represent the static coefficient of friction and the angle of friction or inclination angle at which the fruit starts to roll or fall (°) (Panda et al., 2020).

$$\mu = \tan\alpha \quad [4]$$

Fruit Part Fraction Determination

The fruit fraction was determined to estimate the portion of the rind, arillode, and seed for one individual matoa fruit. Each fruit was manually peeled by hand and separated for

each fruit part, i.e., rind, arillode, and seed (Figure 1). The masses of the rind, arillode, and seed for each fruit were recorded using a digital balance, and their proportions relative to the whole-fruit mass were calculated.

Quantification of Thermal Properties

The four primary thermal properties of matoa, namely specific heat capacity (*Cp*), thermal conductivity (*K*), thermal diffusivity (μ), and latent heat (λ), were estimated using the moisture content of the edible arillode. The moisture content of the edible arillode of matoa fruit was determined using a hot-air oven at 105±5°C for 24 h according to (Association of Official Analytical Chemists, 2000). Since the moisture content (m) of the arillode is higher than 60%, the standard mathematical formulas as in Equations 5-8 proposed by Sweat (1974) were used, where Cp, K, μ, λ, ρ_T, W, and m represent specific heat capacity, thermal conductivity, thermal diffusivity, latent heat of fusion, true density, weight, and moisture content.

$$Cp = 1.675 + 0.025m \quad [5]$$

$$K = 0.148 + 0.00493m \quad [6]$$

$$\mu = \frac{K}{\rho_T Cp} \quad [7]$$

$$\lambda = 335W \quad [8]$$

Mass Modelling of the Matoa Fruit

The non-linear regression analysis of data from the physical properties of matoa fruit was done using the Microsoft Excel 2010. The four most commonly used models to predict the mass of agricultural products Equations 9 to 12, i.e., linear, quadratic, power, and S-curve, were applied to analyse their suitability for mass prediction of matoa fruit, where M is the mass of the fruit, X is the considered physical parameter utilised for evaluating the relationship with mass, and a, b, and c are the constants of curve fitting (Saini et al., 2022).

$$M = a + bX \quad [9]$$

$$M = a + bX + cX^2 \quad [10]$$

$$M = a(X)^b \quad [11]$$

$$M = a + \left(\frac{b}{X}\right) \quad [12]$$

The mass data of matoa fruit was modelled using independent variables from the physical attributes of the matoa fruit, namely dimensional properties, area, and volume (Kamble et al., 2021), and the whole model was classified as follows:

1. Dimension-based models: single or multiple variable regression of dimensional attributes of matoa fruit: *L*, *W*, *T*, *Da*, *Dg*, and *De*;
2. Area-based models: single variable regression of area attributes of matoa fruit: *SA*, PA_L, PA_W, PA_T, and *CPA*;
3. Volume-based models: single variable regression of volume attributes of matoa fruit: *V*, V_{ellip}, V_{osp}, and V_{pro}.

Statistical Analysis and Model Validation

Descriptive statistics (mean, minimum, maximum, and standard deviation) for all measured physical and thermal attributes of the fruit were calculated using the Microsoft Excel 2024 software. The applied model's performance was tested using three statistical indicators, i.e., the coefficient of determination (R^2), root-mean-square error (*RMSE*), and reduced chi-square (χ^2) between actual and predicted mass values. These indicators were used to identify the most suitable model, with R2 serving as primary selection criterion, whereas RMSE and χ^2 were used to evaluate the model's goodness of fit. The regression model which has the higher coefficient of determination (R^2) and lower root-mean-square error (*RMSE*) as well as reduced mean square of deviation (χ^2) were selected as the best fit (Saini et al., 2022). RMSE and χ^2 were computed using Equations 13 and 14, respectively, where $M_{ex,i}$ denotes the *i*-th measured mass, $M_{pr,i}$ represents the *i*-th predicted mass, *N* is the total number of samples, and *z* is the number of model parameters

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(M_{pr,i} - M_{ex,i}\right)^2\right]^{1/2} \quad [13]$$

$$\chi^2 = \frac{\sum_{i=1}^{N}\left(M_{ex,i} - M_{pr,i}\right)^2}{N - z} \quad [14]$$

RESULTS AND DISCUSSION

Physical Properties

The dimensional characteristics of whole matoa fruits were evaluated at a natural moisture content of 37.55±2.75% (w.b.), and the corresponding mean values with standard deviations

are summarised in Table 2. Based on the measurement results, the mean values for fruit length (*L*), width (*W*), and thickness (*T*) were 39.30±3.19 mm, 29.35±1.41 mm, and 28.72±1.38 mm, respectively. The dimensions of matoa fruit are relatively bigger than both dimensions of the fruit with similar shapes, i.e., *Haematocarpus validus* and *Prunus nepalensis* L (Sasikumar et al., 2021; Vivek et al., 2018). The average mass (*M*) of the whole matoa fruit is 15.31±2.44 g, and the average consumable portion available in the fruit, i.e., arillode, is 57.13±4.35%, while the rest is a non-consumable portion in the form of rind and seed.

Table 2
Physical parameters

No.	Physical Properties	N	Mean	Min	Max	SD
1	Mass *M* (g)	183	15.31	6.51	21.31	2.44
2	Length *L* (mm)	183	39.30	27.30	50.20	3.19
3	Width *W* (mm)	183	29.35	24.90	32.60	1.41
4	Thickness *T* (mm)	183	28.72	24.40	31.70	1.38
5	Arithmetic mean diameter D_a (mm)	183	32.46	27.57	37.30	1.78
6	Geometric mean diameter D_g (mm)	183	32.10	27.48	36.29	1.69
7	Equivalent mean diameter D_e (mm)	183	32.10	27.48	36.29	1.69
8	Projected area perpendicular to length *PA-L* (mm^2)	183	908.03	602.50	1226.18	105.95
9	Projected area perpendicular to width *PA-W* (mm^2)	183	678.01	486.95	834.69	64.54
10	Projected area perpendicular to thickness *PA-T* (mm^2)	183	663.48	479.13	811.65	62.19
11	Criteria projected area *CPA* (mm^2)	183	749.84	543.67	911.09	74.12
12	Surface area (mm^2)	183	3246.29	2371.63	4136.25	339.35
13	Aspect ratio (decimal)	183	0.75	0.61	0.90	0.05
14	Sphericity (decimal)	183	0.82	0.71	0.92	0.03
15	Eccentricity (decimal)	183	0.66	0.44	0.79	0.05
16	Radius of curvature minimum R_{min} (mm)	183	17.16	13.85	20.33	1.05
17	Radius of curvature maximum R_{max} (mm)	183	22.79	17.21	28.07	1.61
18	Elongation ratio (decimal)	183	1.34	1.11	1.64	0.08
19	Flakiness ratio (decimal)	183	0.98	0.88	0.99	0.02
20	True volume (mm^3)	183	17628.42	10500	24500	2751.85
21	Ellipsoid volume (mm^3)	183	17463.31	10860.33	25014.06	2720.40
22	Prolate spheroid volume (mm^3)	183	24003.30	10965.55	41036.13	4691.35
23	Oblate spheroid volume (mm^3)	183	17848.64	11037.64	25422.78	2817.40
24	True density (g/cm^3)	183	0.87	0.48	1.22	0.08
25	Bulk density (g/cm^3)	-	0.54	0.53	0.55	0.02
26	Porosity (%)	-	37.90	36.45	39.36	2.05

Table 2 (*continued*)

No.	Physical Properties	N	Mean	Min	Max	SD
27	Fruit part fraction: rind (%)	183	22.30	12.30	52.17	6.27
28	Fruit part fraction: arillode (%)	183	57.13	39.13	68.95	4.35
29	Fruit part fraction: seed (%)	183	20.58	1.83	33.76	4.75
30	Rind thickness (mm)	60	1.43	0.55	3.15	0.52
31	Coefficient of friction					
	Plywood (axial)	30	0.318	0.158	0.488	0.085
	Plywood (radial)	30	0.104	0.070	0.141	0.023
	Galvanized iron sheet (axial)	30	0.348	0.141	0.625	0.110
	Galvanized iron sheet (radial)	30	0.118	0.052	0.194	0.038
	Stainless steel sheet (axial)	30	0.253	0.123	0.554	0.093
	Stainless steel sheet (radial)	30	0.083	0.052	0.141	0.024
	Rubber sheet (axial)	30	0.304	0.087	0.466	0.093
	Rubber sheet (radial)	30	0.087	0.052	0.123	0.022
	Cardboard (axial)	30	0.272	0.158	0.424	0.085
	Cardboard (radial)	30	0.107	0.052	0.213	0.041

A clear representation of the distribution of length (L), width (W), thickness (T), mass (M), and true volume (V) of matoa fruit is depicted in Figure 2. According to Figure 2, the range of approximately 38-41 mm of fruit length is the most frequent length range for matoa fruit (41%). The fruit length also exhibits the highest peak frequency, suggesting that matoa fruits are more uniform in length compared to other parameters. The most frequent width (75%) and thickness (73%) of matoa fruit are falls between approximately 18-31 mm, showing a more symmetrical and narrow distribution compared to length. This alignment of Figure 2 implies that the fruit's mass correlates closely with its volume, reflecting that matoa have a consistent density across the samples. All parameters demonstrate a bell-shaped distribution, with most samples clustering around the central ranges, reflecting natural variability in matoa fruit, which is common in agricultural products.

The mean values for D_a, D_g, and D_e were 32.46±1.78 mm, 32.10±1.69 mm, and 32.10±1.69 mm, respectively. The mentioned average diameters are the basic data needed to determine the aperture size of the machine for segregation and grading based on the dimensions of the fruit (Pathak et al., 2019). The average area parameters of the fruit, namely SA, PA_L, PA_W, PA_T, and CPA, were found to be 3246.29±339.35 mm^2, 908.03±105.95 mm^2, 678.01±64.54 mm^2, 663.48±62.19 mm^2, and 749.84±74.12 mm^2. All mentioned area parameters of the fruit are beneficial in designing sizing systems, dryers and drying systems, storage systems, various aspects of postharvest, and processing of the fruit (Pathak et al., 2019; Pathak et al., 2020).

The shape of the matoa fruit was depicted with its sphericity ($Ø$), aspect ratio (A_R), flakiness ratio (F_R), elongation ratio (E_R), and eccentricity (e) observed as 0.82±0.03,

0.75±0.05, 0.98±0.02, 1.34±0.08, and 0.66±0.05, respectively. According to the obtained data, it indicates that the shape of the fruit seems to exhibit a shape that is closer to an ellipsoid than a perfect sphere due to the higher value of the elongation ratio (E_R). The sphericity ($\varnothing$) and aspect ratio (A_R) of the fruit are responsible for the flow behaviour of the materials, i.e., the higher the sphericity ($\varnothing$), the higher the flowability of the fruit. The flakiness ratio (F_R) and elongation ratio (E_R) of the matoa fruit indicate that the shape of the fruit is not flaked and slightly elongated. All the mentioned shape properties of the matoa fruit indicate that the matoa fruit will roll on the specific test surface and will be a major consideration for designing hoppers. The rolling ability can also be estimated by using the radius of curvature of the matoa fruit, which is very critical for designing conveyors and chutes, i.e., 17.16±1.05 mm of R_{min} and 22.79±1.61 mm of R_{max} for matoa fruit (Pathak et al., 2019; Pathak et al., 2020).

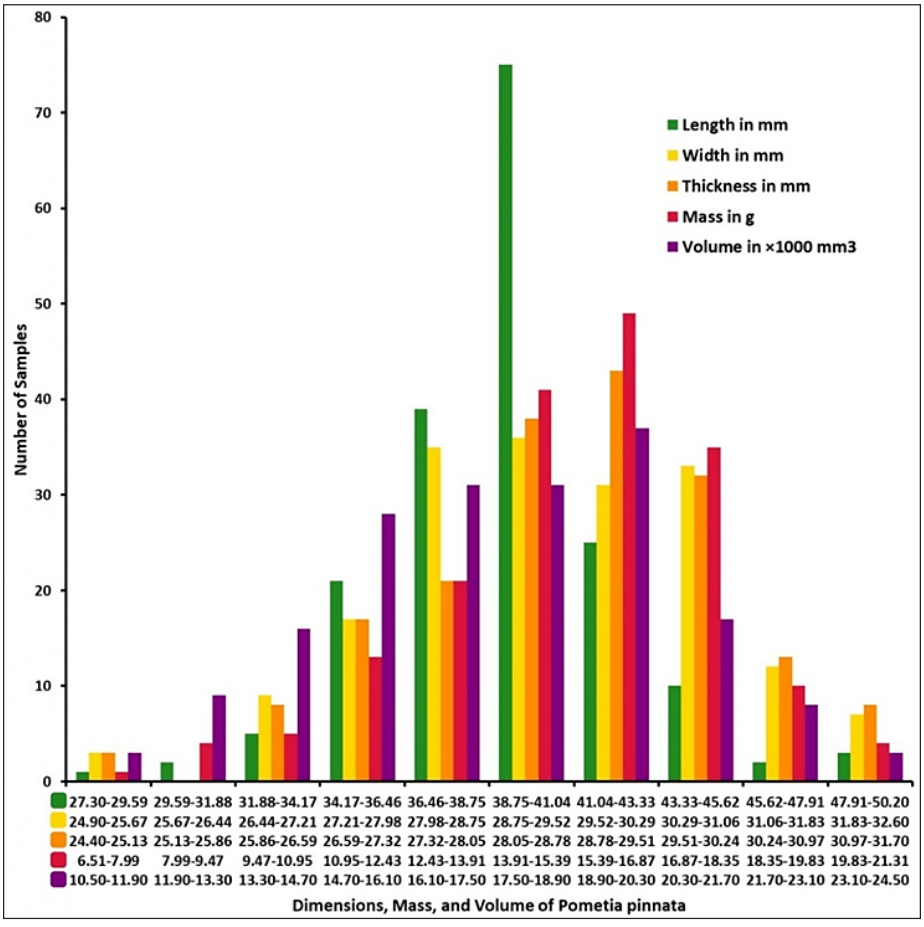


Figure 2. Distribution of length, width, thickness, mass, and true volume

The average volumes for matoa fruit, namely true volume (V), ellipsoid volume (V_{ellip}), prolate spheroid volume (V_{pro}), and oblate spheroid volume (V_{osp}), were recorded as 17628.42±2751.85 mm^3, 17463.31±2720.40 mm^3, 24003.30±4691.35 mm^3, and 17848.64±2817.40 mm^3, respectively. The volume parameters of the fruit are the main factors in calculating the density parameters of the fruit, i.e., bulk density (ρ_B), true density (ρ_T), and porosity (ε). The respective values of average bulk density (ρ_B), true density (ρ_T), and porosity or void fraction (ε) was 0.54±0.02 g/cm3, 0.87±0.08 g/cm3, and 37.90±2.05%. Both volume and density parameters are crucial for designing storage systems such as bins, silos, and storage rooms (Pathak et al., 2019). Both are also very critical inputs to designing efficient packaging, e.g., packaging materials, packaging size, packaging formation, and packaging configuration (Birania et al., 2022; Mahawar et al., 2019). In addition, the volume and density parameters are critical to ensuring the optimum transportation system of the fruit (Sasikumar et al., 2021).

The results also showed that for all tested surfaces, the static friction in the axial orientation was consistently higher than that in the radial orientation. Similar phenomena have been reported previously on Kendu fruit (*Diospyros melanoxylon* Roxb.) (Panda et al., 2020). The obtained average values in the axial direction of the coefficient of friction, from largest to smallest, were galvanised iron sheet, plywood, rubber sheet, cardboard, and stainless-steel sheet, while the coefficient of friction in the radial direction was galvanised iron sheet, cardboard, plywood, rubber sheet, and stainless-steel sheet. The friction coefficients obtained serve as important reference data for the design of conveyors, storage bins, and hopper systems. Similarly, various material handling tools and equipment designs, e.g., chutes and conveyor belts, are completely dependent on the angle of friction to specify their angle of inclination (Vivek et al., 2018).

Thermal Properties

The thermal characteristics of the edible arillode of matoa fruit are summarised in Table 3. The thermal properties of the edible arillode of matoa fruit, namely specific heat capacity (Cp), thermal conductivity (K), thermal diffusivity (μ), and latent heat (λ), were relatively high thanks to the high moisture content, i.e., 73.47±2.50% (w.b.). Thermal properties are beneficial data in food processing and storage, such as fruits, when the contained water in the product is highly responsible (Sasikumar et al., 2021). The thermal conductivity of the edible part of matoa fruit was relatively low (0.51 J/ms °C). This suggests that the edible material has a relatively low capacity to conduct heat, which is important when estimating heat flux during thermal processing. The thermal diffusivity of the arillode was relatively high (1.67×10^{-4} m^2/s), likely due to its high moisture content, which enables heat to move rapidly through the material relative to its bulk capacity. The obtained thermal properties of edible arillode of matoa fruit, i.e., specific heat capacity (Cp), thermal conductivity (K), and

Table 3
Thermal properties

No.	Physical Properties	N	Mean	Min	Max	SD
1	Specific heat capacity *Cp* (kJ/kg °C)	17	3.512	3.456	3.617	0.063
2	Thermal conductivity *K* (J/ms °C)	17	0.510	0.499	0.531	0.012
3	Thermal diffusivity μ ($\times 10^{-4}$ m^2/s)	17	1.670	1.660	1.687	0.010
4	Latent heat of fusion λ (kJ/kg)	17	2.461	2.067	2.832	0.329

thermal diffusivity (μ), were relatively on par with previous reports of fruits, e.g. sohiong fruit (Vivek et al., 2018), cashew apple fruit (Singh et al., 2019), bael fruit (Sonawane et al., 2020), green plantain (Kamble et al., 2021), blood fruit (Sasikumar et al., 2021), and wood apple (Murakonda et al., 2022).

Mass Modelling of the Matoa Fruit

The mathematical models developed to estimate the mass of matoa fruit were analysed based on its dimensions, surface area, and volume. A total of 76 predictive models were evaluated for estimating the mass of matoa fruit, i.e., 40 models for dimensions-based, 20 models for surface area-based, and 16 models for volume-based. The model considered the most suitable when resulting in the highest R^2, as well as the lowest *RMSE* and χ^2. The results showed that the range of R^2 values ranged from 0.423 to 0.915, from 0.668 to 0.843, and from 0.536 to 0.838 for dimension, surface area, and volume-based models, respectively. Table 4 presents the ten best-performing models, which achieved the highest R^2 and the lowest *RMSE* and χ^2 values.

According to Table 4, the dimension-based models are superior to surface area and volume-based models. Based on dimensional properties, the quadratic model Equation

Table 4
The selected ten models with the highest R^2 and the lowest RMSE and χ^2 values

Model-based	Dependent	Independent	Variables	Model	R^2	RMSE	χ^2
Dimension	M	L, W, T	Multivariate	Quadratic	0.915	0.610	0.400
Dimension	M	L, W	Multivariate	Quadratic	0.899	0.660	0.460
Dimension	M	L, W, T	Multivariate	S-curve	0.857	0.790	0.640
Dimension	M	L, W	Multivariate	S-curve	0.852	0.800	0.660
Dimension	M	W, T	Multivariate	Quadratic	0.850	0.810	0.680
Area	M	CPA	Univariate	Quadratic	0.843	0.830	0.700
Dimension	M	L, W, T	Multivariate	Linear	0.842	0.830	0.710
Dimension	M	L, W	Multivariate	Linear	0.838	0.840	0.720
Volume	M	V_{osp}	Univariate	Quadratic	0.838	0.840	0.720
Dimension	M	L, W, T	Multivariate	Power	0.836	0.850	0.730

15 with multiple variables of length (L), width (W), and thickness (T) resulted in the best fit, with 0.915, 0.610, and 0.400 of R^2, *RMSE*, and χ^2 values, respectively. The quadratic model also gives the best fit for the surface area and volume-based models, with the criteria projected area (*CPA*) and volume of the oblate spheroid (V_{osp}) as single independent variables, respectively. The surface area-based model Equation 16 and volume-based model Equation 17 have 0.843 and 0.836, 0.830 and 0.850, and 0.700 and 0.730 for their respective R^2, *RMSE*, and χ^2 values. The selected mass models for matoa fruit based on its dimension, surface area, and volume are shown in Figure 3.

$$M = -84.3474 + 0.1897T + 4.0146W + 0.8720L + 0.0667TL \\ -0.5194TW + 0.0184WT + 0.2116T^2 + 0.1984W^2 - 0.0399L^2 \Rightarrow R^2 = 0.915 \quad [15]$$

$$M = -40.2497 + 0.1217CPA - 6.2726 \times 10^{-5} CPA^2 \Rightarrow R^2 = 0.843 \quad [16]$$

$$M = -17.9169 + 0.0030V_{osp} - 6.4349 \times 10^{-8} V_{osp}^2 \Rightarrow R^2 = 0.838 \quad [17]$$

The previous reports for relatively similar shapes of matoa, i.e., *Haematocarpus validus* and *Prunus nepalensis* L, also indicated the superiority of the quadratic model. For the majority of the mass group of sohiong fruit (*Prunus nepalensis* L), Vivek et al. (2018) recommended the quadratic model for dimension, area, and volume-based model, with length (L) ($R^2 = 0.920$), projected area perpendicular to length (PA_L) ($R^2 = 0.946$), and ellipsoid volume (V_{ellip}) ($R^2 = 0.945$) as independent variables, respectively. Sasikumar et al. (2021) reported that the mass of blood fruit (*Haematocarpus validus*) can be suitably

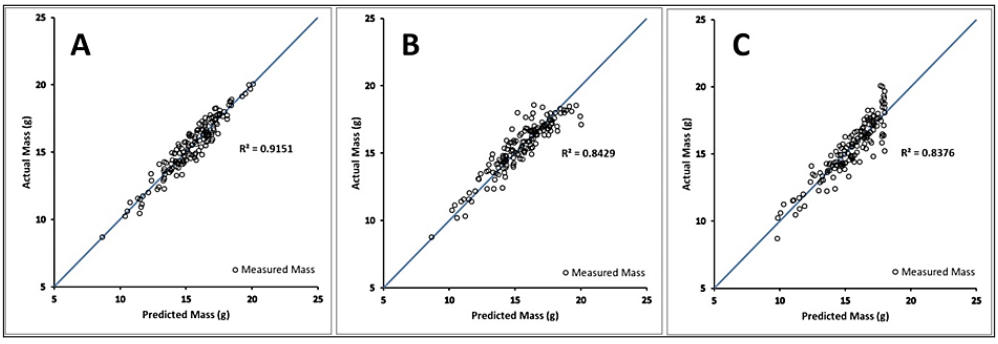


Figure 3. Matoa fruit mass model based on (A) '*LWT*' dimensions; (B) criteria projected area; and (C) volume of an oblate spheroid

predicted using its width (*W*) for dimension-based models in quadratic form ($R^2 = 0.984$), as well as using its criteria projected area (*CPA*) for area-based models in the same form of equation ($R^2 = 0.974$). However, for volume-based models, the linear model was recommended with ellipsoid volume (V_{ellip}) as an independent variable ($R^2 = 0.953$).

CONCLUSION

This study comprehensively quantified the engineering properties of matoa fruit to support the design of appropriate machinery and postharvest handling systems. Physical characterisation at the natural moisture content (37.55 ± 2.75% w.b.) indicated a fruit mass range of 6.51–21.31 g and mean principal dimensions of 39.30 ± 3.19 mm (length), 29.35 ± 1.41 mm (width), and 28.72 ± 1.38 mm (thickness). The calculated shape indices confirmed an ellipsoidal geometry. The static coefficient of friction was greatest on galvanised iron and lowest on stainless steel across both axial and radial orientations. The edible arillode exhibited low thermal conductivity (0.510 J/ms°C), reflecting limited heat-transfer capability. Among the predictive models evaluated, the quadratic model based on length-width-thickness (LWT) yielded the highest accuracy for mass estimation ($R^2 = 0.915$), while the criteria projected area (CPA) and oblate spheroid volume (Vosp) provided the most reliable surface-area- and volume-based predictions ($R^2 = 0.843$ and 0.838, respectively. Overall, the findings offer essential baseline data to facilitate informed machine design and optimised postharvest processing for matoa fruit.

ACKNOWLEDGEMENT

This research received financial support from the Faculty of Agricultural Technology, Universitas Brawijaya, in the framework of the BPPM funding scheme (SK: No. 40/2023 and contract number: 998/UN10.F10.06/TU/2023).

REFERENCES

Altuntaş, E., & Mahawar, M. K. (2022). Mass modelling of potato cultivars with different shape index by physical characteristics. *Journal of Food Process Engineering, 45*(10), e14126. https://doi.org/10.1111/jfpe.14126

Association of Official Analytical Chemists. (2000). *Official methods of analysis of the Association of Official Analytical Chemists*. AOAC.

Azman, P. N. M. A., Shamsudin, R., Che Man, H., & Ya'acob, M. E. (2020). Some physical properties and mass modelling of pepper berries (*Piper nigrum* L.), variety Kuching, at different maturity levels. *Processes, 8*(10), 1314. https://doi.org/10.3390/pr8101314

Barbhuiya, R. I., Nath, D., Singh, S. K., & Dwivedi, M. (2020). Mass modelling of Indian coffee plum (*Flacourtia jangomas*) fruit with its physicochemical properties. *International Journal of Fruit Science, 20*(sup3), S1110-S1133. https://doi.org/10.1080/15538362.2020.1775161

Bibwe, B., Mahawar, M. K., Jalgaonkar, K., Meena, V. S., & Kadam, D. M. (2022). Mass modelling of guava (cv. Allahabad safeda) fruit with selected dimensional attributes: Regression analysis approach. *Journal of Food Process Engineering, 45*(3), e13978. https://doi.org/10.1111/jfpe.13978

Birania, S., Attkan, A. K., Kumar, S., Kumar, N., & Singh, V. K. (2022). Mass modelling of strawberries (*Fragaria × ananassa*) based on selected physical attributes. *Journal of Food Process Engineering, 45*(5), e14023.https://doi.org/10.1111/jfpe.14023

Burley, A. L., Enright, N. J., & Mayfield, M. M. (2011). Demographic response and life history of traditional forest resource tree species in a tropical mosaic landscape in Papua New Guinea. *Forest Ecology and Management, 262*(5), 750-758. https://doi.org/10.1016/j.foreco.2011.05.008

Gade, S. R., Meghwal, M., & Prabhakar, P. K. (2020). Engineering properties of dried ash gourd (*Benincasa hispida* Cogn.) seeds: Mass modelling and its analysis. *Journal of Food Process Engineering, 43*(12), e13545.https://doi.org/10.1111/jfpe.13545

Hanafi, H., Irawan, C., Sirait, S. M., Sulistiawaty, L., & Setyawati, S. R. (2020). Toxicity test with BSLT (brine shrimp lethality test) method on methanol, ethyl acetate extract, hexane on seeds, and rind of matoa extract (*Pometia pinnata*). *Oriental Journal of Chemistry, 36*(6), 1143-1147. https://doi.org/10.13005/ojc/360618

Irawan, C., Hanafi, L. S., Rochaeni, H., & Lestari, P. S. (2017a). Comparison of total phenolic content in seed, flesh fruit and peel of *Pometia pinnata* from Indonesia. *Journal of Medicinal Plants Studies, 5*(4), 163-165. https://doi.org/10.22270/jmpas.v6i4.316

Irawan, C., Sulistiawaty, L., Rochaeni, H., & Lestari, P. S. (2017b). Evaluation of DPPH free radical scavenging activity of *Pometia pinnata* from Indonesia. *Pharma Innovation, 6*(8), 403-406. https://doi.org/10.22271/tpi.2017.v6.i8f.1111

Jacobs, M. (1962). *Pometia* (Sapindaceae), a study in variability. *Reinwardtia, 6*(2), 109-144.

Kamble, M. G., Singh, A., Mishra, V., Meghwal, M., & Prabhakar, P. K. (2021). Mass and surface modelling of green plantain banana fruit based on physical characteristics. *Computers and Electronics in Agriculture, 186*, 106194. https://doi.org/10.1016/j.compag.2021.106194

Lim, T. K. (2013). *Pometia pinnata*. In *Edible medicinal and non-medicinal plants*. Springer. https://doi.org/10.1007/978-94-007-5628-1_15

Mahawar, M. K., Bibwe, B., Jalgaonkar, K., & Ghodki, B. M. (2019). Mass modelling of kinnow mandarin based on some physical attributes. *Journal of Food Process Engineering, 42*(5), e13079. https://doi.org/10.1111/jfpe.13079

Matra, D. D., Ritonga, A. W., Natawijaya, A., Poerwanto, R., Sobir, Siregar, U. J., Widodo, W. D., & Inoue, E. (2019). Datasets for genome assembly of six underutilised Indonesian fruits. *Data in Brief, 22*, 960-963.

Murakonda, S., Patel, G., & Dwivedi, M. (2022). Characterisation of engineering properties and modelling mass and fruit fraction of wood apple (*Limonia acidissima*) fruit for post-harvest processing. *Journal of the Saudi Society of Agricultural Sciences, 21*(4), 267-277. https://doi.org/10.1016/j.jssas.2021.09.005

Panda, G., Vivek, K., & Mishra, S. (2020). Physical characterisation and mass modelling of kendu (*Diospyros melanoxylon* Roxb.) fruit. *International Journal of Fruit Science, 20*(sup3), S2005-S2017.https://doi.org/10.1080/15538362.2020.1851339

Panda, T. C., Thota, N., Dwivedi, M., Pradhan, R. C., & Seth, D. (2022). Mass modelling of engineering properties and characterisation of kadamb fruit (*Neolamarckia cadamba*): An underutilised fruit. *Journal of Food Process Engineering, 45*(11), e14160. https://doi.org/10.1111/jfpe.14160

Pathak, S. S., Pradhan, R. C., & Mishra, S. (2019). Physical characterisation and mass modelling of dried *Terminalia chebula* fruit. *Journal of Food Process Engineering, 42*(3), e12992. https://doi.org/10.1111/jfpe.12992

Pathak, S. S., Pradhan, R. C., & Mishra, S. (2020). Mass modelling of belleric myrobalan and its physical characterisation in relation to post-harvest processing and machine designing. *Journal of Food Science and Technology, 57*, 1290-1300. https://doi.org/10.1007/s13197-019-04162-1

Sahin, S., & Sumnu, S. G. (2007). *Physical properties of foods*. Springer.

Saini, P., Panghal, A., Mittal, V., & Gupta, R. (2022). Hempseed (*Cannabis sativa* L.) bulk mass modelling based on engineering properties. *Journal of Food Process Engineering, 45*(1), e13929.https://doi.org/10.1111/jfpe.13929

Sasikumar, R., Vivek, K., Chakkaravarthi, S., & Deka, S. C. (2021). Physicochemical characterisation and mass modelling of blood fruit (*Haematocarpus validus*)-An underutilised fruit of northeastern India. *International Journal of Fruit Science, 21*(1), 12-25. https://doi.org/10.1080/15538362.2020.1848752

Singh, S. S., Abdullah, S., Pradhan, R. C., & Mishra, S. (2019). Physical, chemical, textural, and thermal properties of cashew apple fruit. *Journal of Food Process Engineering, 42*(5), e13094.https://doi.org/10.1111/jfpe.13094

Singh, H., & Meghwal, M. (2020). Physical and thermal properties of various ajwain (*Trachyspermum ammi* L.) seed varieties as a function of moisture content. *Journal of Food Process Engineering, 43*(2), e13310. https://doi.org/10.1111/jfpe.13310

Sonawane, A., Pathak, S. S., & Pradhan, R. C. (2020). Physical, thermal, and mechanical properties of bael fruit. *Journal of Food Process Engineering, 43*(6), e13393. https://doi.org/10.1111/jfpe.13393

Suzuki, T., Nagata, M., Kagawa, N., Takano, S., & Nomura, J. (2021). Anti-obesity effects of matoa (*Pometia pinnata*) fruit peel powder in high-fat diet-fed rats. *Molecules, 26*(21), 6733.https://doi.org/10.3390/molecules26216733

Sweat, V. E. (1974). Experimental values of thermal conductivity of selected fruits and vegetables. *Journal of Food Science, 39*(6), 1080-1083.https://doi.org/10.1111/j.1365-2621.1974.tb07323.x

Tomar, M. S., & Pradhan, R. C. (2022). Prediction of mass-based process designing parameters of Amla fruit using different modelling techniques. *Journal of Food Process Engineering, 45*(8), e14039. https://doi.org/10.1111/jfpe.14039

Thomas, A., Pulissery, S. K., Vithu, P., & Abdullah, S. (2025). Assessment of engineering and physicochemical properties of gac fruit and predictive mass modelling for physical attributes. *Journal of Food Measurement and Characterisation, 19*, 5394-5410. https://doi.org/10.1007/s11694-025-03304-8

Vivek, K., Mishra, S., & Pradhan, R. C. (2018). Physicochemical characterisation and mass modelling of sohiong (*Prunus nepalensis* L.) fruit. *Journal of Food Measurement and Characterisation, 12*, 923-936. https://doi.org/10.1007/s11694-017-9708-x

Zulfahmi, Z., & Rosmaina, R. (2024). Phenotype variability of *Pometia pinnata* from Riau, Indonesia, based on qualitative characters. *Biodiversitas Journal of Biological Diversity, 25*(10).https://doi.org/10.13057/biodiv/d251055

Pertanika J. Sci. & Technol. 34 (2): 1181 - 1202 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

FCA_Unet: An Enhanced UNet with Feature Coordination Attention for Wheat Leaf Disease Severity Assessment

Hongyan Zang[1,2], Annie Anak Joseph[1]*, Kuryati binti Kipli[1], David Bong Boon Liang[1], Wanzhen Wang[1,2], and Rong Liu[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]College of Computer and Information Engineering, Qilu Institute of Technology, 250200 Jinan, China

ABSTRACT

Accurate monitoring of wheat leaf disease severity, critical for crop protection and precision agriculture, has attracted extensive attention in the context of smart agriculture. Existing methods face notable limitations: manual assessment is inefficient and subjective; classification-based approaches lack generalisability; segmentation models struggle with poor small-lesion detection, low boundary accuracy, and high parameter counts. To tackle these issues, we propose FCA_Unet, an enhanced UNet framework. The model consists of two key components: (1) the integration of the designed A_CBAM attention mechanism module and multi-scale feature fusion (MFFE) module. The attention module, based on deformable convolution, enhances focus on lesion-specific attributes while suppressing irrelevant background information. In contrast, the MFFE enables cross-layer fusion, integrating high-level semantic features for disease classification and low-level contextual features for boundary optimisation. (2) An optimised backbone network: The Inception multi-branch structure, coordinated attention mechanism (INCA), and MFFE are integrated into ResNet50 to reduce parameter redundancy while preserving feature extraction efficiency, thus meeting the lightweight deployment requirements of agricultural field equipment. Experimental results demonstrate that FCA_Unet achieves 89.85% mIoU on wheat stripe rust and powdery mildew, representing an 11.8% improvement over UNet, with a parameter count of 29.95M (68.2% of UNet's). Combined with disease severity indices, the model proves its superiority in both segmentation accuracy and severity assessment, providing robust support for automated wheat disease monitoring in practical agricultural scenarios.

Keywords: Attention mechanism, feature fusion, semantic segmentation, wheat leaf disease severity

ARTICLE INFO

Article history:
Received: 24 October 2025
Accepted: 19 December 2025
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.24

E-mail addresses:
22010336@siswa.unimas.my (Hongyan Zang)
jannie@unimas.my (Annie Anak Joseph)
kkuryati@unimas.my (Kuryati binti Kipli)
bbldavid@unimas.my (David Bong Boon Liang)
22010329@siswa.unimas.my (Wanzhen Wang)
1274657903@qq.com (Rong Liu)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Wheat is a staple crop critical to global food security and agricultural development (Hassan et al., 2024; Kaur et al., 2023). These diseases spread rapidly through leaf lesions, if not detected in time, they can cause yield losses of 10% -70% (Kumar et al., 2024; Kumar & Kukreja, 2024). Accurate segmentation of leaf lesions is essential for early warning and targeted management (Sharma & Sethi, 2024; Shi et al., 2024; Zhang & Zhang, 2022).

Traditional approaches rely on manual inspection, which is time-consuming, subjective, and unsuitable for large-scale monitoring (Mi et al., 2020; Yi et al., 2024). With advances in computer vision and deep learning, image-based segmentation has become a promising tool for monitoring plant disease (Badrinarayanan et al., 2017; Hassan et al., 2024). UNet and its variants are widely used in this field due to their encoder-decoder structure and strong performance in pixel-level annotation (Rai & Pahuja, 2023; Shwetha et al., 2024; Wang et al., 2023; Yi et al., 2023; Zhang & Zhang, 2022; Zhang et al., 2022). Yet, their application in real field conditions remains challenging. Images often suffer from uneven lighting and cluttered backgrounds, while lesion appearance varies across disease stages (Liu et al., 2024; Wang et al., 2023). Moreover, agricultural monitoring frequently relies on drones or portable devices with limited computational resources, making conventional deep models too heavy for real-time deployment (Shi et al., 2024).

Recent advances in attention mechanisms, multi-scale feature fusion, and lightweight designs address these issues (Chen et al., 2021; Mishra et al., 2024; Yang et al., 2024; Yi et al., 2023), but balancing high accuracy with low complexity for practical field use remains an ongoing challenge. In this study, we improved the UNet model to balance segmentation accuracy and efficiency for wheat leaf disease detection. The model incorporates:

1. An A_CBAM attention mechanism for precise lesion location under complex backgrounds.
2. A multi-scale feature fusion module (MFFE) for integrating deep and shallow multi-scale features.
3. A lightweight FCA_ResNet backbone that reduces parameters while preserving feature extraction capacity.

Experiments on stripe rust and powdery mildew datasets demonstrate that FCA_Unet achieves higher accuracy than conventional models while maintaining efficiency suitable for edge deployment. This work provides an effective framework for automated and precise field monitoring of wheat diseases.

The remainder of this paper is organised as follows. Section 2 reviews related work. Section 3 presents the data material and the FCA_Unet model. Section 4 describes the results and analysis. Section 5 reviews the discussion and conclusion.

RELATED WORKS

In the development of smart agriculture, accurate monitoring of leaf diseases, particularly the precise estimation of disease severity, is critical for safeguarding crop yield and quality. Early research mainly adopted image classification approaches. For example, Hayit et al. (2021) proposed the Yellow-Rust-Xception model, which classifies wheat stripe rust into six severity levels with an accuracy of 91%. A model introduced by Nigam et al. (2022) for stem rust severity estimation achieved training and testing accuracies of 98.41% and 96.42%, respectively, with lower computational costs. Similarly, a C-DenseNet model incorporating convolutional block attention modules (CBAM) attained 97.99% accuracy on a six-level stripe rust grading dataset (Mi et al., 2020). While such classification methods achieve high accuracy, they provide only coarse severity estimates rather than precise quantification.

Segmentation-based methods have therefore attracted increasing attention. Semantic segmentation (Li et al., 2022; Zhao et al., 2025), instance segmentation (Sharma et al., 2024), and convolutional neural networks (CNNs) have shown strong potential in disease diagnosis and precision agriculture. In these approaches, lesion areas are segmented to estimate pixel counts or lesion area, which are then mapped to severity levels. For instance, Gao et al. (2023) combined simple linear iterative clustering super pixel segmentation with random forest classification, achieving an accuracy of 93.22% in stripe rust severity estimation. UNet, as a representative encoder-decoder model, has become a dominant choice in leaf disease segmentation due to its efficiency in pixel-level labelling. In stripe rust studies, UNet achieved an IoU of 0.9563 (Hassan et al., 2024). Sharma et al. (2024) integrated PointRend segmentation with EfficientNet classification, reporting an accuracy of 99.43%. Kumar et al. (2024) proposed a hybrid approach (FERSPNET-50) combining panoptic segmentation and object detection, outperforming YOLOv4 and RetinaNet in precision for rust detection.

Attention mechanisms and feature fusion have further advanced segmentation performance. These methods enable models to capture small lesions and irregular lesion boundaries more effectively. For example, Wang et al. (2023) proposed MFBP-UNet with multi-scale feature extraction and dynamic sparse attention, achieving a mean IoU of 86.15% in pear leaf disease segmentation. Rai and Pahuja (2023) improved UNet with attention gates, enhancing northern leaf blight segmentation in maize with 98.97% accuracy. Zhang et al. (2022) and Ge et al. (2023) employed an adaptive fusion module to complementarily integrate global and local attention features. This effectively reduces missed and false detections in multi-target scenarios. Zhang et al. (2023) apply the attention mechanism separately to two dimensions: semantic feature extraction and boundary detail optimisation. Via an attention-driven feature interaction module, the network achieves mutual enhancement between semantic information and boundary features.

Similarly, Yao et al. (2024) and Yi et al. (2024) introduced coordinate attention and dilated convolutions to boost lesion segmentation accuracy in apple and grape leaf diseases. These works demonstrate that combining multi-scale fusion and attention mechanisms enhances robustness in complex disease scenarios.

Lightweight model design has also emerged as a major trend to enable deployment on mobile and edge devices. For instance, WE-DeepLabV3+ employed MobileNetV2 with efficient channel attention, achieving an mIoU of 82.0% with only 5.1M parameters (Wang et al., 2024). Wang et al. (2021) introduced a lightweight UNet with MobileNetV3 and coordinate attention, reducing complexity while achieving 88.87% mIoU. Liu et al. (2024) developed StripeRustNet, a two-stage lightweight model for stripe rust, achieving 98.65% mIoU for leaf segmentation and 86.08% for lesion segmentation, while significantly reducing annotation time. Compared with large models, lightweight designs provide a better balance between accuracy and efficiency, making them more suitable for real-time agricultural monitoring in the field.

The analysis of existing research shows that in the study of various computer vision models, high accuracy and lightweight design are still the two most important indicators pursued by deep learning in practical applications. The balance between the two has always been essential to research content.

MATERIALS AND METHODS

Data

This study takes wheat disease as an example to illustrate the efficiency of our proposed method. The dataset consists of 652 images, including two parts. One part comes from a public wheat stripe rust dataset collected by the Ezhou Research Station of Huazhong Agricultural University in Hubei, China (Liu et al., 2024). The dataset consists of 5,013 high-resolution images with a resolution of 3,472 × 4,624 pixels (all images on a uniform background). There are 542 segmented images in the dataset. Each image is labelled as background, leaf, and diseased spot. Another part comes from a custom dataset captured under field conditions. 110 images of wheat powdery mildew are selected. These images are annotated using LabelMe, also according to the categories of background, leaf, and diseased spot. Image examples are shown in Figure 1. All 652 samples are randomly divided into a training set, validation set, and test set according to the ratio of 8:1:1. The specific division is shown in Table 1. The training set is used to learn and optimise the parameters of the model, the verification set is used to adjust the hyperparameters during the training process and monitor whether the model is overfitting, and the test set is used to evaluate the performance of the model in the final, ensuring the objectivity and reliability of the evaluation results.



Figure 1. Six samples of wheat leaf disease datasets: stripe rust (Liu et al., 2024) and powdery mildew (Jiang et al., 2022)

Table 1
Statistics of the dataset

Dataset	Training set	Validation set	Testing set	Total
Stripe rust	434	54	54	542
Powdery mildew	88	11	11	110
Total	522	349	179	652

Proposed FCA_Unet Model

This section introduces FCA_Unet, an enhanced variant of the traditional UNet designed to improve segmentation performance on wheat leaf diseases. FCA_Unet adopts an encoder-decoder architecture based on UNet. As illustrated in Figure 2, the encoder is composed of the improved FCA_ResNet, derived from ResNet50, together with the attention mechanism A_CBAM, while the decoder consists of up-sampling, deconvolution, and the feature fusion module MFFE. After image resizing, inputs are fed into the backbone FCA_ResNet. Feature maps from four different stages are concatenated with the up-sampled features through A_CBAM. By means of up-sampling, the decoder restores the feature maps extracted by the encoder to the resolution of the input image. Finally, decision-level fusion is achieved through the MFFE module, producing accurate segmentation results.

The overall structure of the INCA module is shown in Figure 3. The coordinate attention (CA) mechanism is a lightweight channel attention approach that overcomes the spatial feature limitations of SE attention and the long-range dependency shortcomings of CBAM. Given the widespread distribution of wheat disease images, a multi-scale feature extraction method is required to capture long-range dependencies. FCA_ResNet (Zang et al., 2025) integrates CA into multi-branch convolutional blocks to construct the INCA module, which effectively emphasises multi-scale information and lesion features.

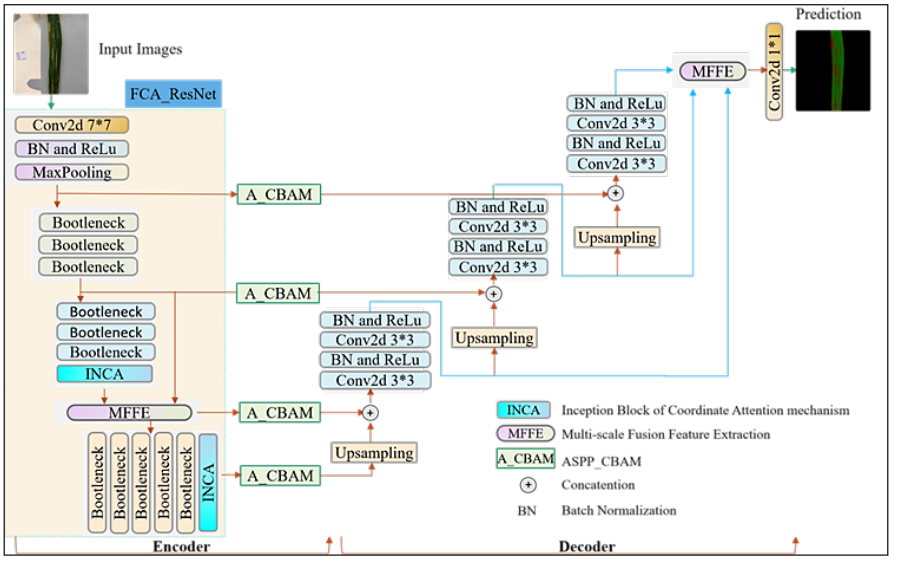


Figure 2. Structure overview of the proposed FCA_Unet segmentation model

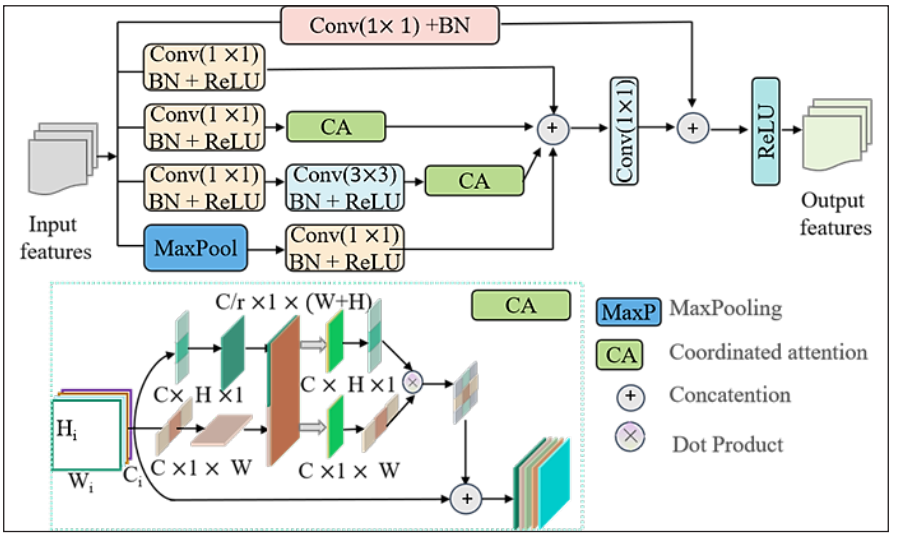


Figure 3. Structure diagram of the INCA module

Attention Mechanism Deform ASPP CBM (A_CBAM)

In standard convolution, kernels have fixed sizes and shapes, which may work well for objects with regular structures. However, in complex scenarios such as the segmentation of irregular disease spots on wheat leaves, deformable convolution offers greater potential than standard convolution. In the A_CBAM module, the 3×3 standard convolution within the ASPP is replaced by deformable convolution.

In a conventional convolution operation, each point on the output feature map corresponds to a sampling region of size K×K on the input feature map. In deformable convolution, each of these K×K sampling points learns an offset. As offsets are represented by coordinates, one output requires learning 2×K×K parameters. For an output of size H×W, this results in 2×K×K×H×W parameters in total, represented as the offset field in Figure 4, where N = K×K.

By allowing kernel positions to adjust dynamically, deformable convolution adapts the convolution weights to the spatial locations of sampling points. This enables the model to better accommodate non-uniform feature distributions in images, thereby enhancing its ability to capture complex structures, an advantage particularly critical in tasks such as object detection and semantic segmentation.

Multi-scale Feature Fusion Module (MFFE)

In segmentation tasks, the objective is to predict the category of each pixel. This requires not only robust representation of global target features but also precise preservation of boundary details, placing high demands on fine-grained feature extraction. Although data from different sources may contain redundancy, they also provide complementary information. In Figure 5, the MFFE module performs decision-level outputs from three different decoding stages, which enhances both the accuracy and the detail integrity of lesion segmentation. It further optimises the segmentation of small lesions and boundary regions.

The following section outlines the experimental results. It begins with providing information about the unique network configurations and related evaluation metrics. It then delves into conducting ablation experiments and comparative studies to analyse the performance of the proposed FCA_Unet segmentation model.

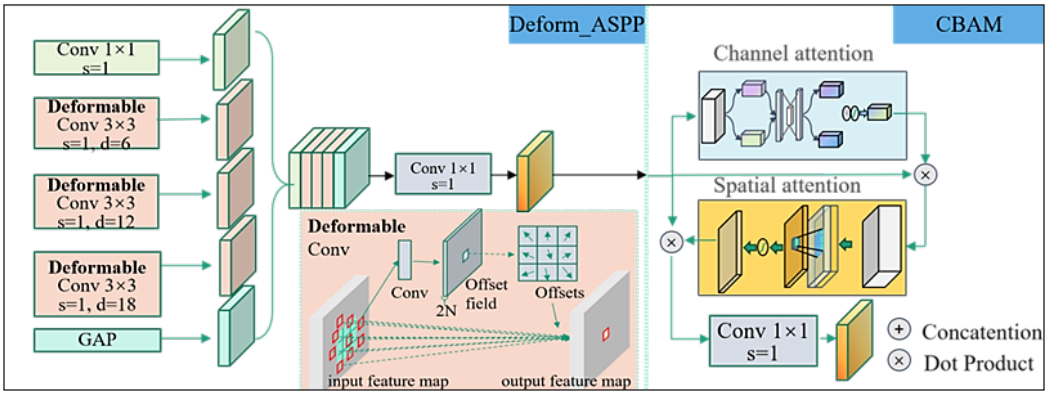


Figure 4. Structure diagram of A_CBAM module

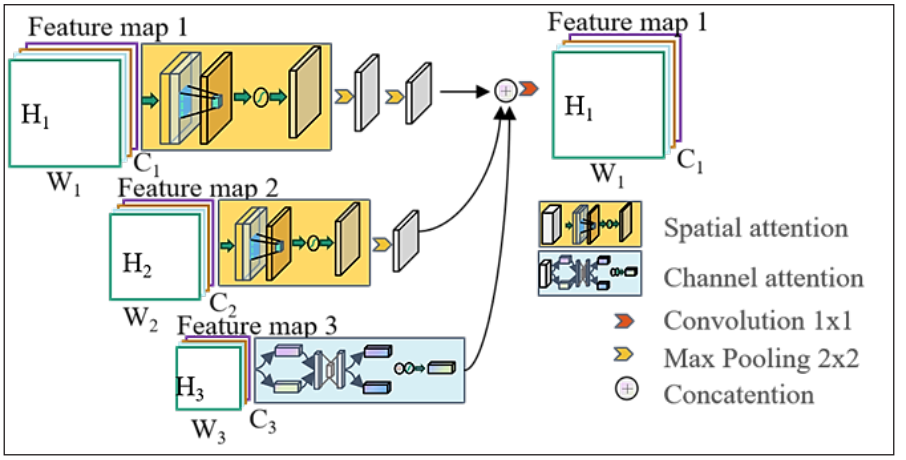


Figure 5. Structure diagram of the MFFE module

Evaluation Metrics

In the experiment, to quantitatively evaluate the performance of the proposed method and other comparison methods, mean pixel accuracy (mPA), mean intersection over union (mIoU), Disease Severity (DS), optimal model comprehensive score (OMCS), and other evaluation metrics were employed to assess the segmentation performance of the proposed model on wheat stripe rust.

Assuming there is a total of k classes (including a background class), P_{ij} denotes the number of pixels belonging to class i that are predicted as class j. Thus, P_{ii} represents the true positives (TP) and true negatives (TN), while P_{ij} and P_{ji} correspond to the false positives (FP) and false negatives (FN), respectively.

mPA

Pixel accuracy (PA) represents the ratio of correctly predicted pixels to the total number of pixels. Its calculation formula is given in Equation 1: -

$$\mathrm{PA}=\frac{\sum_{i=1}^{k} P_{ii}}{\sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}}=\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad [1]$$

mPA is defined as the mean value of the ratio between the number of correctly predicted pixels for each class and the total number of pixels of that class. The calculation formula is as shown in Equation 2: -

$$\mathrm{mPA}=\frac{1}{k+1}\sum_{i=1}^{k}\frac{P_{ii}}{\sum_{j=1}^{k} P_{ij}} \quad [2]$$

mIoU

The intersection over union (IoU) is the ratio of the intersection between the real and predicted sets to the union of the real and predicted sets for each category. The mIoU is a widely used evaluation metric in experimental studies of semantic segmentation and is calculated based on the average of the ratio of intersection and union of all categories. The calculation formulas are as shown in Equations 3 and 4: -

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP+FP+FN}} \quad [3]$$

$$\mathrm{mIoU}=\frac{1}{k}\sum_{i=1}^{k}\frac{\mathrm{TP}(i)}{\mathrm{TP}(i)+\mathrm{FP}(i)+\mathrm{FN}(i)} \quad [4]$$

DS

The DS index is used to quantify disease severity. DS value is the ratio of leaf spot area to leaf area, which is used to measure the severity of wheat leaf disease, as shown in Equation 5. This helps to assess the health status of the leaves in time and take appropriate measures.

$$\mathrm{DS}=\frac{S_{\mathrm{Disease}}}{S_{\mathrm{leaf}}}\times 100\% \quad [5]$$

OMCS

This paper proposes an OMCS. The core objective of OMCS is to identify the optimal model among multiple candidates that achieves both a high mIoU and a predicted DS value close to the ground-truth DS value.

For a single sample, the composite score of model *m* is defined as OMCS_m, as shown in Equation 6: -

$$\mathrm{OMCS}_m=\mathrm{NormmIoU}_m+(1-\mathrm{NormDS}_m) \quad [6]$$

Where: $\mathrm{NormmIoU}_m$ (as shown in Equation 7) denotes the normalised mIoU of model *m* (range: 0-1), and NormDS_m (as shown in Equation 8) represents the normalised DS of model *m* (range: 0-1).

$$\mathrm{NormmIoU}_m = \frac{\mathrm{mIoU}_m - \min(\mathrm{mIoU})}{\max(\mathrm{mIoU}) - \min(\mathrm{mIoU}) + \varepsilon} \quad [7]$$

mIoU_m: mIoU of model *m* on the current sample.

$\max(\mathrm{mIoU}), \min(\mathrm{mIoU})$: maximum and minimum mIoU among all models on the current sample.

ε: a very small value (e.g., 10^{-8}) to prevent division by zero.

$$\mathrm{NormDS}_m = \frac{\mathrm{DS}_m - \min(\mathrm{DS})}{\max(\mathrm{DS}) - \min(\mathrm{DS}) + \varepsilon} \quad [8]$$

$\max(\mathrm{DS}), \min(\mathrm{DS})$: the highest and lowest DS among all models for the current sample.

DS_m: as shown in Equation 9, the relative error between the predicted severity DS of model *m* and the ground truth is suitable for scenarios where the true DS values of different samples vary significantly. In scenarios where the true DS values are relatively small, absolute error can be used instead, i.e., $|\mathrm{PredDS}_m - \mathrm{TrueDS}|$.

$$\mathrm{DS}_m = \frac{|\mathrm{PredDS}_m - \mathrm{TrueDS}|}{\mathrm{TrueDS} + \varepsilon} \quad [9]$$

PredDS_m, TrueDS: the predicted DS of model *m* and the ground-truth DS label of the sample.

ε:a very small value (e.g., 10^{-8}) to avoid division by zero.

For each test sample, compare the comprehensive scores of all models. The model with the highest score is considered the optimal model for that sample, and its score is the OMCS of the sample. So, the best model can be expressed as Equation 10.

$$\text{Best Model} = \arg\max(\mathrm{OMCS}_m) \quad [10]$$

In addition, the number of parameters and FLOPs are also important indicators to evaluate the lightness of the model. They are used to determine whether the algorithm can be deployed on mobile devices.

RESULTS AND ANALYSIS

Ablation Experiment

To assess the effectiveness of the attention mechanism A_CBAM and the feature fusion module MFFE, we conducted comparative experiments and ablation studies. These two modules were embedded into different structures of the base model, and the relevant experimental comparison results are presented in Table 2.

By embedding the A_CBAM attention mechanism and MFFE feature fusion module into the base model and setting up different module combination schemes (such as embedding A_CBAM alone, embedding MFFE alone, and embedding both simultaneously), the impact of each module on model performance can be intuitively compared. The experimental results in Table 2 reflect the practical effects of A_CBAM in optimising feature attention and MFFE in multi-scale feature fusion from the perspective of quantitative metrics (e.g., mIoU, mAP, model parameter count, and inference time), providing data support for verifying the effectiveness of the modules.

Table 2
Evaluation of the effectiveness of the attention mechanism and feature fusion module

Number	Method	CBAM	A_CBAM	MFFE	mIoU	mAP	Parameter (MB)	Inference Time (ms)
1	UNet	×	×	×	0.7805	0.9744	43.93	**8.99**
2	FCA_Unet_1	×	×	×	0.7956	0.9751	39.57	11.63
3	FCA_Unet_2	√	×	×	0.7923	0.9780	39.64	13.57
4	FCA_Unet_3	×	√	×	0.8067	0.9781	**29.91**	20.08
5	FCA_Unet_4	×	×	√	0.7952	0.9782	39.61	14.41
6	FCA_Unet_5	×	√	√	**0.8077**	**0.9794**	29.95	20.68
7	FCA_Unet_6	√	×	√	0.8036	0.9738	39.68	14.57

Comparison with Baseline Model

When compared to the baseline model UNet, FCA_Unet_1, devoid of additional modules, outperformed UNet in both mIoU (0.7956 vs. 0.7805) and mAP (0.9751 vs. 0.9744), while reducing parameters by 9.9% (39.57 MB vs. 43.93 MB). This highlights the inherent superiority of FCA_Unet's basic architecture over the original UNet.

Impact of Individual Modules on Performance

FCA_Unet_3 with CBAM exhibited a marginal decrease in mIoU (0.7923 vs. 0.7956) but a modest improvement in mAP (0.9780 vs. 0.9751) compared to FCA_Unet_1 without CBAM.

However, integrating CBAM slightly increased parameters (39.64 MB vs. 39.57 MB) and significantly prolonged inference time (13.57 ms vs. 11.63 ms). This suggests CBAM did not effectively enhance the core segmentation metric (mIoU) under current experimental conditions, instead introducing additional computational overhead.

FCA_Unet_4 with A_CBAM achieved the highest mIoU (0.8067) among single-module models, representing a 1.11% improvement over FCA_Unet_1, with a slight mAP increase (0.9781 vs. 0.9751). Notably, A_CBAM substantially reduced model parameters (by ~24.4%, 29.91 MB vs. 39.57 MB), demonstrating its lightweight design advantage. Inference time increased significantly (20.08 ms vs. 11.63 ms), likely due to the computational costs of the self-attention mechanism.

FCA_Unet_5 with MFFE showed comparable mIoU to FCA_Unet_1 (0.7952 vs. 0.7956) but higher mAP (0.9782 vs. 0.9751). Parameters remained nearly unchanged (39.61 MB vs. 39.57 MB), though inference time increased (14.41 ms vs. 11.63 ms), indicating MFFE can improve classification accuracy (mAP) without increasing model size at the cost of slightly higher computational load.

Impact of Module Combinations

FCA_Unet_6, combining A_CBAM and MFFE, achieved the highest mIoU (0.8077) and mAP (0.9794) across all models, representing modest improvements over standalone A_CBAM (mIoU +0.10%, mAP +0.13%) and standalone MFFE (mIoU +1.25%, mAP +0.12%). With 29.95 MB parameters (like Model 4) and only a marginal inference time increase, the computational overhead is justified by performance gains. These results suggest A_CBAM and MFFE are complementary: A_CBAM enhances feature attention, while MFFE strengthens multi-scale feature integration.

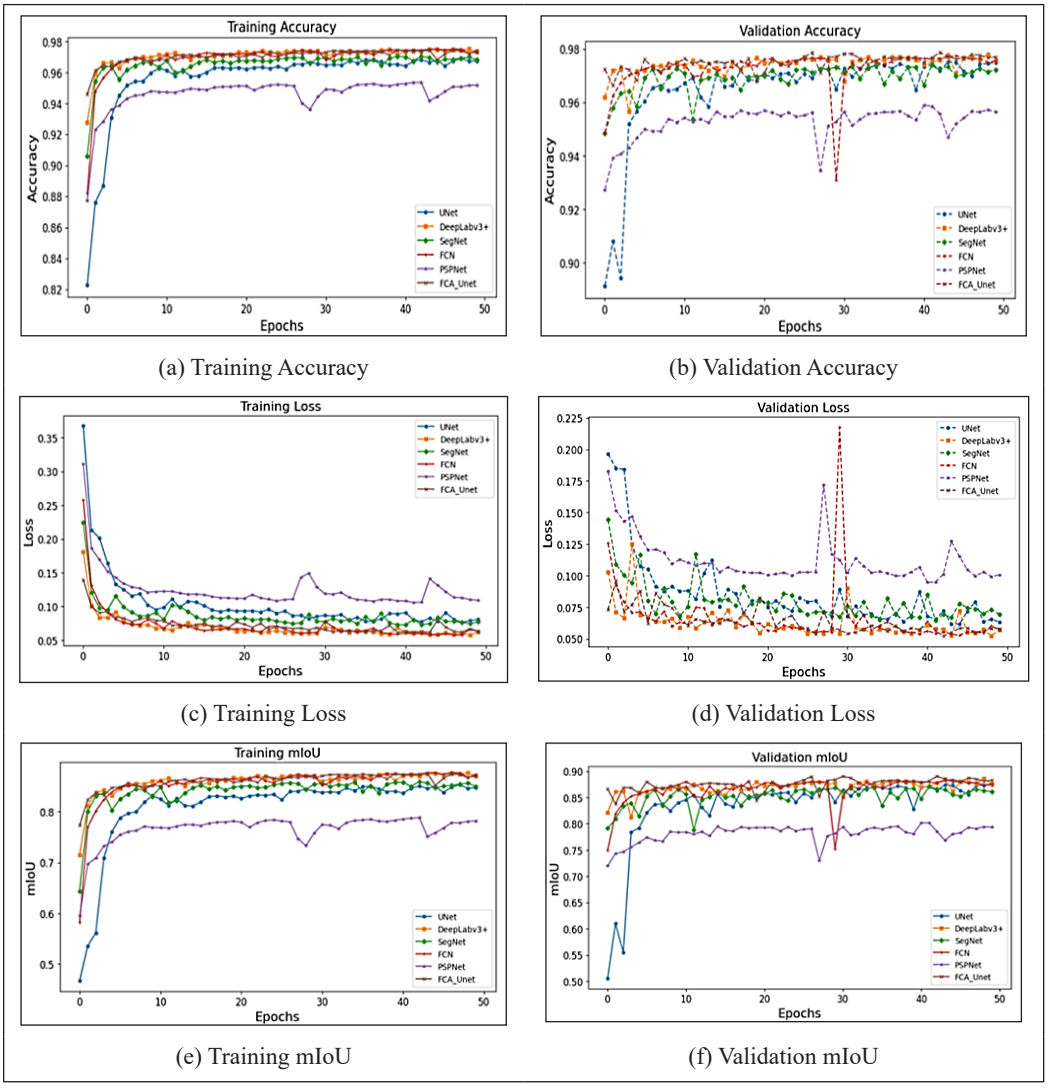
FCA_Unet_7, combining CBAM and MFE, achieved an mIoU of 0.8036, higher than standalone CBAM (0.7923) or MFFE (0.7952) but lower mIoU and mAP than the A_CBAM + MFFE combination. With parameters and inference times comparable to Models 3 and 5, this combination proved less effective in balancing performance and efficiency.

Experimental Comparison of Different Models

To demonstrate FCA_Unet's capability to capture feature information in wheat leaf disease images, four pre-trained segmentation models were compared. Comparative experiments used the same dataset under identical environmental conditions to ensure network performance comparability. Image size was fixed at 256×256 pixels for uniform processing, with a batch size of 4 and 50 epochs to avoid GPU memory issues and unstable training.

Training and Validation

This setup balanced efficiency and accuracy based on experimental results. A classic default learning rate of 0.001 was used to prevent slow convergence from low rates, suitable for different training stages. The Adam optimiser adjusted network parameters to minimise loss and accelerate convergence, facilitating deep learning model training. In (a) and (b) of Figure 6, the training accuracy and validation accuracy of FCA_Unet both remain around 97%. For the training loss and validation loss shown in (c) and (d), the losses of FCN and FCA_Unet are close, both reaching approximately 6%. In (e) and (f), FCA_Unet and FCN achieve the highest training mIoU, which can reach around 85%; however, in terms of validation mIoU, FCA_Unet and DeepLabv3+ perform the best, approaching 87%.



(a) Training Accuracy (b) Validation Accuracy

(c) Training Loss (d) Validation Loss

(e) Training mIoU (f) Validation mIoU

Figure 6. Comparative performance of different models in terms of accuracy, loss, and mIoU on the stripe rust training and validation sets

Testing

Differences in model performance can be visually assessed by randomly selecting several images for comparison. Comparative experiments between the proposed model and other approaches further confirm the efficacy and feasibility of the enhanced architecture. On the test set, the metrics recorded include mPA, mDice, mIoU, FLOPs, and model parameters. Here, mPA refers to the average pixel accuracy across all test images in the dataset.

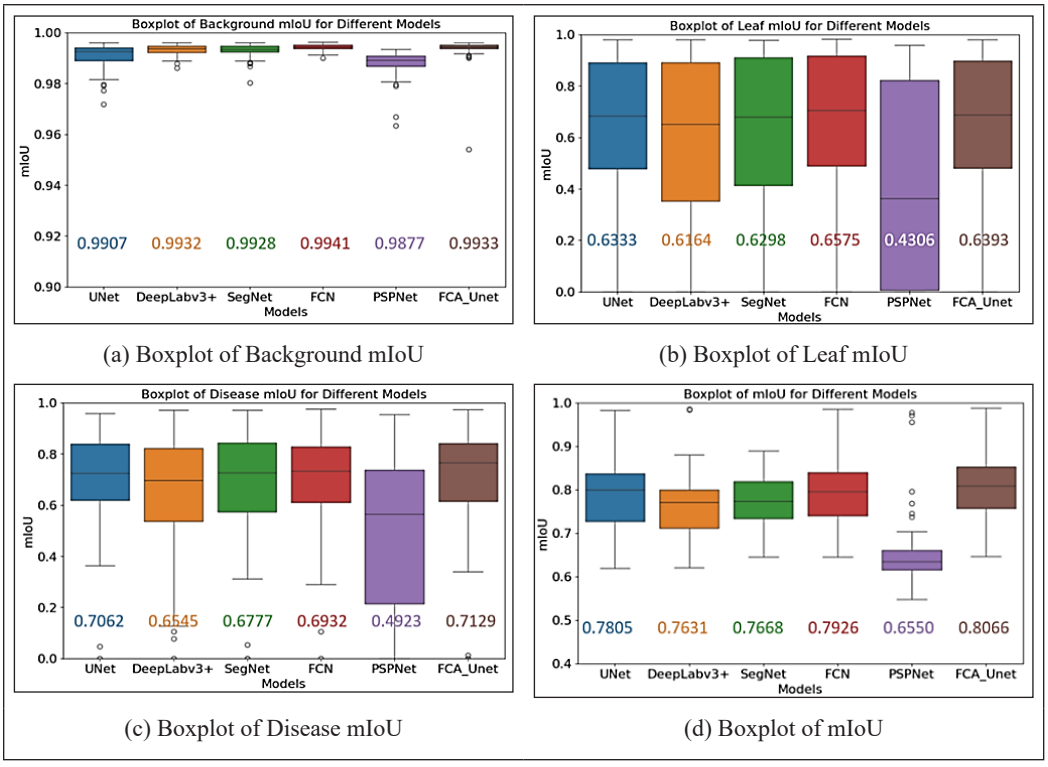
As Table 3 shows, FCA_Unet outperforms other models in mIoU, Dice, FLOPs, parameters, and best pixel accuracy. Notably, its top-1, top-5, and top-10 mIoU values across all tests are 89.85%, 88.54%, and 85.84%, respectively, demonstrating superior segmentation capability. The only exception is mAP, where FCA_Unet scores 97.57% second only to SegNet.

FCA_Unet's FLOPs reach 35.85G (55.69% higher than baseline UNet), but its parameter count is only 29.95M-68.18% of UNet's 43.93M, and comparable to FCN (23.52M), SegNet (29.44M), and PSPNet (15.18M). This efficiency stems from replacing UNet's ResNet50 backbone with FCA_ResNet, which has 10.76M parameters—57.9% fewer than ResNet50 (25.56M) (Zang et al., 2025).

Table 3
Performance comparison between FCA_Unet and mainstream segmentation models on the stripe rust test set

Model	Backbone	mPA	mIoU			FLOPs	Parameters	Best Pixel Accuracy
			Top 1	Top 5	Top 10			
UNet	ResNet50	97.44%	88.98%	85.81%	84.13%	23.02G	43.93M	99.56%
DeepLabv3+	ResNet50	97.42%	87.96%	82.21%	81.03%	45.77G	40.35M	99.65%
FCN	ResNet50	**97.76%**	89.60%	86.72%	84.20%	**5.55G**	23.52M	99.63%
SegNet	VGG16	97.59%	88.92%	83.56%	82.86%	40.14G	29.44M	99.59%
PSPNet	ResNet50	95.50%	88.19%	85.26%	83.20%	15.05G	**15.18M**	99.64%
FCA_Unet	ResNet_F_INCA	97.57%	**89.85%**	**88.54%**	**85.84%**	35.84G	29.95M	99.66%

Figure 7 compares IoU performance for wheat leaf disease spots. In background segmentation, FCN achieves the highest mean IoU (0.9941), followed by FCA_Unet and DeepLabv3+, with PSPNet lowest (0.9878). For leaf segmentation, PSPNet fluctuates significantly while FCA_Unet stays stable. FCA_Unet also leads in lesion segmentation. Overall, PSPNet is the weakest and least stable; FCN excels in background segmentation; SegNet shows consistent performance; and FCA_Unet delivers strong overall results with potential for greater stability.



(a) Boxplot of Background mIoU

(b) Boxplot of Leaf mIoU

(c) Boxplot of Disease mIoU

(d) Boxplot of mIoU

Figure 7. The performance of different models on the test set regarding the IoU and mIoU of the background, leaves, and disease spots

Visualisation

To provide a more intuitive comparison of segmentation results across different networks, we randomly selected wheat leaf disease images. The segmentation results for stripe rust and powdery mildew on the test set are shown in Figures 8 and 9, respectively, illustrating the performance of different network architectures. The first column displays the original wheat leaf images, the second column shows the manually annotated masks, and each subsequent column represents the outputs of different models. The mIoU and DS values for each sample under different models are summarised in Tables 4 and 5. As shown in Table 4, the proposed FCA_Unet consistently achieved the highest mIoU across both disease types, particularly on stripe rust samples.

Compared with other models, FCA_Unet produced more complete lesion segmentation, finer boundaries, and higher-quality masks. These gains likely result from its dilated convolutions and attention mechanisms, which capture contextual and small-lesion features, and from a lightweight fusion module in the decoder that refines boundaries. Its superior mIoU further confirms robustness and effectiveness in wheat leaf disease severity assessment.

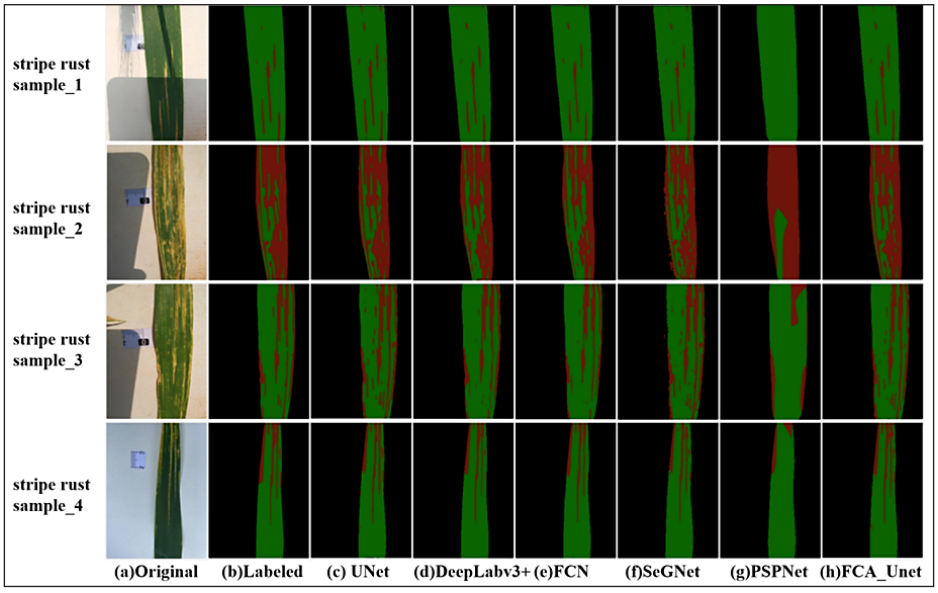


Figure 8. Example images of stripe rust demonstrating visual results obtained by segmentation models. Masks for background, normal leaf and disease spots are represented by black, green and red, respectively

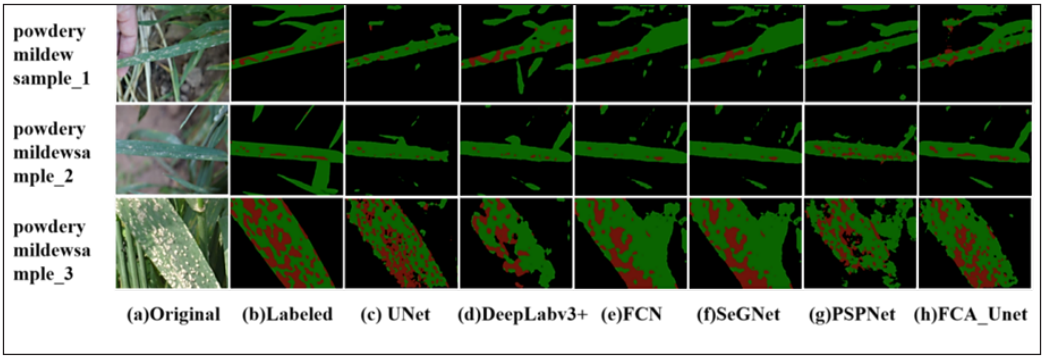


Figure 9. Example images of powdery mildew demonstrating visual results obtained by segmentation models. Masks for background, normal leaf and disease spots are represented by black, green and red, respectively

From Table 5, it can be said that different models produce markedly different DS predictions. For instance, in the case of stripe rust sample_1, the ground truth DS is 5.62%. The FCA_Unet predicts a DS of 5.86%, with an absolute error of only 0.24%. In contrast, the predictions from UNet, DeepLabv3+, FCN, SegNet, and PSPNet are 6.48%, 5.05%, 4.90%, 5.27%, and 0%, with absolute errors of 0.86%, -0.57%, -0.72%, -0.35%, and -5.62%, respectively.

Since this study focuses on disease severity, DS accuracy is crucial. However, Tables 4 and 5 show that a DS value close to the ground truth does not necessarily correspond to a high mIoU.

Table 4
Comparison of mIoU performance of different segmentation models on the test set

Number	Test sample	UNet	DeepLabv3+	FCN	SeGNet	PSPNet	FCA_Unet
1	stripe rust sample_1	0.8898	0.8796	0.896	0.8892	0.632	**0.8984**
2	stripe rust sample_2	0.8557	0.8103	0.8039	0.8351	0.6866	**0.8725**
3	stripe rust sample_3	0.8581	0.849	0.8833	0.8742	0.6603	**0.8852**
4	stripe rust sample_4	0.8824	0.8717	0.8774	0.8632	0.7041	**0.8985**
5	powdery mildew sample_1	0.4815	0.6176	0.6031	0.6012	0.4523	**0.624**
6	powdery mildew sample_2	0.5094	0.5737	0.6400	0.5512	0.4129	**0.6413**
7	powdery mildew sample_3	0.6796	0.5706	0.5999	0.4813	0.5936	0.7307

Table 5
Comparison of Dice Score (DS) performance of different segmentation models on the test set

Number	Test sample	TRUE	UNet	DeepLabv3+	FCN	SeGNet	PSPNet	FCA_Unet
1	stripe rust sample_1	5.62%	6.48%	5.05%	4.90%	5.27%	0.00%	**5.86%**
2	stripe rust sample_2	73.81%	74.17%	68.02%	61.66%	68.86%	86.54%	**72.10%**
3	stripe rust sample_3	21.99%	27.53%	**21.75%**	19.95%	26.96%	14.58%	24.84%
4	stripe rust sample_4	17.19%	15.37%	14.00%	14.13%	13.50%	7.13%	**16.84%**
5	powdery mildew sample_1	8.92%	5.80%	15.65%	**9.79%**	5.51%	2.14%	11.29%
6	powdery mildew sample_2	5.10%	2.82%	4.03%	8.79%	13.17%	0.00%	**6.13%**
7	powdery mildew sample_3	41.88%	55.27%	**38.29%**	32.50%	26.90%	28.69%	33.68%

A single DS value can be unreliable because DS is a relative measure: it only reflects the agreement between numerical proportions but cannot determine whether each diseased or healthy pixel is correctly predicted. By contrast, mIoU evaluates pixel-level agreement and thus provides a more rigorous measure.

To address this, we propose the OMCS, which integrates both mIoU and DS into a single, more intuitive indicator for identifying the best-performing model. The OMCS results derived from Tables 4 and 5 are illustrated in Figure 10. Here, the normalised ranges of Equations 8 and 9 are both in 0-1. Therefore, the OMCS in this case ranges from 0 to 2, where a larger value indicates better model performance. As shown in Figure 10, FCA_Unet is the optimal model in 5 out of 7 samples, demonstrating that FCA_Unet is the best-performing model among all those compared.

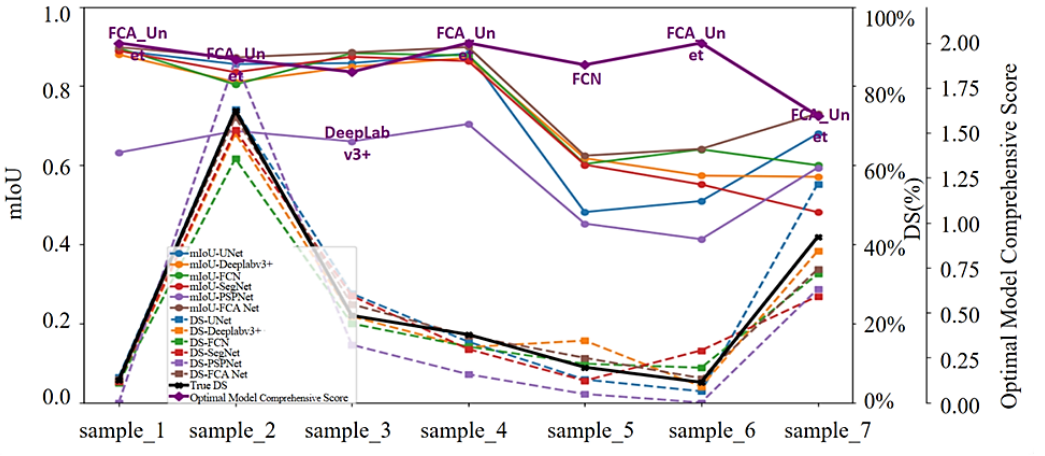


Figure 10. Comprehensive score analysis of the optimal model

Discussion

This study introduces FCA_Unet, a lightweight and high-performance model for wheat leaf disease segmentation. By integrating the A_CBAM attention mechanism and MFFE feature fusion module with an optimised FCA_ResNet backbone, the model achieves superior lesion segmentation accuracy and detail preservation while reducing parameters by 31.8% compared with UNet. Experimental results demonstrate its advantage over benchmark models, reaching a maximum mIoU of 89.85% and ranking as the optimal model in most test samples through OMCS evaluation.

Comparative experiments show that FCA_Unet achieves the highest mIoU of 89.85% on the test set, outperforming classical models such as FCN, PSPNet, and SegNet. Although FLOPs increased by 55.69% compared to UNet due to attention and dilated convolutions, the reduced parameter size facilitates deployment on resource-limited devices, a trade-off that aligns with the practical needs of agricultural on-site detection (Liu et al., 2025). To avoid the high computational cost incurred by traditional attention mechanisms, Al-Gaashani et al. (2025) introduced the multi-scale convolutional PoolFormer block to enhance the model's ability to capture both local and global dependencies.

In this paper, the OMCS, which integrates mIoU and DS, FCA_Unet demonstrates balanced advantages in both pixel-level segmentation and disease quantification, addressing the limitations of relying solely on DS. Notably, discrepancies observed between DS predictions and mIoU highlight that accurate severity assessment depends on precise pixel-level segmentation rather than ratio-based measures alone. By enhancing boundary delineation and small-lesion detection, FCA_Unet provides a more reliable foundation for disease severity evaluation, offering practical value for precision pesticide application in agriculture.

CONCLUSION

This study proposes the FCA_Unet model, an improved variant of UNet, which significantly enhances segmentation performance for wheat leaf diseases. The model's superiority arises from two main innovations: (1) the integration of the A_CBAM attention mechanism and MFFE fusion module, which improve lesion segmentation by capturing contextual information and refining multi-scale features, particularly for small targets and boundary details; and (2) the adoption of FCA_ResNet as the backbone, reducing parameters to 42.1% of ResNet50 and 68.2% of baseline UNet, thereby achieving model lightweighting without compromising accuracy. These findings highlight FCA_Unet as an efficient solution for automated wheat disease monitoring, with potential for broader application to other crop disease segmentation tasks.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), Qilu Institute of Technology, and the StripeRust-Pocket dataset for providing partial support for our experimental data.

REFERENCES

Al-Gaashani, M. S. A. M., Alkanhel, R., Ali, M. A. S., Muthanna, M. S. A., Aziz, A., & Muthanna, A. (2025). MSCPNET: A multi-scale convolutional pooling network for maize disease classification. *IEEE Access, 13*, 11423-11446. https://doi.org/10.1109/ACCESS.2024.3524729

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

Chen, S., Zhang, K., Zhao, Y., Sun, Y., Ban, W., Chen, Y., Zhuang, H., Zhang, X., Liu, J., & Yang, T. (2021). An approach for rice bacterial leaf streak disease segmentation and disease severity estimation. *Agriculture, 11*(5), Article 420. https://doi.org/10.3390/agriculture11050420

Gao, R., Jin, F., Ji, M., & Zuo, Y. (2023). Research on the method of identifying the severity of wheat stripe rust based on machine vision. *Agriculture, 13*(12), Article 2187. https://doi.org/10.3390/agriculture13122187

Ge, Y., Zhang, Q., Xiang, T., Zhang, C., & Bi, H. (2023). TCNet: Co-salient object detection via parallel interaction of transformers and CNNs. *IEEE Transactions on Circuits and Systems for Video Technology, 33*(6), 2600-2615. https://doi.org/10.1109/TCSVT.2022.3225865

Hassan, A., Mumtaz, R., Mahmood, Z., Fayyaz, M., & Naeem, M. K. (2024). Wheat leaf localisation and segmentation for yellow rust disease detection in complex natural backgrounds. *Alexandria Engineering Journal, 107*, 786-798. https://doi.org/10.1016/j.aej.2024.09.018

Hayit, T., Erbay, H., Varçın, F., Hayit, F., & Akci, N. (2021). Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks. *Journal of Plant Pathology, 103*(3), 923-934. https://doi.org/10.1007/s42161-021-00886-2

Jiang, J., Liu, H., Zhao, C., He, C., Ma, J., Cheng, T., Zhu, Y., Cao, W., & Yao, X. (2022). Evaluation of diverse convolutional neural networks and training strategies for wheat leaf disease identification with field-acquired photographs. *Remote Sensing, 14*(14), Article 3446. https://doi.org/10.3390/rs14143446

Kaur, P., Harnal, S., Gautam, V., Singh, M. P., & Singh, S. P. (2023). Performance analysis of segmentation models to detect leaf diseases in tomato plant. *Multimedia Tools and Applications, 83*(6), 16019-16043. https://doi.org/10.1007/s11042-023-16238-4

Kumar, D., & Kukreja, V. (2024). Image segmentation, classification, and recognition methods for wheat diseases: Two decades' systematic literature review. *Computers and Electronics in Agriculture, 221*, Article 109005. https://doi.org/10.1016/j.compag.2024.109005

Kumar, D., Kukreja, V., & Singh, A. (2024). A novel hybrid segmentation technique for identification of wheat rust diseases. *Multimedia Tools and Applications, 83*(29), 72221-72251. https://doi.org/10.1007/s11042-024-18463-x

Li, Y., Qiao, T., Leng, W., Jiao, W., Luo, J., Lv, Y., Tong, Y., Mei, X., Li, H., Hu, Q., & Yao, Q. (2022). Semantic segmentation of wheat stripe rust images using deep learning. *Agronomy, 12*(12), Article 2933. https://doi.org/10.3390/agronomy12122933

Liu, K., Li, L., & Zhang, X. (2025). Fast and accurate identification of kiwifruit diseases using a lightweight convolutional neural network architecture. *IEEE Access, 13*, 84826-84843. https://doi.org/10.1109/ACCESS.2025.3564355

Liu, W., Chen, Y., Lu, Z., Lu, X., Wu, Z., Zheng, Z., Suo, Y., Lan, C., & Yuan, X. (2024). StripeRust-Pocket: A mobile-based deep learning application for efficient disease severity assessment of wheat stripe rust. *Plant Phenomics, 6*, Article 0201. https://doi.org/10.34133/plantphenomics.0201

Mi, Z., Zhang, X., Su, J., Han, D., & Su, B. (2020). Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. *Frontiers in Plant Science, 11*, Article 558126. https://doi.org/10.3389/fpls.2020.558126

Mishra, M., Choudhury, P., & Pati, B. (2024). IoT-enabled plant leaf disease segmentation and multi-classification using mayfly bald eagle optimisation-enabled machine learning. *Multimedia Tools and Applications, 83*(21), 59747-59781. https://doi.org/10.1007/s11042-023-17680-0

Nigam, S., Jain, R., Prakash, S., Marwaha, S., Arora, A., Singh, V. K., Singh, A. K., & Prakasha, T. L. (2022). Wheat disease severity estimation: A deep learning approach. In R. Misra, N. Kesswani, M. Rajarajan, B.

Veeravalli, & A. Patel (Eds.), *Lecture Notes in Networks and Systems* (Vol. 340, pp. 185-193). Springer. https://doi.org/10.1007/978-3-030-94507-7_18

Rai, C. K., & Pahuja, R. (2023). Northern maize leaf blight disease detection and segmentation using deep convolution neural networks. *Multimedia Tools and Applications, 83*(7), 19415-19432. https://doi.org/10.1007/s11042-023-16398-3

Sharma, J., Kumar, D., Chattopadhay, S., Kukreja, V., & Verma, A. (2024). Automated detection of wheat powdery mildew using YOLACT instance segmentation. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimisation (Trends and Future Directions)* (pp. 1-4). https://doi.org/10.1109/ICRITO61523.2024.10522394

Sharma, T., & Sethi, G. K. (2024). Improving wheat leaf disease image classification with point rend segmentation technique. *SN Computer Science, 5*(2), Article 224. https://doi.org/10.1007/s42979-023-02571-w

Shi, L., Liu, Z., Yang, C., Lei, J., Wang, Q., Yin, F., & Wang, J. (2024). Lightweight U-Net-based method for estimating the severity of wheat fusarium head blight. *Agriculture, 14*(6), Article 938. https://doi.org/10.3390/agriculture14060938

Shwetha, V., Bhagwat, A., Laxmi, V., & Shrivastava, S. (2024). A custom backbone UNet framework with DCGAN augmentation for efficient segmentation of leaf spot diseases in jasmine plant. *Journal of Computer Networks and Communications, 2024*, Article 15. https://doi.org/10.1155/2024/5057538

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., & Zhu, H. (2021). A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Computers and Electronics in Agriculture, 189*, Article 106373. https://doi.org/10.1016/j.compag.2021.106373

Wang, H., Ding, J., He, S., Feng, C., Zhang, C., Fan, G., Wu, Y., & Zhang, Y. (2023). MFBP-UNET: A network for pear leaf disease segmentation in natural agricultural environments. *Plants, 12*(18), Article 3209. https://doi.org/10.3390/plants12183209

Wang, Z., Yang, L., Wang, R., Lei, L., Ding, H., & Yang, Q. (2024). WE-DeepLabV3+: A lightweight segmentation model for *Panax notoginseng* leaf diseases. *Computers and Electronics in Agriculture, 227*, Article 109612. https://doi.org/10.1016/j.compag.2024.109612

Yang, T., Wang, Y., & Lian, J. (2024). Plant diseased lesion image segmentation and recognition based on improved multi-scale attention net. *Applied Sciences, 14*(5), Article 1716. https://doi.org/10.3390/app14051716

Yao, C., Yang, Z., Li, P., Liang, Y., Fan, Y., Luo, J., Jiang, C., & Mu, J. (2024). Two-stage detection algorithm for plum leaf disease and severity assessment based on deep learning. *Agronomy, 14*(7), Article 1589. https://doi.org/10.3390/agronomy14071589

Yi, X., Wang, J., Wu, P., Wang, G., Mo, L., Lou, X., Liang, H., Huang, H., Lin, E., Maponde, B. T., & Lv, C. (2023). AC-UNet: An improved UNet-based method for stem and leaf segmentation in *Betula luminifera*. *Frontiers in Plant Science, 14*, Article 1268098. https://doi.org/10.3389/fpls.2023.1268098

Yi, X., Zhou, Y., Wu, P., Wang, G., Mo, L., Chola, M., Fu, X., & Qian, P. (2024). U-Net with coordinate attention and VGGNet: A grape image segmentation algorithm based on fusion pyramid pooling and the dual-attention mechanism. *Agronomy, 14*(5), Article 925. https://doi.org/10.3390/agronomy14050925

Zang, H., Joseph, A. A., Zhang, S., Fu, H., Huang, L., & Huang, Z. (2025). FCA-ResNet: An improved model with enhanced multi-scale feature fusion and coordinate attention for wheat leaf disease classification. *International Journal of Engineering and Technology Innovation, 15*(2), 195-209. https://doi.org/10.46604/ijeti.2024.14304

Zhang, Q., Ge, Y., Zhang, C., & Bi, H. (2022). TPRNet: Camouflaged object detection via transformer-induced progressive refinement network. *The Visual Computer, 39*, 4593-460. https://doi.org/10.1007/s00371-022-02611-1

Zhang, Q., Sun, X., Chen, Y., Ge, Y., & Bi, H. (2023). Attention-induced semantic and boundary interaction network for camouflaged object detection. *Computer Vision and Image Understanding, 233*, Article 103719. https://doi.org/10.1016/j.cviu.2023.103719

Zhang, S., & Zhang, C. (2022). Modified U-Net for plant diseased leaf image segmentation. *Computers and Electronics in Agriculture, 204*, Article 107511. https://doi.org/10.1016/j.compag.2022.107511

Zhao, C., Li, C., Wang, X., Wu, X., Du, Y., Chai, H., Cai, T., Xiang, H., & Jiao, Y. (2025). Plant disease segmentation networks for fast automatic severity estimation under natural field scenarios. *Agriculture, 15*(6), Article 583. https://doi.org/10.3390/agriculture15060583

Pertanika J. Sci. & Technol. 34 (2): 1203 - 1230 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

An Interpretable Random Forest with SHAP Explanations for Multiclass Skill Level Classification Model in Malaysia's Labour Market

Rabi'atul'adawiah Shabli [1,2], Ahmad Zia Ul-Saufie[1]*, Arief Gusnanto[3], and Nurain Ibrahim[1]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[2]Department of Statistics Malaysia, Ministry of Economy, Federal Government Administrative Centre, 62514 Putrajaya, Wilayah Persekutuan Putrajaya, Malaysia

[3]Department of Statistics, School of Mathematics, Faculty of Engineering and Physical Sciences, University of Leeds, LS2 9JT Leeds, United Kingdom

ABSTRACT

Skill level classification is essential for effective human capital interventions, wage policy formulation and long-term workforce development. However, the black-box nature of machine learning algorithms limits the transparency and interpretability required for policymaking. This study aims to develop an accurate and interpretable multiclass skill level classification model for the Malaysian labour market. The model was constructed using secondary data from the Department of Statistics Malaysia's 2023 Salaries and Wages Survey, covering 120,518 Malaysian workers. Data preprocessing involved cleaning, transformation and addressing class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). Using the top nineteen XGBoost feature importance scores, a Random Forest model was selected due to its robust ensemble mechanism and trained using these features. The proposed model exhibits strong predictive performance across all metrics, achieving an accuracy of 0.8754, sensitivity of 0.8485, specificity of 0.9384 and F1-score of 0.8378. The interpretability of the most influential predictors was investigated using SHapley Additive exPlanations (SHAP) at both global and local levels, revealing the importance of salaries and wages activity, highest certificate and education level. These features provide deeper insights into understanding skill level classification across skilled, semi-skilled and low-skilled categories. The findings emphasise the significance of combining high predictive

ARTICLE INFO

Article history:
Received: 18 October 2025
Accepted: 04 March 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.25

E-mail addresses:
adawiah.dosm@gmail.com (Rabi'atul'adawiah Shabli)
ahmadzia101@uitm.edu.my (Ahmad Zia Ul-Saufie)
a.gusnanto@leeds.ac.uk (Arief Gusnanto)
nurainibrahim@uitm.edu.my (Nurain Ibrahim)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

accuracy with transparent feature interpretability in labour market analysis. This study represents one of the initial labour market studies in Malaysia integrating Random Forest with SHAP. The proposed interpretable skill level framework can be utilised to facilitate evidence-based policymaking, targeted human capital interventions and workforce development planning in Malaysia.

Keywords: Classification, explainable artificial intelligence, feature selection, labour market, random forest, SHAP, skill level, XGBoost feature importance

INTRODUCTION

Machine learning algorithms are extensively applied to complex prediction tasks, yet they are often criticised for functioning as "black boxes", where linkages between input features and target variables remain opaque (Neubauer et al., 2025). This lack of transparency posed a challenge among policymakers and decision-makers who required not only accurate predictions but also clear interpretability to cultivate trust and accountability. As a result, researchers have aimed to enhance machine learning algorithms by combining high predictive accuracy and techniques that can clarify the model behaviour in a way that is easily accessible and actionable (Neubauer et al., 2025; Sani et al., 2025).

Explainable Artificial Intelligence (XAI) has become an effective solution to address this issue, as this technique allows the operational mechanisms of black-box models to be made more transparent (Dhungana, 2025). Implementation of XAI will foster user confidence, model refinement and strengthen the governance framework. A considerable amount of literature has been published on the utilisation of interpretability techniques for classification purposes, such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), Explain Like I'm Five (ELI5), Quantum Lattice (QLattice), and Anchor (Arravalli et al., 2025).

Nowadays, SHAP has been successfully integrated with machine learning algorithms in various fields. SHAP has been employed to clarify input predictors of therapy dropout (Visibelli et al., 2025), chronic kidney disease (Elshewey et al., 2025), and breast cancer symptoms (Arravalli et al., 2025). In medicine, it enables clinicians to discover non-linear associations among the features and facilitate precision medicine. Likewise, Sani et al. (2025) adopted the Random Forest and SHAP to clarify sociodemographic characteristics that affected the fertility preferences of Somalian women. SHAP has been used in the classification of industrial safety, including ExAIRFC-GSDC, to identify input predictors of gas leakage accidents with an accuracy of more than 98 per cent (Lalithadevi & Krishnaveni, 2025). Taken together, this application suggests that integrating SHAP can strengthen predictive accuracy and interpretability, with the intention that predictions are not only strong but also explainable.

Recent labour market studies that apply XAI methods have indicated an increase in the ability to overcome the limitations of traditional statistical analysis. This can be seen in other studies that have applied SHAP interpretation on the job losses among the elderly (Celbiş et al., 2023), unemployment rates in Europe (Celbiş, 2023), occupational injuries in South Korea (Lee et al., 2023), gig workers' employment (Liu & Wang, 2025), employee attrition (Manafi Varkiani et al., 2025) and online vacancies (Dossche et al., 2025). Together, these studies outline that tree-based algorithms, specifically Random Forest and XGBoost were found to be the most appropriate techniques due to their ability to predict, while SHAP enables revealing of the input predictors, including age, education, activity and employment status.

Taken together, prior research underscores the growing relevance of SHAP-enhanced machine learning in fields where both predictive performance and interpretability are essential. Until recently, amidst advances in medicine, industrial safety, workplace injuries and the labour market, there has been little research concentrated on the classification of skill level. Accurate and interpretable categorisation of workers into skilled, semi-skilled and low-skilled is vital for shaping wage policies, human capital interventions and long-term workforce planning. This study, therefore, set out to address the gap by combining the Random Forest with SHAP analysis to enhance skill level prediction in Malaysia. By combining strong predictive performance with feature-level transparency, the study aimed to provide actionable insights for policymakers to design evidence-based and targeted labour market initiatives.

Consequently, this study provides several key contributions. From the methodological perspective, this study incorporates an ensemble Random Forest model with SHAP explanations, enabling an accurate prediction and interpretability for multiclass skill level classification. Additionally, this study leverages national representative data of the 2023 Salaries and Wages Survey, consisting of 120,518 Malaysian workers, to identify the most influential skilled, semi-skilled and low-skilled input predictors across demographic, socioeconomic and employment-related features. Lastly, in terms of policy implications, this proposed skill level framework offers transparent and empirical information which may support human capital strategies pertaining to labour, employment and social protection for a dynamic and competitive Malaysian labour market.

METHODOLOGY

This study utilised the cross-sectional 2023 *Salaries and Wages Survey* dataset, a household survey approach of Labour Force/ Salaries & Wages/ Migration Survey from the Department of Statistics Malaysia (Department of Statistics Malaysia, 2024a). The dataset consisted of 120,518 workers with nineteen features and could be segmented into three components encompassing identification particulars, labour force particulars and salaries

and wages (Department of Statistics Malaysia, 2024a). It adopted a stratified two-stage sampling design to cater to the national representatives throughout every state across the nation (Department of Statistics Malaysia, 2024b). The research workflow was illustrated in Figure 1, which describes a sequence of processes encompassing data preprocessing, feature selection, model training and model interpretation.

This study is based on a cross-sectional dataset that captures the workforce in a single year of 2023. Given that the dataset is cross-sectional, the analysis assumes that the features reflect stable and underlying patterns in the labour market. The assumption of limited short-term mobility is supported by the Thirteenth Malaysia Plan 2026-2030, which indicates that social mobility indicators remain at a low level and that young workers continue to experience constraints in reaching the skills required for the job market (Ministry of Economy, 2025). This classification model serves as a structural benchmark that identifies fundamental requirements, such as education level and the highest certificate, to define skill level classification in the Malaysian labour market at a particular time.

Skill Level Categories

The target variable, namely the skill level, is a categorical feature. In the Malaysian context, skill level can be classified into skilled, semi-skilled and low-skilled categories (Department of Statistics Malaysia, 2025; Ministry of Human Resources, 2020). The nine major occupational groups in the Malaysia Standard Classification of Occupations (MASCO) are used to disaggregate labour into three skill level categories as depicted in Table 1. The definition of skill level is determined based on a single feature, namely, the occupational classification based on MASCO 2020 (Department of Statistics Malaysia, 2025). As a result, the current level of skill categorisation before model training is not influenced by other input predictors, including the income level. The income level is defined as an outcome measure that indicates the labour market returns to skill and is not included as an element in the classification of skill level categories (Zhao et al., 2025).

According to Figure 2, the semi-skilled category registered the highest cases at 74,717 with a percentage share of 62.00%. This was followed by the skilled category with 30,969 (25.70%) and the low-skilled category at 14,832 cases (12.31%). The skill level composition faced an imbalance in class structure, where classification models had a prediction preference for the majority category, causing lower prediction accuracy for the minority categories (Zhao & Li, 2025).

Data Preprocessing

A data preprocessing procedure was required, which included data cleaning, data transformation, data partitioning and handling imbalanced classes within the target variable. Data cleaning contained four procedures, namely eliminating duplicate cases, managing

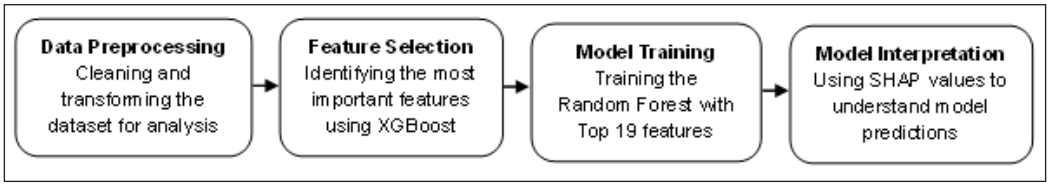


Figure 1. Research flow illustrating data preprocessing, feature selection, and model training and model interpretation

Table 1
Skill level categories based on major groups of MASCO 2020

Skill level categories	Major occupational group
Skilled	1: Managers
	2: Professionals
	3: Technicians and Associate Professionals
Semi-skilled	4: Clerical Support Workers
	5: Service and Sales Workers
	6: Skilled Agricultural, Forestry, Livestock and Fishery Workers
	7: Craft and Related Trades Workers
	8: Plant and Machine Operators and Assemblers
Low-skilled	9: Elementary Occupations

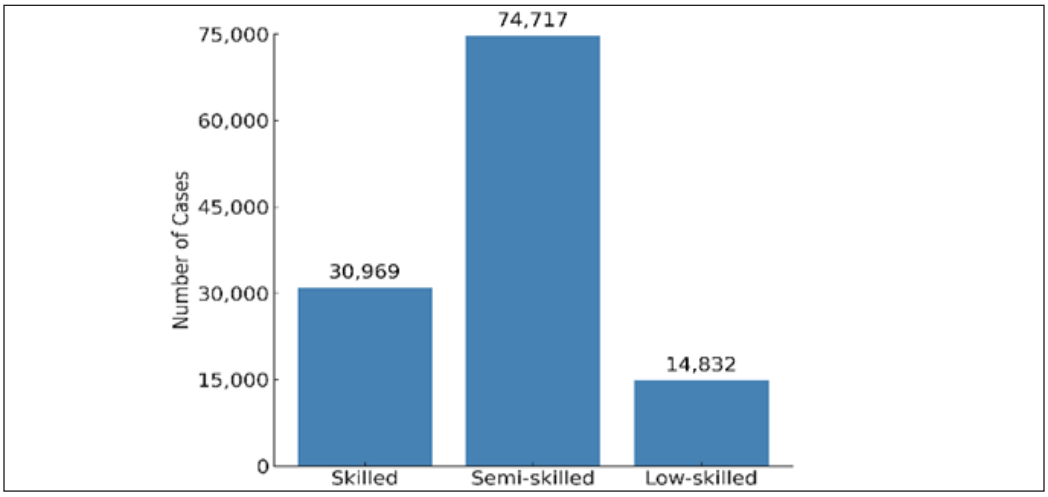


Figure 2. Distribution of skill level by skilled, semi-skilled and low-skilled categories

missing values, detecting outliers and eliminating irrelevant features. The dataset was randomly split into a ratio of 80:20, where 80 per cent of observations were assigned for training and 20 per cent for testing (Khalfallah et al., 2025; Sakyi-Yeboah et al., 2025) before the execution of the machine learning technique (Shaha et al., 2024).

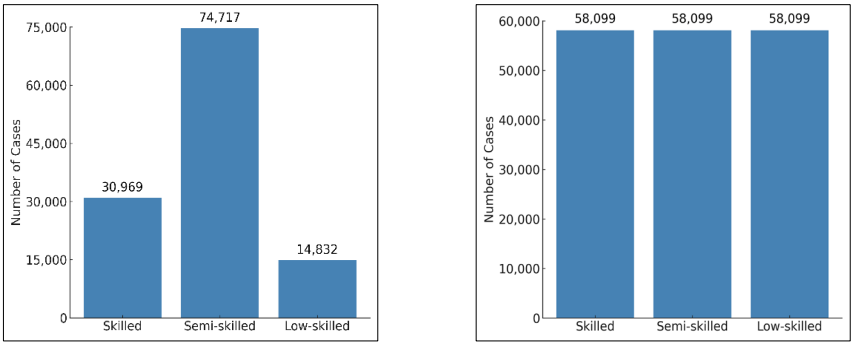
An imbalanced dataset may limit decision-making validity, as it introduces bias in machine learning models towards majority categories while reducing the effectiveness of predictive performance for minority categories (Sheokand & Beniwal, 2025). Two of the most frequent practices used to handle imbalanced datasets are oversampling and undersampling. Oversampling generates synthetic data for minority categories while undersampling reduces the cases in majority categories, resulting in a similar distribution across all categories (Unlu & Subasi, 2025). In the meantime, undersampling may lead to the loss of information on the dominant category.

To address the multiclass imbalance classification problem within the target variable, the Synthetic Minority Oversampling Technique (SMOTE), one of the most prominent methods, was employed to ensure an equal distribution among the three skill level classes, namely skilled, semi-skilled and low-skilled, by generating synthetic cases for the minority classes (Unlu & Subasi, 2025). This finding is consistent with Tsai et al. (2024) who reported that employing SMOTE is effective in addressing multiclass imbalanced datasets.

In this study, the SMOTE was applied specifically to the training dataset to balance the samples (Wang et al., 2025). In labour market data, minority categories like low-skilled may represent a structurally different worker population instead of having a smaller number of samples due to undersampling. As a result, SMOTE was applied only to the training dataset to enhance classification performance. The testing dataset remained unchanged to maintain the actual labour market distribution.

Figure 3 illustrates the comparison of skill level distributions before and after implementing SMOTE. The original distribution indicated an imbalanced class with the semi-skilled component dominating the dataset at 74,717 cases, representing 62 per cent of the overall dataset. A dominant category of semi-skilled individuals resulted in high accuracy of prediction on the dominant class and low accuracy of prediction on the skilled and low-skilled individuals (Hasan et al., 2025). After the SMOTE technique implementation, these three categories obtained a balanced distribution with 58,099 cases for each category. This procedure was to ensure that the prediction model was not biased towards the majority class and to strengthen the model's effectiveness in correctly classifying minority skill level categories (Park et al., 2024). Further analysis of classification performance was conducted after addressing class imbalance using the SMOTE technique.

Subsequently, a prediction model was developed using Random Forest, as it was one of the most prominent and powerful ensemble machine learning algorithms. This bagging algorithm was integrated with SHAP to determine important features influencing skill level



(a) Before SMOTE (b) After SMOTE

Figure 3. Distribution of skill level categories before and after applying SMOTE, demonstrating the imbalanced dataset into equal representation between skilled, semi-skilled and low-skilled workers

classification, thereby providing an explanation of the factors distinguishing the skilled, semi-skilled and low-skilled categories. A further explanation pertaining to the Random Forest algorithm and the SHAP interpretability was discussed in the subsequent segment of Model Training using Random Forest and Model Interpretation using Explainable SHAP analysis.

Feature Selection

Feature selection is an essential procedure for determining quality features in developing a reliable and strong skill level classification model (Hasan et al., 2025). These methods are categorised into filter, wrapper and embedded techniques. This process specified the relationship between input features and the target variables (Pedersen et al., 2025). It makes the selection of the most relevant input predictors that contribute to the skill level model prediction while removing the irrelevant features (Hasan et al., 2025).

Tree-based algorithms like Random Forest and XGBoost have built-in mechanisms for measuring the importance of features, which improve the efficiency in analysing relevant predictors in a classification model (Méndez-Astudillo, 2024). The main difference between the feature importance computation in Random Forest and XGBoost, as stated by Méndez-Astudillo (2024), is that Random Forest uses the total impurity reduction while XGBoost uses the average gain of the tree splits. However, XGBoost strength is when dealing with large volumes of data, various input predictors and determining the important features (Danquah et al., 2025; Méndez-Astudillo, 2024).

Based on the advantage of feature importance offered by XGBoost, a built-in feature selection method was employed to determine the most relevant input predictors to classify the skill level (Pedersen et al., 2025). XGBoost computes the feature importance score

using the gain method during its boosting process (Wang et al., 2024). According to Wang et al. (2024), XGBoost computes the feature importance score naturally in the training process and ranks nineteen features according to their contribution value. Empirical data presented by Wang et al. (2024) show that built-in feature importance is effective when dealing with large datasets, where there are multiple input predictors.

Accordingly, the XGBoost and Random Forest algorithms utilise different ensemble structures where boosting reduces bias and bagging reduces variance. However, both share similar inductive biases as they rely on decision tree-based learners. In line with the findings by Kurniadi and Saputri (2024), the combination of XGBoost and Random Forest leverages the high predictive accuracy and feature selection effectiveness of XGBoost, with the stability and robustness of Random Forest.

In accordance with the statement by Sariman et al. (2024), feature selection reduces the inclusion of irrelevant input features, which minimises the complexity of the model, but does not affect the skill level classification performance. As a consequence, the nineteen features ranked by built-in XGBoost were added to a Random Forest algorithm to improve the accuracy of skill level classification (Pourdaryaei et al., 2024). The deployment was carried out in nineteen iterations, starting with the highest importance score and proceeding to the least important score. These nineteen iterations were then evaluated through performance metrics to determine the proposed model. The following segment describes the Random Forest procedure and evaluation on these ranked input predictors.

Model Training using Random Forest

The skill level classification prediction used in this study applied the proposed Random Forest algorithm, which has been recognised due to its robust performance compared to other tree-based algorithms (Visibelli et al., 2025). This ensemble machine learning algorithm, based on the bagging technique, was introduced by Breiman (2001) and can be implemented to perform both classification and regression problems (Bodilsen et al., 2023; Liu & Wang, 2025). The algorithm constructs multiple decision trees separately using bootstrap samples of the original dataset. Those trees are trained on a random subset of input predictors at each split.

The randomness mechanism minimises the correlation between the single trees and increases the model generalisation, particularly when dealing with a large sample size, noisy data and high-dimensional datasets (Liu & Wang, 2025). The final Random Forest classification outcome is constructed by combining all the trees' predictions using a majority voting process, which helps to enhance accuracy and stability (Elshewey et al., 2025; Sakyi-Yeboah et al., 2025; Visibelli et al., 2025). This ensemble structure generates several decision trees with different split criteria, enabling stronger classification performance and removing irrelevant predictions (Salah-Ud-Din et al., 2024).

Random Forest is known for its effectiveness in handling complex datasets and robustness against overfitting (Visibelli et al., 2025). Additionally, this algorithm demonstrates strong generalisation, requires minimal data preprocessing procedures and achieves greater accuracy against the traditional algorithms even with smaller datasets (Zahra et al., 2025).

In this study, the Random Forest was implemented with the number of decision tree fixed at 100. The minimum sample size allowed to split an internal node was two, while the minimum number of samples at a leaf node was one. The Gini index was used as the splitting criterion to measure node impurity (Zeffora & Shobarani, 2022). This is due to its computational efficiency and is widely used in classification problems. There was no limit required on the maximum depth of the trees or even the number of leaf nodes, which allows the model to grow at full capacity and learn the complex decision boundaries within the skill level dataset. The random state was fixed at 42 to ensure reproducibility of the results (Hsieh, 2025). The set of parameters used was selected to balance the flexibility of the model and predictive power in the classification of skill level in the context of the Malaysian labour market.

Model Interpretation using Explainable SHAP Analysis

Interpretability techniques are an instrument to distinguish the underlying mechanisms of machine learning algorithms (Liu & Wang, 2025). It could be implemented before or during the model training (Neubauer et al., 2025). XAI approaches were developed to provide clarification into the operations of the models to facilitate end-user understanding (Neubauer et al., 2025). Those methods provided model transparency, strengthening trust in the predictive procedure and revealing the relationships between the target variable and the predictors (Liu & Wang, 2025). Leveraging XAI techniques was for the purpose of obtaining deeper explanations of the impacted features contributing to the performance of Random Forest algorithms in skill level classification. The most well-known model-agnostic approaches for interpreting the prediction mechanisms are SHAP, LIME, and Integrated Gradients (Joshi et al., 2025).

One of the prominent XAI techniques was SHAP analysis, introduced in 2017 by Lundberg and Lee (2017). The computation was an extension of the Shapley value measurement developed by Shapley (1953). The concept of SHAP was based on cooperative game theory, where each individual had a different contribution, and that allocation was distributed among the players following the proportion of their contributions (Liu & Wang, 2025). It relies on Shapley values to illustrate the feature's importance at both the global and local levels (Pedersen et al., 2025). SHAP supported interpretability at the global level by revealing a comprehensive understanding of the overall model behaviour, whereas the local level provided contribution scores of individual features (Dossche et al., 2025).

Considering these approaches, SHAP has been particularly prominent because of its transparency, reliability, and ability to explain model predictions according to individual characteristics using Shapley values (Adalarasu et al., 2025; Arravalli et al., 2025). Unlike traditional regression methods that provide only average effects through coefficients, SHAP illustrated how each feature contributes positively or negatively to individual predictions (Sani et al., 2025). The computation of SHAP values can be performed via interventional algorithms (Tree SHAP, Linear SHAP, Kernel SHAP) or conditional extensions of Kernel SHAP (Adalarasu et al., 2025). This framework supported both global explanations, summarising overall feature importance and local explanations, clarifying feature impacts for specific classes in multiclass classification tasks. Such interpretability is needed to comprehend the predictors of skill level classification as well as to facilitate evidence-based guidance on human capital initiatives.

Global SHAP provided the overall view of the skill level model predictions. The SHAP values described the importance of input features by providing a contribution score for individual features that reflected their impact on the predicted skill level categories (Ogungbire & Pulugurtha, 2025). The formula of SHAP was given in Equation 1 (Ogungbire & Pulugurtha, 2025).

$$\phi_i = \sum_{s \in F_{\{i\}}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{s\cup\{i\}}(X_{s\cup\{i\}}) - f_s(X_s) \right] \quad [1]$$

where ϕ refers to the SHAP value of i^{th} feature, F refers to the total input features, S refers to the feature subsets excluding the i^{th} feature, $f_{s\cup\{i\}}$ and f_s refers to the model prediction without and with i^{th}feature, X_s refers to the input values of features for the subset S. In terms of multiclass classification, a positive number of SHAP values described the feature contribution to drive the classifier towards the exact category (Elshewey et al., 2025). On the contrary, a negative SHAP value showed the impact of a feature that influenced the prediction towards choosing a negative category (Elshewey et al., 2025). Additionally, a study by Sakyi-Yeboah et al. (2025) indicates that greater SHAP values reflected the level of importance of feature contribution towards the model classifier.

Local SHAP explained insights on the contribution of features to the specific skill level classes. The visualisation of SHAP values, highlighting the importance of features, was presented using dotted blue and red colours (Kaur et al., 2025). The effect of features could be analysed through red dots, which described a positive contribution, while blue dots indicated a negative contribution to the model prediction (Kaur et al., 2025; Ogungbire & Pulugurtha, 2025). Computing SHAP values underlined the most important features along with the impact and direction by offering an in-depth perspective on the mechanism of the model behaviour.

To overcome the problem of attribution inflation between correlated predictors such as education level and highest certificate, this study applies Tree SHAP by estimating conditional Shapley values based on the model outlined by Berge et al. (2024). This method explicitly considers inter-feature dependencies, whereby the observed values of the remaining input predictors are conditioned during the explanatory procedure. As a result, the model distributes contributions more fairly among related predictors and reduces the likelihood of inflating feature importance scores, a situation in which features are assumed to be independent.

Even though SHAP is computationally intensive, regular use by government bodies can be enabled through the implementation of less computationally intensive procedures, optimised algorithms and sampling methods (Mitchell et al., 2022). These strategies can reduce computational requirements and make SHAP a feasible solution for agencies with limited analytical resources. Furthermore, SHAP does not require extensive infrastructure, as this is not a real-time dataset, whereby the analysis is not needed frequently and does not require high-end infrastructure. The study offers some important insights, which not only enhance the accuracy of prediction but also offer a clear indication that could be used in human capital interventions and employer strategies.

Model Evaluation

The effectiveness of the machine learning algorithm was evaluated using classification metrics, including accuracy, sensitivity, specificity and F1-score (Bağcı & Kaya Soylu, 2025; Wang et al., 2024). In relation to multiclass classification, Bağcı and Kaya Soylu (2025) and Wang et al. (2024) indicates that a confusion matrix was implemented to determine true predictions involving true positives (TP) and true negatives (TN), whereas false predictions were defined as false positives (FP) and false negatives (FN).

The formula shown in Equation 2 specified the accuracy, quantified as the ratio of true predictions generated by the model to the total number of classification predictions (Rafrastara et al., 2025). In this context, accuracy reflected the ability of the Random Forest algorithm to effectively categorise workers into skilled, semi-skilled and low-skilled levels.

$$Accuracy = \frac{(TP1 + TP2 + TP3) + (TN1 + TN2 + TN3)}{(TP1 + TP2 + TP3) + (FP1 + FP2 + FP3) + (TN1 + TN2 + TN3) + (FN1 + FN2 + FN3)} \quad [2]$$

The F1-score was described as a combined metric between recall and precision, measured through its harmonic mean as computed in Equation 3 (Bağcı & Kaya Soylu, 2025). It described the model's capability to accurately determine the worker's skill level classification while minimising misclassification across the three skill level categories.

$$F1\ score = \frac{2 * Precision\ * Recall}{(Precision\ + Recall\ )} \quad [3]$$

Meanwhile, sensitivity measured the effectiveness of the classification model in capturing true positive cases for each skill level category, as given in Equation 4 (Ul Haq et al., 2024). It examined how effectively the model recognised the worker's actual skill level category.

$$Sensitivity = \frac{(TP1 + TP2 + TP3)}{(TP1 + TP2 + TP3) + (FN1 + FN2 + FN3)} \quad [4]$$

In this study, specificity was examined by the ability of the classification model to distinguish true negative cases among the three skill level categories, which were defined as Equation 5 (Ul Haq et al., 2024). It prevented workers from being wrongly classified into other skill level categories beyond their actual skill category.

$$Specificity = \frac{(TN1 + TN2 + TN3)}{(TN1 + TN2 + TN3) + (FP1 + FP2 + FP3)} \quad [5]$$

RESULTS AND DISCUSSION

This section highlights three primary outcomes. It begins with the descriptive analysis of the 2023 Salaries and Wages dataset, followed by the feature selection outcomes based on XGBoost feature importance, Random Forest classification performance and interpretability of XAI using the SHAP technique.

Descriptive Analysis

The Salaries and Wages Survey for the reference year 2023 comprised 120,518 workers. The illustration, as shown in Figure 4, emphasises the composition of educational attainment, with the majority having completed secondary education at 54.7 per cent, followed by tertiary education (33.9 per cent). Only a small segment had completed primary education at 9.3 per cent, whereas a relatively minimal share of 2.2 per cent reported having no formal education. This finding describes an educated workforce recommending potential and readiness for greater skill enhancement initiatives aimed at increasing skilled labour in the country.

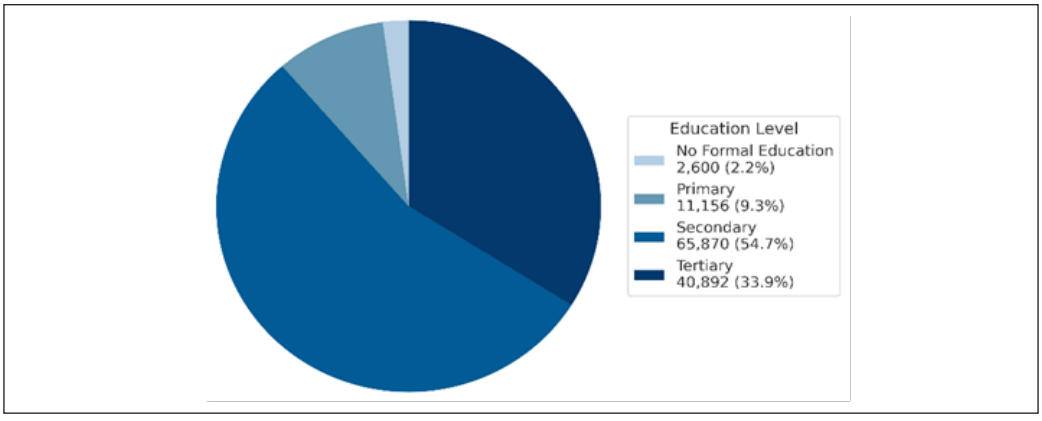


Figure 4. Composition of educational attainment

XGBoost Feature Importance

The XGBoost feature importance analysis indicated the importance scores for nineteen features in the dataset, as depicted in Figure 5. It demonstrated that education level, with a score of 0.213, was the most relevant predictor to determine skill level categories. Salaries and wages ranked second (0.146), followed by the highest certificate (0.118). Concurrently, employment status (0.095) and activity (0.067) also played important roles during the model training. The demographic and labour market characteristics, including strata, gender, hours worked and marital status, had a moderate influence on the skill level categories. Features such as district and disability status had minimal effects, whereas citizenship did not contribute to the skill level prediction. These findings confirm that labour market indicators, particularly education, salaries and wages, and activity are the primary drivers in the model for further SHAP interpretability analysis.

Selected feature importance rankings have been utilised in other labour market studies. Jamaludin et al. (2022) highlighted that salaries and wages were an important indicator as they influenced labour mismatch in the labour market scenario. According to Mansor and Othman (2024), education was critical as the higher level of education influenced the employability in the labour market demand. From the unemployment perspective, Celbiş (2023) indicated that age and education influenced the opportunities for an individual to participate in the labour market.

Despite the dataset being from the 2023 reference year, it shows stability because the statistics reflect the underlying structural dynamics of the Malaysian labour market. Macroeconomic shocks can affect the volume of employment (Department of Statistics Malaysia, 2025), but they do not cause immediate changes in entry-level, including education level and the highest certificate, which are prerequisites for a specific skill level (Ministry of Human Resources, 2020). Additionally, the Thirteenth Malaysia Plan

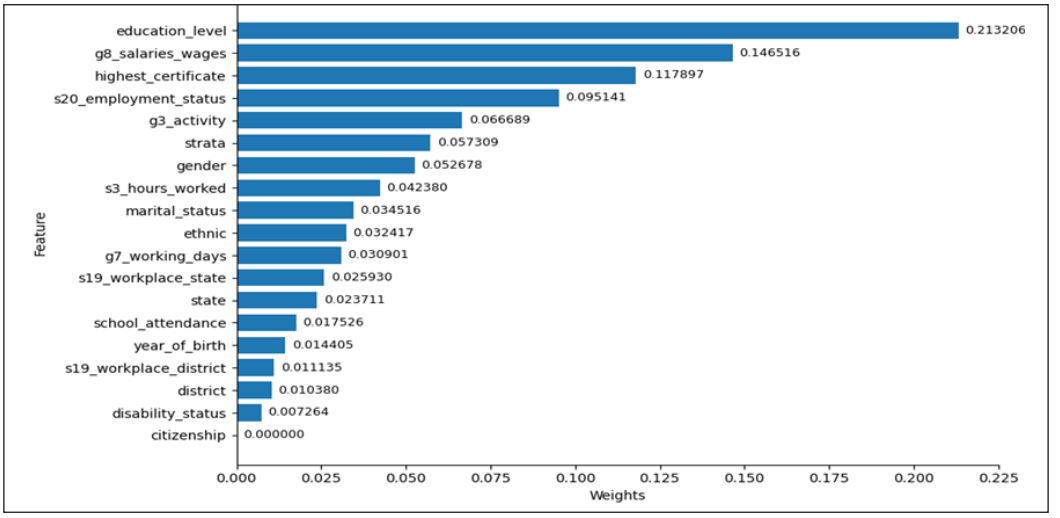


Figure 5. List of XGBoost Feature Importance based on nineteen features

2026-2030 indicates that challenges such as skill mismatch and concentration in low-skill occupations have been persistent issues over a long period, thus confirming the stability of these trends (Ministry of Economy, 2025). Consequently, the findings reflect structural association patterns at one reference period and should not be interpreted as causal effects between skill level categories.

This study employed XGBoost feature importance ranking due to its computational efficiency and suitability for large datasets (Danquah et al., 2025). Different feature selection techniques may produce different feature subsets and interpretation outcomes (Wang et al., 2024). This study considers skill level as a fixed classification based on cross-sectional data. It implicitly assumes limited short-term mobility across skill level categories. The use of longitudinal data for further studies would allow skill development analysis as a dynamic trajectory.

Random Forest Performance

Despite the skill level being determined by MASCO occupational guidelines, this study does not attempt to reproduce these guidelines. Instead, it aims to estimate the association between socioeconomic characteristics and skill level categories. The predictive model identifies the most important determinants of skill level status and assigns weights to their relative importance. It also enables estimation when occupational information is incomplete. For instance, the model can predict skill level when occupational information is not provided in administrative datasets. As a result, the model provides explanatory and policy-relevant insights that extend beyond the fixed MASCO classification.

This study employed a powerful machine learning technique to operationalise this predictive framework and capture the non-linear relationships among the input predictors.

To verify the effectiveness of the Random Forest classification model, the performance was examined in previous studies. In the prior literature, a comparable analysis of five tree-based algorithms consisting of Decision Tree, Random Forest, Gradient Boosting Machines, Adaptive Boosting and Extreme Gradient Boosting was examined for multiclass skill level prediction (Rabi'atul'adawiah Shablia et al., 2025). The results demonstrated that Random Forest recorded an outstanding predictive performance across accuracy (0.8645), specificity (0.9089), sensitivity (0.8645) and F1-score (0.8636) when dealing with imbalanced skill level categories. Thus, the combination of Random Forest with SHAP explanations was further investigated to enhance transparency and interpretability.

This study determined Random Forest as the most effective classification model for classifying skill level categories in the Malaysian labour market. The findings were consistent with prior classification studies across several domains, including medical and other labour market studies, where Random Forest consistently outperformed other tree-based algorithms (Arravalli et al., 2025; Bağcı & Kaya Soylu, 2025; Khalfallah et al., 2025; Rafrastara et al., 2025; Visibelli et al., 2025). Among others were predictions of therapy dropout among chronic pain patients (Visibelli et al., 2025), detection of breast cancer symptoms (Arravalli et al., 2025), fertility preferences among Somalian women (Sani et al., 2025) and employee attrition (Manafi Varkiani et al., 2025). Taken together, these studies suggest that the robust performance of Random Forest was mainly attributed to its ensemble structure (Khalfallah et al., 2025), efficiency in handling feature redundancy, and the ability to recognise complex classification patterns in the dataset (Lalithadevi & Krishnaveni, 2025) and the ability to mitigate overfitting (Visibelli et al., 2025).

Thus, in this study, the Random Forest classifier was constructed using the top nineteen features obtained from XGBoost feature importance. The selection of sixteen features was made when the predictive accuracy had reached its stability, as illustrated in Figure 6. To provide a statistical rationale for the selected cut-off, rather than relying on visual inspection, a formal stopping criterion was employed, since adding more features to the model produced only minimal increases in accuracy (Syeda & Asghar, 2023). As highlighted by Kempny et al. (2025). there is no common cut-off threshold that is reliable across datasets for defining the stopping criterion. This demonstrates that incorporating additional features only provides a marginal performance improvement while increasing model complexity. This exhibits a balance between model parsimony and sufficient explanatory information.

The feature subset comprised socioeconomic, demographic and employment-related characteristics. These include education attainment, salaries and wages, highest certificate, employment status, activity, strata, gender, hours worked, marital status, ethnic, working days, workplace state, state, school attendance, year of birth and workplace district.

Utilising multidimensional predictors assisted the algorithm's ability to recognise the patterns and relationships affecting skill level categorisation. The effectiveness of this algorithm was further assessed through model evaluation.

To validate that SMOTE did not distort the underlying structure of the training dataset, the distributions of the primary input predictors were compared between the original training dataset and the SMOTE dataset (Mujahid et al., 2024). Continuous features, including salaries and wages hours worked, and year of birth, showed similar patterns in the kernel density plots. For the categorical features, including education level and activity, only slight differences were observed. No artificial clusters or unrealistic values were found. These findings suggest that SMOTE enhanced class balance without changing the heterogeneity of the original data (Mujahid et al., 2024).

To ensure that the predictive model is not driven primarily by salaries and wages, an additional analysis was conducted by removing the salaries and wages feature from the model. The classification accuracy decreased slightly from 0.8645 to 0.8334 (Figure 6). This emphasises that although salary and wages are informative, the overall model performance does not change substantially when the feature is excluded. Accordingly, the predictive outcome is not driven solely by salaries and wages as other socioeconomic and employment characteristics also play an important role in defining the skill level categories. These findings are consistent with evidence from South Korean where skill level does not always improve employability or salaries, indicating that labour market performance, such as salaries, is not necessarily determined solely by individual skills (Ban & Choi, 2025).

After the robustness checks, the predictive performance is summarised through a confusion matrix as exhibited in Figure 7(a). This multiclass classification was further examined by using four evaluation metrics of accuracy, sensitivity, specificity and F1-score, as illustrated in Figure 7(b). The accuracy score of the proposed algorithm was 0.8754, which indicates that most predictions were in concordance with the appropriate skill level categories. As specified by Halias et al. (2023), an accuracy of more than 70 per cent is typically considered to indicate high classification performance.

The F1-score of 0.8378 reflects a balance between recall and precision, which demonstrates Random Forest's strong performance in all categories. The high sensitivity score at 0.8485 illustrates the ability of the Random Forest algorithm to classify each category correctly. A scoring of 0.9384 for specificity reveals the effectiveness of the algorithm in reducing the misclassification of skill level categories. Collectively, these findings suggest that Random Forest is efficient at classifying skill level into the three categories and is suitable for multiclass classification prediction tasks involving skill level. Based on the confusion matrix, most misclassifications occur at the boundaries between categories, particularly between the semi-skilled category and its neighbouring categories.

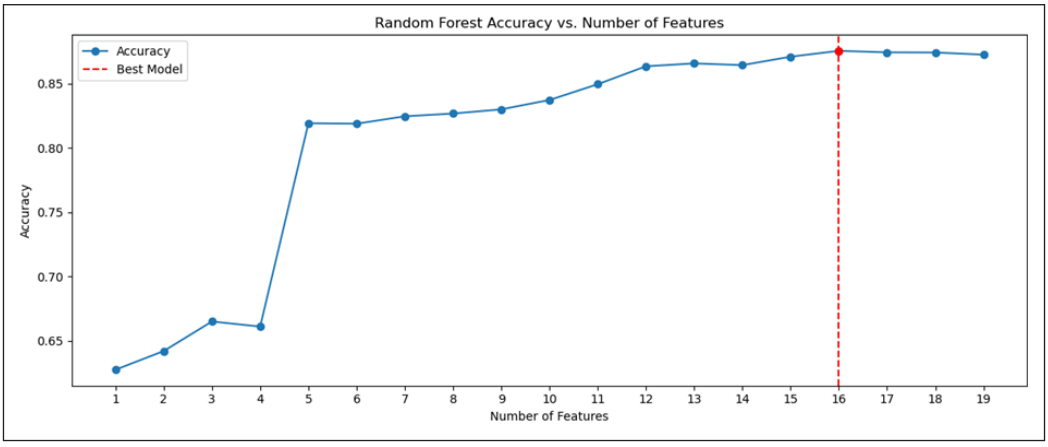


Figure 6. The prediction accuracy of the Random Forest algorithm across various feature subsets in the testing dataset. The red dotted line portrayed the optimum Random Forest algorithm with sixteen features, which had the highest accuracy score

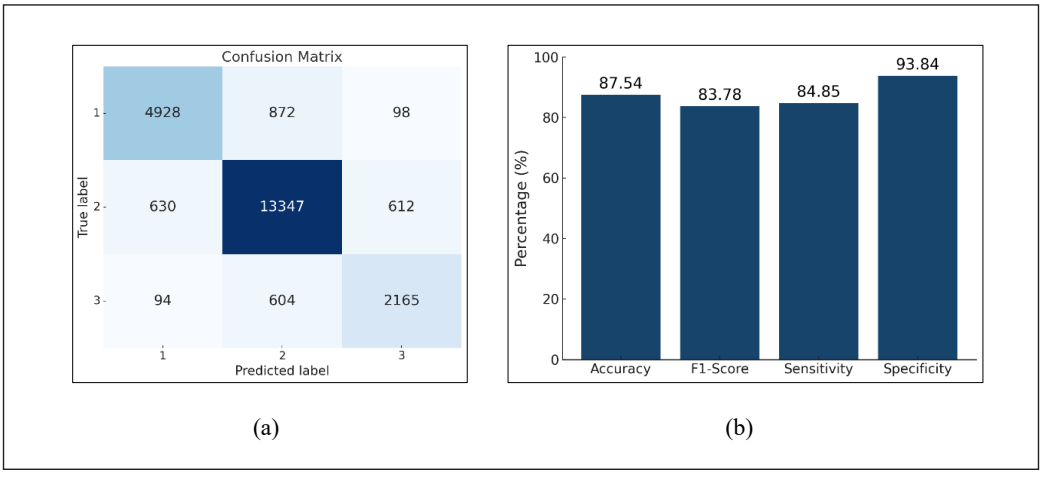


Figure 7. Random Forest classification results (a) Top 16 features confusion matrix and (b) Model evaluation of the proposed algorithm.

This reflects the nature of the MASCO occupational taxonomy, where clearly defined categories usually share similar skill profiles, rather than a weakness of the predictive model (Ministry of Human Resources, 2020).

The predictive performance for the three skill level categories is presented in Table 2. For the skilled category, the model demonstrates strong classification capabilities with an accuracy of 0.9275. The high sensitivity of 0.8719 and specificity of 0.9452 resulted in an F1-score of 0.8533.

Table 2
Predictive performance by skill level categories

Category	Accuracy	Sensitivity	Specificity	F1-score
Skilled	0.9275	0.8719	0.9452	0.8533
Semi-skilled	0.8836	0.9004	0.8543	0.9076
Low-skilled	0.9397	0.7530	0.9659	0.7546
Skill Level	0.8754	0.8485	0.9384	0.8378

This reflects that the model can distinguish skilled workers from other categories with minimal false positives. In the case of the semi-skilled category, the model registered an accuracy of 0.8836. The higher sensitivity of 0.9004 shows strong predictive ability for this category. However, the lower specificity of 0.8543 indicates a moderate false positive rate, where workers from other categories are classified as semi-skilled. For the minority category, the low-skilled category, the model achieved the highest accuracy of 0.9397 and specificity of 0.9659. This is largely due to the correct rejection of skilled and semi-skilled categories, the true negative cases. On the contrary, sensitivity decreased to 0.7530, resulting in a lower F1-score of 0.7546. This finding shows that the model can prevent false classification of other categories as low-skilled, but it is prone to missing actual low-skilled workers.

To evaluate sensitivity to hyperparameter choices, the Random Forest algorithm was estimated using alternative tree structures. These included shallow trees, which often provide good generalisation with a maximum depth of 10 (Zhou & Mentch, 2023) and regularised trees that prevent overfitting with a maximum depth of 20 and a minimum leaf size of five (Hancock & Khoshgoftaar, 2022). The same SMOTE dataset and validation dataset were utilised for all configurations. The baseline setting achieved the best performance with an accuracy of 0.8754 and an F1-score of 0.8378. Similar results were observed with moderate regularisation, where the accuracy was 0.8391. However, when using heavily constrained shallow trees, the accuracy decreased to 0.7774. The feature importance rankings remained stable across configurations with a correlation value of 0.9828, indicating consistent interpretation. These findings demonstrate that the predictive performance and SHAP explanations remain strong when the hyperparameter settings are slightly changed and do not require extensive tuning. The following segment explains how SHAP was applied to interpret the Random Forest classification model.

The Interpretability of SHAP Analysis

Although the goal is to obtain high accuracy, interpretability is essential in making policy decisions. The XAI technique, particularly SHAP analysis, was employed to interpret the impact of features on the algorithm while enhancing the model transparency (Arravalli et

al., 2025; Visibelli et al., 2025). Despite offering valuable information, the implementation of SHAP analysis requires high computational resources (Joshi et al., 2025). Integration of SHAP analysis with the proposed Random Forest algorithm offered valuable information on the most important factors impacting skill level within the Malaysian labour market. Further analysis on the illustration of skill level can be described through red dots (positive outcome) and blue dots (negative outcome) as portrayed in Figure 8. The plot of SHAP values clearly portrayed the importance of the features for predicting the skill level classification. Greater SHAP values describe higher contributions of the specific features in influencing the classification model outcomes into a particular skill level category.

SHAP analysis determined the most influential factors for predicting skill level. The summary from the SHAP global perspective presented in Figure 8(a) identified nine input predictors as the most impactful features in skill level classification. The subsets were salaries and wages activity, highest certificate, education level, working days, gender, employment status, strata and year of birth. Educational attainment represents the highest education level completed by an individual, comprising no formal education, primary, secondary and tertiary level. Meanwhile, the highest certificate is closely related but refers to specific formal qualifications, which consist of no certificate/ not applicable, SPM and below, STPM/ certificate, diploma and degree (Department of Statistics Malaysia, 2025).

Salaries and wages were demonstrated as the most dominant predictor that was strongly associated with higher skill level prediction. This pattern is consistent with the labour market scenario where skilled occupations are often associated with higher remuneration (Zhao et al., 2025). The dominance of salaries and wages in SHAP interpretability may reflect institutional and job allocation mechanisms in the labour market, rather than individual skill differences alone (Zhao et al., 2025). As a result, salaries and wages are interpreted as indicators of labour market performance instead of as direct measures of individual skill, consistent with the objective of this study.

Likewise, the type of activity performed by workers also impacted the prediction of skill level, as only certain sectors are demanding skilled workers. In the meantime, the highest certificate and education level were associated with discovering the type of skill level. It was due to skilled workers who had undergone formal education and academic criteria, which is important for career development. The findings of this study are consistent with Koltai et al. (2025), who highlighted through principal component analysis that the education level, income and occupational prestige are essential indicators in socio-economic studies, regardless of gender and age. Another five influencing predictors included working days, gender, employment status, strata and the year of birth.

Despite education level, highest certificate and salaries and wages being conceptually related, tree-based algorithms such as Random Forest are less sensitive to multicollinearity, as Random Forest can tolerate correlated input predictors and maintain its stable

classification performance (Soni et al., 2025). The SHAP values represent the conditional impact of these features on the model and do not indicate independent causal relationships. The interpretation focuses on their overall explanatory role in classifying the skill level and not on their individual effects.

Although SHAP identifies several demographic features like gender and year of birth as statistically related to the model predictions, they do not directly cause skill level classification. However, these input predictors reflect broader labour market conditions, including differences across the age groups and types of jobs. To reduce the risk of misuse in sensitive policy areas, the findings presented in this study are strictly limited to macro-level information intended for strategic human capital planning. According to the Thirteenth Malaysia Plan 2026-2030 document, human capital development is designed to support the national agenda rather than individual-level decisions (Ministry of Economy, 2025). Therefore, the findings should not be applied to individual evaluation or unfair decision processes. This is consistent with the nature of SHAP explanations in emphasising patterns in model behaviour rather than causal relationships (Joshi et al., 2025). SHAP is employed in this study as a tool to assist in explaining the model behaviour, particularly on the contribution of individual input predictors to the skill level classification (Joshi et al., 2025; Neubauer et al., 2025). Meanwhile, the economic interpretation should not rely on feature importance alone. The SHAP explanations analysed in this study are intended to support policy discussion by emphasising observed patterns in skill level model behaviour, but not to imply causal relationships for policy initiatives (Neubauer et al., 2025).

Exploring the local SHAP analysis revealed the variations across those three skill level categories. The SHAP summary showed that the most relevant predictors influencing skilled and semi-skilled categories were largely the same features. Nevertheless, the importance score varies depending on the SHAP values. As presented in Figure 8(b), the top three highest SHAP values indicated skilled category predictors were salaries and wages highest certificate and activity, whereas the semi-skilled in Figure 8(c) demonstrated highest certificate, salaries and wages and activity in determining the worker's skill. The outcomes were consistent with Jamaludin et al. (2022), which indicated salaries and wages as the primary indicators of labour market mismatch. In addition to income factors, educational attainment also played a vital role as it impacted graduate employability performance and the labour market requirements (Mansor & Othman, 2024). The activity performed by the worker was identified as one of the relevant features in occupational injury studies (Zermane et al., 2023).

Unlike the skilled and semi-skilled categories, Figure 8(d) shows that the classification of low-skilled workers did not rely on working days, year of birth and strata but rather on marital status, state and workplace. A study carried out by Kaboth et al. (2023) showed that low-skilled occupations typically did not require any vocational training, and their

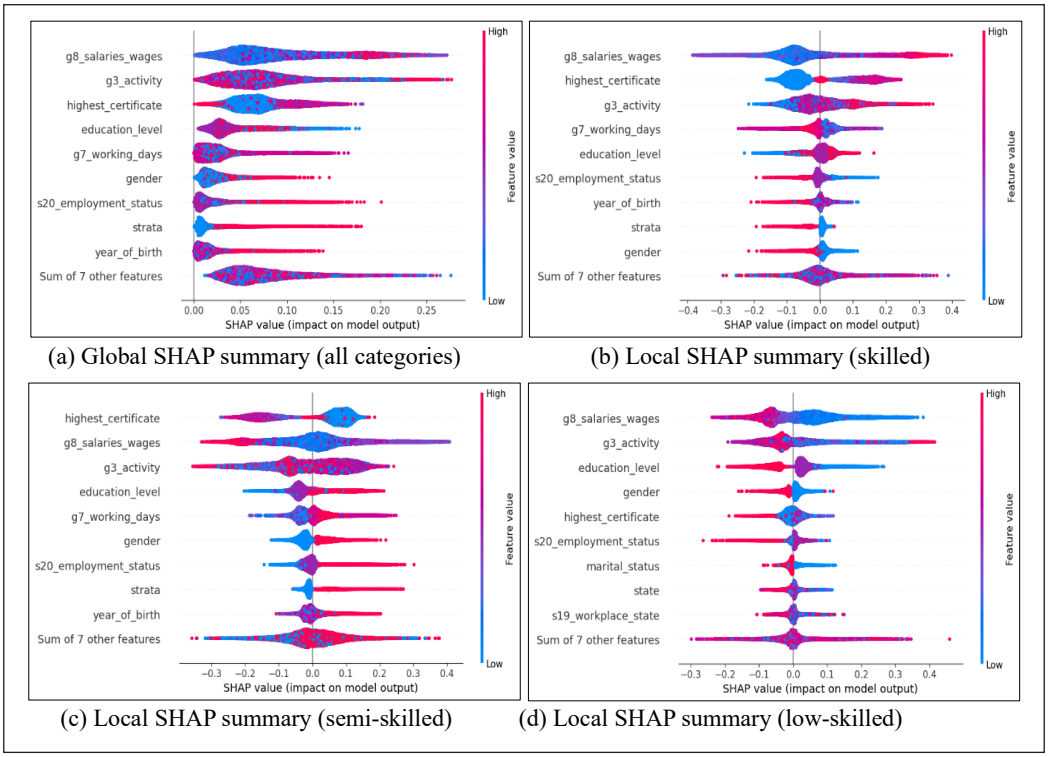


Figure 8. Summary of Global and Local SHAP (a) Global SHAP summary (all categories); (b) Local SHAP summary (skilled); (c) Local SHAP summary (semi-skilled); and (d) Local SHAP summary (low-skilled)

positions were frequently replaceable. This situation was attributed to the nature of low-skilled occupations, which typically require minimal training and experience to perform the required tasks (Kaboth et al., 2023).

The local SHAP analysis discloses how similar input predictors operate across the three skill level categories. The visualisation shows that salaries and wages received, and the highest certificate obtained are the most prominent predictors of the skilled category. Higher salaries and wages and higher certificates show a stronger positive impact on the model, demonstrating that they push the classification model towards categorising workers as skilled. In the semi-skilled category, the highest certificate and activity are the key drivers of classification. The SHAP pattern assumes that semi-skilled workers are identified by specific job activities, together with an average level of academic qualifications. For the low-skilled category, the model largely relies on the salaries and wages and education level to make predictions. The findings show that earning lower salaries and wages and a lower educational level are the main characteristics and are associated with a higher likelihood of being classified as low-skilled.

Overall, these results indicate that SHAP analysis determined salaries and wages activity and highest certificate as the main predictors in predicting skill level categories, despite their importance scores varying across the categories. Taken together, the interpretation technique from two perspectives, including global and local levels, enhances model understanding and provides beneficial insights that can be applied to labour market policy.

CONCLUSION

This study develops an interpretable skill level classification model for the Malaysian labour market by integrating the Random Forest algorithm with XAI techniques. The results of this study indicate that with the use of the top sixteen important features, the Random Forest algorithm achieves a high predictive accuracy of 0.8754. It shows a strong performance in all the evaluation metrics and thus outperformed other tree-based algorithms. Predictive accuracy combined with the incorporation of SHAP analysis also increases the trustworthiness, transparency and interpretability of the proposed model. The SHAP values determine the most affecting features by measuring both positive and negative impacts, with salaries and wages being the most influential global predictors, followed by activity and highest certificate. SHAP explanations at the local level reveal the feature contributions for each category of skilled, semi-skilled and low-skilled, offering in-depth insights into the underlying causes of classification differences. As this study uses a single cross-sectional reference year dataset, the findings are reliable and applicable to situations where the economic structure and labour market characteristics are like those in 2023.

Based on these findings, the interpretable framework supports policymakers in comprehending the rationale and justification of the predicted outcomes and enables the identification of socioeconomic, demographic, and employment characteristics that shape the classification of skill level. This information promotes evidence-based justification of human capital development and is consistent with current labour market practices. SHAP visualisations should be used as a tool to support model interpretability and interpreted together with the labour market context. It is not appropriate to be analysed alone for decision-making purposes.

However, the framework still relies on MASCO occupational groupings. The list of occupations only appears in the Labour Force/ Salaries & Wages/ Migration Survey conducted by the Department of Statistics Malaysia. Other labour market information does not contain occupational data, and the skill level cannot be measured directly. This study, therefore, focuses on identifying relevant input predictors that are strongly associated with skill level. These predictors can be applied in future studies to predict skill level using secondary data that do not capture occupational information. As skill level in this study is defined using the MASCO occupational classification, the findings should be interpreted within this definition.

To enhance the validity of this study, future studies should adopt other feature selection methods, other advanced interpretability approaches, and time-series analyses to strengthen generalisability and reflect the transitions in the Malaysian labour market. To measure the utility of policy interventions, an extended study should use a longitudinal comparison framework to evaluate whether SHAP-based interventions yield better human capital outcomes and reduce skill mismatches compared to traditional approaches. This study proposes a new framework to identify important features for skill level classification in Malaysia, addressing current approaches that rely on a single feature. Taken together, this result suggests that the combination of the Random Forest algorithm and SHAP explanations enables an accurate and interpretable prediction of skill level classification by addressing the gap between data-driven insights and the practical application in the Malaysian labour market.

ACKNOWLEDGEMENT

The authors are very grateful to the Department of Statistics Malaysia for providing the annual labour market statistics. Heartfelt appreciation is also extended to Universiti Teknologi MARA (UiTM) and the University of Leeds for their support, recommendations and guidance in the research undertaking.

REFERENCES

Adalarasu, K., Chetty, R. M. K., Begum, K. G., Harini, S., & Janardhanan, M. (2025). An explainable machine learning (XAI) framework for classification of intricate dancing posture among Indian Bharatanatyam dancers. *Applied Soft Computing, 171*, Article 112817. https://doi.org/10.1016/j.asoc.2025.112817

Arravalli, T., Chadaga, K., Muralikrishna, H., Sampathila, N., Cenitta, D., Chadaga, R., & Swathi, K. S. (2025). Detection of breast cancer using machine learning and explainable artificial intelligence. *Scientific Reports, 15*(1), Article 26931. https://doi.org/10.1038/s41598-025-12644-w

Bağcı, M., & Kaya Soylu, P. (2025). The optimal threshold selection for high-frequency pairs trading via supervised machine learning algorithms. *Computational Economics*. Advance online publication. https://doi.org/10.1007/s10614-025-10958-5

Ban, G. W., & Choi, H. (2025). Skills and labour market performance in Korea: An international comparison perspective. *International Economic Journal, 39*(2), 357-382. https://doi.org/10.1080/10168737.2025.2463372

Berge, H., Glad, I. K., Jullum, M., & Aas, K. (2024). A comparative study of methods for estimating model-agnostic Shapley value explanations. *Data Mining and Knowledge Discovery*. Advance online publication. https://doi.org/10.1007/s10618-024-01016-z

Bodilsen, S., Albeck Nielsen, S., & Rosholm, M. (2023). Measuring employment readiness for hard-to-place individuals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4646348

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Celbiş, M. G. (2023). Unemployment in rural Europe: A machine learning perspective. *Applied Spatial Analysis and Policy, 16*(3), 1071-1095. https://doi.org/10.1007/s12061-022-09464-0

Celbiş, M. G., Wong, P., Kourtit, K., & Nijkamp, P. (2023). Impacts of the COVID-19 outbreak on older-age cohorts in European labour markets: A machine learning exploration of vulnerable groups. *Regional Science Policy & Practice, 15*(3), 559-585. https://doi.org/10.1111/rsp3.12520

Danquah, L. K. G., Appiah, S. Y., Mantey, V. A., Danlard, I., & Akowuah, E. K. (2025). Computationally efficient deep federated learning with optimised feature selection for IoT botnet attack detection. *Intelligent Systems with Applications, 25*, Article 200462. https://doi.org/10.1016/j.iswa.2024.200462

Department of Statistics Malaysia. (2024a). *Labour force/salaries & wages/migration survey 2024*. https://www.dosm.gov.my/portal-main/census-surveys-list/labour-force-social-statistics

Department of Statistics Malaysia. (2024b). *Salaries & wages survey 2023*. https://www.dosm.gov.my/portal-main/release-content/salaries-and-wages-survey-report-2023

Department of Statistics Malaysia. (2025). *Salaries & wages survey report 2024*. https://www.dosm.gov.my/portal-main/release-content/salaries-and-wages-survey-report-2024

Dhungana, H. (2025). A machine learning approach for wind turbine power forecasting for maintenance planning. *Energy Informatics, 8*(1), Article 2. https://doi.org/10.1186/s42162-024-00459-4

Dossche, W., Vansteenkiste, S., Baesens, B., & Lemahieu, W. (2025). Anticipating delays in recruitment: Explainable machine learning for the prediction of hard-to-fill online job vacancies. *European Journal of Operational Research*. Advance online publication. https://doi.org/10.1016/j.ejor.2025.06.027

Elshewey, A. M., Selem, E., & Abed, A. H. (2025). Improved CKD classification based on explainable artificial intelligence with extra trees and BBFS. *Scientific Reports, 15*(1), Article 17861.https://doi.org/10.1038/s41598-025-02355-7

Halias, A. F., Saiful, N. H., Ibrahim, N., Muhamad Jamil, S. A., Mansor, M. M., Ul-Saufie, A. Z., & Md Ghani, N. A. (2023). Type 2 diabetes mellitus prediction using data mining approach. In *2023 IEEE International Conference on Computing (ICOCO)* (pp. 29-34). https://doi.org/10.1109/ICOCO59262.2023.10398078

Hancock, J., & Khoshgoftaar, T. M. (2022). Optimising ensemble trees for big data healthcare fraud detection. In *Proceedings of the International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 65-72). https://doi.org/10.1109/IRI54793.2022.00061

Hasan, Y., de Lima, A., Namjoo, E., de Bulnes, D. F., Albarracín, J. F. H., & Ryan, C. (2025). Improving breast cancer diagnosis using grammatical evolution-based feature selection. *SN Computer Science, 6*(4), Article 306. https://doi.org/10.1007/s42979-025-03840-6

Hsieh, P.-H. (2025). ML-based validation of experimental randomisation in learning games. *Frontiers in Artificial Intelligence, 8*, Article 1541087. https://doi.org/10.3389/frai.2025.1541087

Jamaludin, S., Safian, S. S., Said, R., & Muhamed, A. A. (2022). The incidence of overeducated and undereducated worker for manufacturing sector in Malaysia. *International Journal of Academic Research in Business and Social Sciences, 12*(5), 810-823. https://doi.org/10.6007/IJARBSS/v12-i5/12942

Joshi, A., Pradhan, B., Chakraborty, S., Varatharajoo, R., Alamri, A., Gite, S., & Lee, C.-W. (2025). An explainable Bi-LSTM model for winter wheat yield prediction. *Frontiers in Plant Science, 15*, Article 1491493. https://doi.org/10.3389/fpls.2024.1491493

Kaboth, A., Hünefeld, L., & Himmelreicher, R. (2023). Employment trajectories of workers in low-skilled jobs in Western Germany. *Journal for Labour Market Research, 57*(1), Article 21. https://doi.org/10.1186/s12651-023-00351-w

Kaur, M., Valderrama, C. E., & Liu, Q. (2025). Hybrid vision transformer framework for efficient and explainable SEM image-based nanomaterial classification. *Machine Learning: Science and Technology, 6*(1), Article 015066. https://doi.org/10.1088/2632-2153/adc072

Kempny, C., Annac, K., Wahidie, D., Yilmaz-Aslan, Y., & Brzoska, P. (2025). Validation of stop criteria in ASReview for optimising title and abstract screening. *European Journal of Public Health, 35*(Supplement_4). https://doi.org/10.1093/eurpub/ckaf161.1276

Khalfallah, S., Puech, W., Tlija, M., & Bouallegue, K. (2025). Exploring the effectiveness of machine learning and deep learning techniques for EEG signal classification in neurological disorders. *IEEE Access, 13*, 17002-17015. https://doi.org/10.1109/ACCESS.2025.3532515

Koltai, J., Rakovics, Z., Kmetty, Z., Számel, K., Ungvári, B., Váradi, B., & Huszár, Á. (2025). Classifying social position with social media behavioural data. *EPJ Data Science, 14*(1), Article 60. https://doi.org/10.1140/epjds/s13688-025-00578-2

Kurniadi, F. I., & Saputri, H. A. (2024). Feature selection using XGBoost on METABRIC dataset for survivability breast cancer detection. In *2024 International Conference on Information Management and Technology (ICIMTech)* (pp. 259-262). https://doi.org/10.1109/ICIMTech63123.2024.10780791

Lalithadevi, B., & Krishnaveni, S. (2025). ExAIRFC-GSDC: An advanced machine learning-based interpretable framework for accurate gas leakage detection and classification. *International Journal of Computational Intelligence Systems, 18*(1), Article 16. https://doi.org/10.1007/s44196-025-00742-6

Lee, J.-Y., Lee, W., & Cho, S. (2023). Characteristics of fatal occupational injuries in migrant workers in South Korea: A machine learning study. *Heliyon, 9*(9), Article e20138. https://doi.org/10.1016/j.heliyon.2023.e20138

Liu, J., & Wang, H. (2025). Impact of personal information and reputation system on gig workers' employment status: An interpretable machine learning-based approach. *Journal of Computational Social Science, 8*(3), Article 59. https://doi.org/10.1007/s42001-025-00371-1

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (arXiv:1705.07874). *arXiv*. https://doi.org/10.48550/arXiv.1705.07874

Manafi Varkiani, S., Pattarin, F., Fabbri, T., & Fantoni, G. (2025). Predicting employee attrition and explaining its determinants. *Expert Systems with Applications, 272*, Article 126575. https://doi.org/10.1016/j.eswa.2025.126575

Mansor, A., & Othman, Z. (2024). Malaysian community college graduates' employability prediction model using a machine learning approach. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 16*(3), 1-7. https://doi.org/10.54554/jtec.2024.16.03.001

Méndez-Astudillo, J. (2024). The impact of comorbidities and economic inequality on COVID-19 mortality in Mexico: A machine learning approach. *Frontiers in Big Data, 7*, Article 1298029. https://doi.org/10.3389/fdata.2024.1298029

Ministry of Economy. (2025). *Thirteenth Malaysia plan 2026-2030*

Ministry of Human Resources. (2020). *Malaysia standard classification of occupations 2020*.

Mitchell, R., Frank, E., & Holmes, G. (2022). GPUTreeShap: Massively parallel exact calculation of SHAP scores for tree ensembles. *PeerJ Computer Science, 8*, Article e880. https://doi.org/10.7717/peerj-cs.880

Mujahid, M., Kına, E., Rustam, F., Villar, M. G., Alvarado, E. S., Diez, I. D. L. T., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data, 11*(1), 1-32. https://doi.org/10.1186/s40537-024-00943-4

Neubauer, A., Brandt, S., & Kriegel, M. (2025). Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. *Energy and AI, 20*, Article 100480. https://doi.org/10.1016/j.egyai.2025.100480

Ogungbire, A., & Pulugurtha, S. S. (2025). Spatiotemporal risk mapping of statewide weather-related traffic crashes: A machine learning approach. *Machine Learning with Applications, 20*, Article 100642. https://doi.org/10.1016/j.mlwa.2025.100642

Park, J., Choi, S., & Feng, Y. (2024). Predicting startup success using two bias-free machine learning: Resolving data imbalance using generative adversarial networks. *Journal of Big Data, 11*(1), Article 122. https://doi.org/10.1186/s40537-024-00993-8

Pedersen, S. M., Damslund, N., Kjær, T., & Olsen, K. R. (2025). Optimising test intervals for individuals with type 2 diabetes: A machine learning approach. *PLOS ONE, 20*(2), Article e0317722. https://doi.org/10.1371/journal.pone.0317722

Pourdaryaei, A., Mohammadi, M., Mubarak, H., Abdellatif, A., Karimi, M., Gryazina, E., & Terzija, V. (2024). A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market. *Expert Systems with Applications, 235*, Article 121207. https://doi.org/10.1016/j.eswa.2023.121207

Rabi'atul'adawiah Shablia, Ul-Saufie, A. Z., & Ibrahim, N. (2025). Improving the prediction of the labour skill classification model in Malaysia using tree-based machine learning algorithms. *Malaysian Journal of Fundamental and Applied Sciences, 21*(5), 2771-2789. https://doi.org/10.11113/mjfas.v21n5.4435

Rafrastara, F. A., Ghozi, W., Sani, R. R., Handoko, L. B., Abdussalam, A., Pramudya, E. R., & Abdollah, F. M. (2025). Integrating information gain and chi-square for enhanced malware detection performance. *Journal of Information and Communication Technology, 24*(1), 80-104. https://doi.org/10.32890/jict2025.24.1.4

Sakyi-Yeboah, E., Agyemang, E. F., Agbenyeavu, V., Osei-Nkwantabisa, A., Kissi-Appiah, P., Moshood, L., Agbota, L., & Nortey, E. N. N. (2025). Heart disease prediction using ensemble tree algorithms: A supervised learning perspective. *Applied Computational Intelligence and Soft Computing, 2025*(1), Article 1989813. https://doi.org/10.1155/acis/1989813

Salah-Ud-Din, M., S. S., B. T. L., & Al Ali, H. (2024). Exploratory data analysis and prediction of passenger satisfaction with airline services. In *2024 New Trends in Civil Aviation (NTCA)* (pp. 295-302). https://doi.org/10.23919/NTCA60572.2024.10517814

Sani, J., Halane, S., Ahmed, A. M., & Ahmed, M. M. (2025). Application of machine learning algorithms and SHAP explanations to predict fertility preference among reproductive women in Somalia. *Scientific Reports, 15*(1), Article 26301. https://doi.org/10.1038/s41598-025-04704-y

Sariman, S. R., Ab Jalil, H., & Marlisah, E. (2024). Prediction model of school dropout factors using classification techniques in Selangor. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 9*(6), Article e002867.https://doi.org/10.47405/mjssh.v9i6.2867

Shaha, T. R., Begum, M., Uddin, J., Torres, V. Y., Iturriaga, J. A., Ashraf, I., & Samad, M. A. (2024). Feature group partitioning: An approach for depression severity prediction with class balancing using machine learning algorithms. *BMC Medical Research Methodology, 24*(1), Article 123. https://doi.org/10.1186/s12874-024-02249-8

Shapley, L. S. (1953). *A value for n-person games*. Defence Technical Information Centre. https://doi.org/10.1515/9781400881970-018

Sheokand, S., & Beniwal, S. (2025). Credit card fraud detection using machine learning and deep learning techniques: A review. *Pertanika Journal of Science and Technology, 33*(6). https://doi.org/10.47836/pjst.33.6.11

Soni, K., Frew, R., & Kebede, B. (2026). Interpretable machine learning for origin classification of Brazilian soybeans: A random forest and XAI-based approach. *Journal of Food Composition and Analysis, 150*, Article 108896. https://doi.org/10.1016/j.jfca.2026.108896

Syeda, D. Z., & Asghar, M. N. (2023). Enhancing malware classification: A comparison of filter and embedded feature selection with a novel dataset. In *Proceedings of the Cyber Research Conference* (pp. 1-8). https://doi.org/10.1109/Cyber-RCI59474.2023.10671445

Tsai, C.-F., Chen, K.-C., & Lin, W.-C. (2024). Feature selection and its combination with data over-sampling for multi-class imbalanced datasets. *Applied Soft Computing, 153*, Article 111267. https://doi.org/10.1016/j.asoc.2024.111267

Ul Haq, E., Yong, Q., Yuan, Z., Jianjun, H., Ul Haq, R., & Qin, X. (2024). Accurate multiclassification and segmentation of gastric cancer based on a hybrid cascaded deep learning model with a vision transformer from endoscopic images. *Information Sciences, 670*, Article 120568. https://doi.org/10.1016/j.ins.2024.120568

Unlu, A., & Subasi, A. (2025). Substance use prediction using artificial intelligence techniques. *Journal of Computational Social Science, 8*(1), Article 21. https://doi.org/10.1007/s42001-024-00356-6

Visibelli, A., Finetti, R., Roncaglia, B., Poli, P., Spiga, O., & Santucci, A. (2025). Predicting therapy dropout in chronic pain management: A machine learning approach to cannabis treatment. *Frontiers in Artificial Intelligence, 8*, Article 1557894. https://doi.org/10.3389/frai.2025.1557894

Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data, 11*(1), Article 44. https://doi.org/10.1186/s40537-024-00905-w

Wang, P., Chen, T.-H., Chang, M.-Y., Hsia, H.-Y., Dai, M., Liu, Y., Hsu, Y.-L., Fu, F., & Zhao, Z. (2025). An explainable artificial intelligence framework for weaning outcomes prediction using features from electrical impedance tomography. *Computer Methods and Programs in Biomedicine, 267*, Article 108811. https://doi.org/10.1016/j.cmpb.2025.108811

Wang, Z., He, C., Hu, Y., Luo, H., Li, C., Wu, X., Zhang, Y., Li, J., & Cai, J. (2024). A hybrid deep learning scheme for MRI-based preliminary multiclassification diagnosis of primary brain tumours. *Frontiers in Oncology, 14*, Article 1363756. https://doi.org/10.3389/fonc.2024.1363756

Zahra, S. T., Imdad, S. K., Khan, S., Khalid, S., & Baig, N. A. (2025). Power transformer health index and life span assessment: A comprehensive review of conventional and machine learning-based approaches. *Engineering Applications of Artificial Intelligence, 139*, Article 109474. https://doi.org/10.1016/j.engappai.2024.109474

Zeffora, J., & Shobarani, S. (2022). Optimising random forest classifier with Jenesis-index on an imbalanced dataset. *Indonesian Journal of Electrical Engineering and Computer Science, 26*(1), 505-511. https://doi.org/10.11591/ijeecs.v26.i1.pp505-511

Zermane, A., Mohd Tohir, M. Z., Zermane, H., Baharudin, M. R., & Mohamed Yusoff, H. (2023). Predicting fatal fall from heights accidents using random forest classification machine learning model. *Safety Science, 159*, Article 106023. https://doi.org/10.1016/j.ssci.2022.106023

Zhao, G., & Li, D. (2025). TreeEM: Tree-enhanced ensemble model combining with feature selection for cancer subtype classification and survival prediction. *Results in Applied Mathematics, 27*, Article 100605. https://doi.org/10.1016/j.rinam.2025.100605

Zhao, X., Wu, D., & Wang, L. (2025). Wage effect and workers' skill mismatch: Paths and mechanisms. *SAGE Open, 15*(1), Article 21582440251318540. https://doi.org/10.1177/21582440251318540

Zhou, S., & Mentch, L. (2023). Trees, forests, chickens, and eggs: When and why to prune trees in a random forest. *Statistical Analysis and Data Mining, 16*(1), 45-64. https://doi.org/10.1002/sam.11594

Pertanika J. Sci. & Technol. 34 (2): 1231 - 1263 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

How Gamification Motivates in Quranic Memorisation: A Method for Design and Suggestions for Future Research

Siti Hasrinafasya Che Hassan[1,2]*, Syadiah Nor Wan Shamsuddin[2], and Nor Hafizi Yusof[3]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Branch, 22200 Besut, Terengganu, Malaysia
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kelantan Branch, Bukit Ilmu, 18500 Machang, Kelantan, Malaysia
[3]Faculty of Islamic Contemporary Studies, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia

ABSTRACT

Building enthusiasm for memorising the Quran has become challenging, as interest and motivation have been found to decline. However, there is still no clear guideline on how to combine the different gamification components systematically within the environment of Quranic memorisation (Hifz), despite the significant role of gamification in the learning process. The proposed research employs the Design and Development Research (DDR) methodology, aiming to create a learning content prototype that ideally integrates gamification components and learning analytics into Quranic learning and evaluation activities. The study has four objectives structured within the three phases of the DDR method. The first phase, analysis, aims to achieve two research objectives: integrating different multimedia components within memorisation strategies and identifying the appropriate components of the gamified process. The second phase, design and development, involves designing a gamified learning application within the MDA framework. The third phase, evaluation, examines how gamification elements influence students' motivation and engagement. A survey was distributed to 46 undergraduates. The respondents' perceptions of using a gamified system yielded high overall mean scores ranging from 4.55 to 4.91 on a 5-point Likert scale. The standard deviations ranged from 0.29 to 0.66, indicating low variability and minimal disagreement. The reliability test revealed very high internal consistency of all the items, with Cronbach's alpha values ranging from 0.965 to 0.970. This research has contributed to the field of Islamic education by demonstrating

ARTICLE INFO

Article history:
Received: 23 October 2025
Accepted: 13 April 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.26

E-mail addresses:
hasrina581@uitm.edu.my (Siti Hasrinafasya Che Hassan)
syadiah@unisza.edu.my (Syadiah Nor Wan Shamsuddin)
nhafizi@unisza.edu.my (Nor Hafizi Yusof)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

the potential of gamification to make the memorisation of the subjects more interactive, engaging and effective.

Keywords: Engagement, gamification, learning analytics, motivation, Quranic memorisation

INTRODUCTION

The growing number of learning institutions offering educational services to learn the Quran through memorisation shows that the Muslim community's level of interest to memorise the Quran has escalated remarkably (Haryono et al., 2022). Hifz denotes the activity involved in learning 30 chapters or parts of the Quran. Memorising the Quran has challenges and difficulties that require a systematic, effective and continuous repetition technique to prevent the memoriser from forgetting (Ahmad et al., 2024). It has been acknowledged that memorising the Quran is not something that can be done easily by every Muslim (Aziz et al., 2019). According to Misnan et al., (2021), during the process of Quran memorisation, the students may experience different stages of learning, which include the acquisition stage, the retaining stage and the retrieval stage. The acquisition stage is where the students are supposed to devote all their attention to the stimulation by encoding it into their memory. The retaining stage is the process whereby the minds of the students can retain the stimulation. The retrieval stage is the process of obtaining stimulation from memory whenever needed.

Unfortunately, according to Misnan et al. (2021) as cited in Ahmad et al. (2024) not all Quran learners were able to retain their memorisation. Ambo and Mokhsein (2019) also reported that most of the Quran learners are unable to memorise all 30 chapters of the Quran within 3 years. Many complain about forgetting the memorised verses (Almosallam et al., 2016; Almosallam, 2015) and many of them find it difficult to maintain the quality of their memorisation once it is completed, despite having spent a great deal of time learning and repeating them. The daily routine of memorising different verses and the consistent repetition of the recitation of the verses resulted in the reciters feeling bored (Sabbri, 2016; Mustafa & Basri, 2014; Sevutra & Erlinda, 2019). Among the causes of students' failure to memorise and maintain good memorisation are the lack of motivation and the lack of effective learning style strategies (Abdullah et al., 2021; Ahmad et al., 2024; Rahman et al., 2022), as well as the lack of appropriate teaching strategies that can facilitate students' interest in the learning process (Musa et al., 2018). As mentioned by Misnan et al. (2021), the best memorisation method and technique should be developed to encourage students to memorise the Quran. These factors have been identified as the cause of weak Tahfiz achievements, which hinder students from directly engaging in the process of memorising the Quran, believed to harm students' memorisation.

Urh et al. (2015) introduced the model for the implementation of gamification in the e-learning domain for achieving a positive outcome on the process of learning, such as increased satisfaction, motivation and participation of the students. In the systematic literature review by Akbari et al. (2024), it is stated that there are three types of Quran application models, namely, voice technology-based model, gamification-based model and modular-based model. In terms of the model's effectiveness, it was revealed that the gamification-based model is effective for supporting the process of learning. The next section discusses the potential integration of gamification in the field of Quranic memorisation.

The Potential of Motivation and Engagement in Quranic Memorisation with Gamification

Gamification is essential for increasing students' motivation and engagement. It seeks to combine intrinsic motivation with extrinsic factors to enhance motivation and engagement (Noradzan et al., 2019). Motivation and engagement are key values in learning and among the most desired experiential qualities for memorising the Quran. Currently, gamification in mobile learning is increasing and becoming popular in response to the needs of Education 4.0 within the context of the Industrial 4.0 (IR4.0), as it aligns with the motivation and engagement of the younger generation in self-directed learning. One important trend in Education 4.0 that has been frequently discussed is students' preference for learning tools that adapt to their abilities and achievements (Hussin, 2018). This allows students who encounter difficulties with a particular subject to practice more until they reach the required level.

Gamification involves using game design elements in non-game contexts to stimulate and encourage individuals to pursue their goals. By integrating these strategies, memorisation of the Quran can be more participatory, motivational, and effective, enabling students not only to reach their memorisation goals but also to connect more deeply with the Quran. Various previous studies, such as the work by Shamsuddin et al. (2016), postulated a conceptual framework called e-Hafiz for modelling the memorisation process of the Quran using gaming components such as avatar, level, points, timing and awards. According to Moulana (2017), the conceptual framework aimed to provide a more engaging and motivational experience in the Quran memorisation process by implementing gaming components, such as points, levels, challenges, achievements, and narratives, in addition to traditional memorisation strategies. Senan et al. (2017) introduced an interactive Quran mobile app called iHafiz, incorporating gamification aspects such as levels, points and rewards to assist autistic children in reciting and memorising the Quran. Rosmansyah & Rosyid (2017) implemented gamification aspects such as badges, points, levels, leaderboard and progress bar in a mobile learning environment for memorising the Quran, aiming to

enhance the fun aspect. By applying the principle of gamification, such as levels, challenges, points, badges, leaderboard or ranking, Sevutra & Erlinda (2019) were able to make the task of memorising the Quran into interesting games. This would enable them to be monitored in memorising and earning points towards ranking for the motivation of memorising the Quran. To benefit students in memorising the Quran and enhancing their understanding in reading the Quran extensively, the iLearning Management System was designed by Pranata et al. (2020) with gamified quiz activities which include levels, challenges, points, awards, progress bar and narratives. The QuranLab mobile application was designed by Siregar et al. (2023) based on the Octalysis gamification framework. Anwar & Inayati (2025) studied the utilisation of Wordwall as a feasible tool for encouraging increased engagement of students in the study of Al-Quran Hadith. In a separate study, Ni'mah et al. (2025) highlighted the implementation of gamification technology learning using Wordwall and Gimkit platforms to aid the learning and assessment processes in the study of the Quran and Hadith. Zein & Syahrudin (2025) proposed the use of gamification elements such as points, badges, leaderboards, and interactive challenges in the traditional method of Quranic study to improve engagement, motivation and learning outcomes.

However, many students who attempt to memorise the Quran do not achieve the desired objectives due to non-compliance and a lack of user satisfaction with the learning experience, resulting in some users discontinuing using the learning application after their initial experience. The traditional "one size fits all" gamification approach fails to consider the variety of users' needs, talents, and preferences (Alomair & Hammami, 2020). Therefore, the authors of this study contend that gamification does not always produce predictable results. Another issue is that many studies treat gamification as a single, uniform concept, even though in practice, gamification environments can feature a wide range of individual designs and implementations. As gamification can incorporate game design components in various ways and take many distinct forms, it is inappropriate to study the motivational benefits of gamification as a general construct. Instead, research should focus on how specific game design components affect a particular setting. Gamification scholars argue that gamification is not simply a matter of adding PBL (points, badges, and leaderboards) to a product or service. Instead, gamification involves a broader discussion, including motivational theories, to foster intrinsic motivation as a task is carried out (Yahya & Salamun, 2021).

This research aims to create a learning content prototype that ideally integrates gamification components and learning analytics into Quranic learning and evaluation activities. The following objectives were formulated to investigate the relevant research questions, which are stated below:

RQ1 : What are the cognitive principles and multimedia elements that can improve memorisation strategies?

RO1 : To integrate the multimedia elements into memorisation strategies that can reduce the cognitive load in memorising the Quran.

RQ2 : What are the appropriate components of gamification that can be integrated to support sustainable motivational learning theory in memorising the Quran?

RO2 : To identify the appropriate gamification elements that can be integrated to facilitate sustainable motivational learning theory in memorising the Quran.

RQ3 : How can we design and develop a gamified Quranic memorisation application that influences students' motivation and engagement?

RO3 : To design and develop a gamified Quranic memorisation application that affects students' motivation and engagement.

RO4 : To evaluate the effects of gamification elements on the students' motivation and engagement in Quranic memorisation.

The remainder of this paper is structured as follows. Section 2 provides a brief literature review on the integration of multimedia elements in memorisation strategies, the adaptation of motivational theory in the context of gamification for Quranic memorisation, and the relevant studies in Quranic memorisation to identify potential behaviours that may be influenced when learning with appropriate gamification features. Section 3 discusses the design principles for the gamification of Quranic memorisation, along with the development of the proposed GLAM-Q application displays. Section 4 covers the findings from a survey on the applicability of the suggested game elements. Finally, Section 5 presents concluding remarks and highlights the research outcomes.

LITERATURE REVIEW

Integration of Multimedia Elements in Memorisation Strategies

Memorisation is a mental process used to recall and retain learning. Its success depends on an individual's ability to retrieve information from memory spontaneously, either with or without cues (Salehuddin et al., 2019). However, according to Schunk (2012), during cognitive load, a student can only memorise a limited amount of knowledge at one time. For example, it is commonly agreed that short-term memory (STM) holds information for a brief period, approximately 2 seconds (Saleem, 2015), if a systematic and effective rehearsal mechanism is not used to transfer the information to long-term memory (LTM). Additionally, individuals exposed to numerous simultaneous stimuli may miss most of them due to the limited attentional capacity of working memory (WM).

Cognitive load theory (CLT) accounts for these processing constraints in instructional design. CLT emphasises the idea of internal mental processes and tackles the problem of how information is received, organised, stored, and retrieved in one's mind (Schunk, 2012). In addition, it has been emphasised that the primary focus of CLT is to ensure that the vast

store of knowledge contained in LTM is optimally acquired. As a result, cognitive load theorists devote much of their attention to determining how knowledge should be organised when presented to learners and which activities learners should engage in when acquiring information. Therefore, ensuring that learners' working memory (WM) is not overburdened with the information presented is one goal of CLT-based instruction. Creating external supporting structures to facilitate the transfer of information into LTM through learning strategies can significantly impact learning outcomes, including enjoyment, engagement, motivation, results and achievements, satisfaction, and other positive attitudes towards learning, particularly in memorisation. Consequently, it can be inferred that from the viewpoint of Quranic memorisation, CLT focuses more on how knowledge can be encoded, stored, and retrieved. For example, in memorising the Quran, the cognitive approach will include step-by-step descriptions with strategies relevant to its approach, employing images or pictures relevant to the interpretation of verses, requiring the students to recite as many verses as they can, as well as providing continuous learning feedback from the formative and summative assessments to strengthen their memorisation progression. Likewise, the cognitive strategies for learning, for example, elaboration, organisation, and rehearsal, are essential in enhancing memorising tasks such as Quranic memorisation. Table 1 presents several cognitive principles associated with memorisation techniques or strategies that may alleviate cognitive load during Quran memorisation.

According to Umar & Aziz (2015), multimedia features assist in creating and storing information using different human senses, known as multisensory, for a long period of time and are quickly accessible to attract the attention of students and make them focus on acquiring knowledge. Mustafa et al. (2019) proposed in their research a VARK Quran memorisation model consisting of four main components, namely, VARK learning style

Table 1
Application of cognitive principles in Quranic memorisation

Cognitive principle	Application example in Quranic memorisation
Chunking	Breaking down long verses into smaller phrases for easier recall (Almosallam et al., 2016; Almosallam, 2015; Moulana, 2017; Mustafa et al., 2019; Shamsuddin et al., 2016)
Repetition and rehearsal	Regularly reviewing memorised verses previously will enhance memory retention (Almosallam et al., 2016; Almosallam, 2015; Moulana, 2017); Mustafa et al., 2019; Rosmansyah & Rosyid, 2017; Shamsuddin et al., 2016; Senan et al., 2017).
Use of mnemonics and imagery	Associating certain sounds or meanings with visual or auditory cues to aid recall (Moulana, 2017; Mustafa et al., 2019; Senan et al., 2017).
Scaffolding	Starting with shorter Surahs and gradually progressing to longer ones with support (Moulana, 2017; Mustafa et al., 2019).
Organising knowledge	Using mind maps or verse structures to show the relationship between themes in Surah (Almosallam et al., 2016; Almosallam, 2015; Mustafa et al., 2019).

modes, sensory memory, short-term memory and long-term memory. The VARK learning style model serves as an input that has four modes, namely, visual, aural, read/write and kinesthetic. These modes of input are sensed through sensors, such as the eye, ear, and body motion/gesture, in the sensory memory. These are eventually converted into image, audio, text and signal representations in the short-term memory. Finally, to retain memory in the long-term memory, there are several memorisation methods that must all be employed as mentioned in Mustafa et al. (2021), and in Table 1 above. Since the educational content should be interactive and engaging, all these combinations may well be attributed to the contribution of multimedia technology in the learning and memorising process of the Quran readings.

Adaptation of Motivational Theory in the Context of Gamification for Quranic Memorisation

Motivation is often considered a prerequisite for sustainable learning in Quranic memorisation. One promising approach shown to support learning and motivation in Quranic memorisation is the use of games and game elements. Gamification has been identified as an effective tool for encouraging students to participate enthusiastically in learning activities (Ishak et al., 2022). When games have an educational purpose and are not played primarily for amusement, they may be called serious games. The use of game design elements in non-game contexts, primarily to motivate desired user behaviour, is known as gamification. Students can remember 90% of the content if they engage with or participate in gamified educational materials or simulations (Noradzan et al., 2019).

Gamification has become a popular strategy in this new digital age to encourage specific behaviours and have a positive impact on learning outcomes in many ways, such as motivation (Noradzan et al., 2019; Nurtanto et al., 2021; Yildiz et al., 2021; Yushaa et al., 2021), engagement (Noradzan et al., 2019; Nurtanto et al., 2021; Yushaa et al., 2021), outcomes and achievements (Karamert & Vardar, 2021; Yushaa et al., 2021), enjoyment (Noradzan et al., 2019; Yushaa et al., 2021), and other positive attitudes. According to Rahman et al. (2018), achievement and engagement of students are the most common areas of research, whereas motivation ranks third. High frequency across the three components affirms that gamification can transform students into more motivated and engaged learners in class, thereby enabling them to perform at the highest possible level in the assessment of their learning. It also demonstrates that gamification can catalyse engagement and motivation (both intrinsic and extrinsic), which are indispensable for creating quality learning and are usually considered prerequisites for completing a task or encouraging specific behaviour, such as the time and effort learners invest in learning.

To elucidate motivation and engagement, researchers have drawn on motivation theory (proposed by Thomas Malone in 1981) and the game flow model (Sweetser & Wyeth, 2005)

to guide the design and development of gamification for Quranic memorisation. In simple terms, the concept of flow describes the mental state of a person who is fully immersed, engaged, and enjoying an activity. However, the activity must be regular and match the learner's abilities; that is, the perceived challenge should balance the actual level of difficulty. Previous studies have shown that when challenges are too easy or too difficult for the user's skill level, they can lead to frustration and anxiety during the activity. Second, regarding the learning goal, the game should provide immediate feedback on what must be completed and on the learner's progress. Intrinsic motivation is essential to ensure that users enjoy the game and remain engaged and immersed while interacting with it, which can affect all phases of learning and performance.

Game Element-affected Behaviours

These relevant studies on the memorisation of the Quran are reviewed in this section to determine the possible outcomes of gamification features that may affect learning. Possible behaviours include engagement, achievement, and motivation. The gamification components used to influence learning outcomes are shown in Table 2.

Based on the studies discussed, the majority of the gamified Quranic memorisation apps have employed the main components of gamification such as points, badges, levels, challenges, achievements, progress bar and leaderboard. However, they are not considering structured learning activities incorporated with formative and summative assessments. Employing gamification elements in assessments ensures that students remain actively engaged while being tested, thus reducing the usual pressure caused during the process of traditional assessment (Zulkefli & Jamil, 2025).

Table 2
List of recent articles related to gamification in Quranic memorisation from 2016 to 2025

Researcher(s)	Research focus	Gamification elements	Affected learning outcome
Zein & Syahrudin (2025)	To examine the concept of gamification for the learning of the Quran. The research proposed integrating gamification elements such as points, badges, leaderboards, and challenges with the traditional learning process. Nevertheless, the development of gamification-based applications has yet to be digitised.	Points, Badges, Leaderboard, Challenge	Engagement Motivation
Ni'mah et al. (2025)	To explore the effectiveness of gamification in engaging students' interest and motivation in learning the Quran Hadith through two applications, namely Wordwall and Gimkit. Both applications are based on gamification, to enable teachers to develop different learning games such as quizzes, puzzles and other interactive learning exercises.	Points, Levels, Awards, Challenges, Competition	Engagement Motivation

Table 2 (*continued*)

Researcher(s)	Research focus	Gamification elements	Affected learning outcome
Anwar & Inayati (2025)	To investigate Wordwall's effectiveness as a means of encouraging student engagement in learning the Al-Quran Hadith. This non-monotonous and interactive game-based learning platform provides 18 various game features, including pairing questions with definitions, guessing questions in boxes, dragging images to identify labels, grouping objects into categories, multiple-choice in quiz format, matching images with descriptions, dragging words into blank spaces and so on.	Challenges Feedback	Engagement Motivation
Ahmad et al. (2024)	In this study, the learning approach for the revision of the Quranic text involves the use of a fun-learning gamification board (which is not a digital application).	Levels, Points, Awards, Challenges, Status (similar as progress bar)	Achievement
(Siregar et al. 2023)	This study explored the effects of gamification on learning the Quran by using the QuranLab application. The application is based on the Waterfall model and incorporates gamification based on Octalysis framework. The effectiveness of the application has been tested through a descriptive analysis of 40 people between the ages of 12 and 22 years by purposive sampling techniques, by using the Hedonic-Motivation System Adoption Model (HMSAM) questionnaire.	Levels, Rewards / Missions, Badges, Points, Coins, Leaderboard Progress bar, Progress tracker, Feedback, Share button Timer	Engagement Motivation
Rahman et al. (2022)	This research aimed at assessing the effectiveness of the Global Tahfiz Game (GTG) in enhancing the students' comprehension of Tajwid and Tahfiz, comprising three sets of questions with differing levels of difficulty based on the 30 Juzuk of the Quran.	Levels	Achievement Motivation
Pranata et al. (2020)	To support students in memorising the Quran and improving their reading comprehension, the iLearning Management System has been developed, offering a range of information, including learning resources, tasks, and quiz activities (with gamification elements).	Levels, Challenges, Points, Awards, Narratives	Motivation Achievement Engagement
Fanani et al. (2019)	Evaluation of usability aspects (usefulness, ease of use, ease of learning, and satisfaction) of a mobile-based application for Al-Quran writing learning with gamification.	Levels, Challenges, Points, Leaderboard	Engagement Achievement

Table 2 (*continued*)

Researcher(s)	Research focus	Gamification elements	Affected learning outcome
Sevutra & Erlinda (2019)	By adhering to the principle of Gamification – turning activities into engaging games – makes it easier to monitor memorisation and accumulate points for ranking, thereby driving enthusiasm and motivation to memorise the Quran.	Levels, Challenges, Points, Badges, Ranking	Motivation
Noor et al. (2019)	To utilise the gamification and augmented reality as an attempt to attract children to learn Tajweed.	Challenges, Points, Badges, Leaderboard, Progression, Immersion, Feedback	Engagement
Rosmansyah & Rosyid (2017)	This study applied gamification in mobile learning to increase the fun factor in memorising the Quran. The results showed good performance in terms of usability, fun, and ease of use.	Badges, Points, Levels, Leaderboard, Progress bar	Achievement Motivation
Moulana (2017)	The goal of this research was to create a more engaging and motivational Quran memorisation experience by combining game design principles with existing memorisation strategies.	Points, Levels, Challenges, Achievement, Narratives	Engagement Motivation
Senan et al. (2017)	This research proposed an interactive Quran mobile application called iHafaz. This consists of three main modules: Hafaz (memorising 10 Surahs), Latihan (exercises with three difficulty levels), and Audit (monitoring progress) to facilitate autistic children in reciting and memorising the Quran.	Levels, Points, Awards	Engagement Motivation
Shamsuddin et al. (2016)	This research proposed a conceptual framework, e-Hafiz, for modelling mobile Quranic memorisation by applying learning theories such as game-based learning, chunking, and rote learning. Three essential modules were designed: recitation (Talaqqi), memorisation and evaluation, to help young learners memorise the Quran.	Avatar, Levels, Points, Timing, Awards	Engagement Motivation

Summative assessment provides students' overall achievement, such as grading or certification. The formative assessment, however, includes learning activities where students interact during lessons, such as answering questions and receiving feedback on the level of efficacy of their interaction. In developing and applying technical enjoyment in a gamified learning system, as illustrated in Table 3, gamification is achievable in Quranic memorisation, where all features in the game were adapted from the previous works.

Table 3
Suggestion on applying game elements in Quranic memorisation

No.	Game element	Suggestion of innovation
1.	Points and rewards system (Al Rumaisa et al., 2025; Pranata et al., 2020; Rosmansyah & Rosyid, 2017; Senan et al., 2017; Shamsuddin et al., 2016)	• Points for progress: Students earn points for each verse or page they memorise. This creates a sense of achievement and progress. • Badges and certificates: Award badges for milestones (e.g., completion of a Juz) and certificates for even bigger accomplishments (e.g., completion of half of the Quran), displaying concrete evidence of effort.
2.	Levels and challenges (Al Rumaisa et al., 2025; Moulana, 2017; Pranata et al., 2020; Sevutra & Erlinda, 2019; Shamsuddin et al., 2016)	• Progressive levels: The more the students memorise, the more they can move to higher levels of difficulty. This maintains the challenge at their skill level. • Daily or weekly challenges: Having short-term targets, such as memorising several verses in a week, may make the students active all the time.
3.	Leaderboards (Al Rumaisa et al., 2025; Rosmansyah & Rosyid, 2017; Sevutra & Erlinda, 2019)	• Healthy competition: Healthy competition among students can be encouraged by displaying leaderboards. Special care needs to be taken so that this does not lead to negative feelings, but compels everyone to perform better. • Group leaderboards: Creating group or class leaderboards can promote teamwork and collective motivation.
4.	Interactive learning tools (Khaleel et al., 2020)	• Mobile Apps: Apps that have gamification elements can be used to make Quran memorisation interesting and fun. Tools such as audio playback, practice reminders, and tracking progress can be an immense help • Virtual rewards: Virtual rewards, such as gifts or unlocking new features, might help in sustaining students' motivation.
6.	Feedback and reflection (Siregar et al., 2023)	• Prompt feedback: Giving prompt responses on the precision retention or memorisation of the Quran assists learners in amending their errors fast and gaining a sense of achievement. • Reflection time: Providing time to reflect on the lessons and meanings of the verses consolidates the religious bond and increases motivation.
7.	Personalised learning paths (Adhoni et al., 2014)	• Specification of learning goals: Allow students to set their goals on the Surah and the number of verses they will memorise, and enable them to modify these goals at any time, to permit students to feel proficient in their learning. This will also be useful because it is important to track the learning goals based on the targeted verses to be memorised. • It then tests the memorised verses of the student by providing questions to be answered by the students based on the level of difficulty. This is centred on the pace of the student and thus can ease the process by making it less cumbersome.
8.	Incorporating stories and context (Moulana, 2017; Pranata et al., 2020)	• Historical context: Sharing stories and historical background related to the verses being memorised can make the content more engaging and meaningful. • Moral lessons: Highlighting the moral and ethical lessons within the verses can inspire students to connect more deeply with the material.

METHODOLOGY

This study used a design and development research (DDR) method to incorporate gamification principles in memorising the Quran. The DDR method that provides orderly procedures for developmental research is divided into three phases: 1) analysis, 2) design and development, and 3) evaluation. Table 4 summarises the mapping of DDR stages according to the research objectives.

The following is the step-by-step description involved in this research methodology.

Table 4
The mapping of DDR stages according to research objectives

DDR Stage	Research Objective (RO)
1. Analysis	RO1: To integrate the multimedia elements into memorisation strategies that can reduce the cognitive load in memorising the Quran. RO2: To determine the appropriate gamification elements that can be integrated to facilitate sustainable motivational learning theory in memorising the Quran.
2. Design and Development	RO3: To design and develop a gamified learning application that can improve students' motivation and engagement in memorising the Quran.
3. Evaluation	RO4: To assess the effects of gamification elements on the students' motivation and engagement in Quranic memorisation.

Step 1: Analysis (Awareness of the Problems)

Analysis is the first step, where the most important issues and challenges faced by students in memorising the Quran were identified. Although Quranic memorisation is a well-established tradition in Malaysia, contemporary memorisation methods and initiatives among Malaysian teenagers in tertiary education encounter numerous problems and challenges, such as:

1. Insufficient pedagogical diversity, reflecting the absence of multiple learning styles (Mustafa et al., 2019), refers to an overreliance on a narrow set of teaching and learning methods, often focusing on traditional, teacher-centred, and rote-based techniques, without incorporating varied, learner-centred, and adaptive strategies that address different learning needs and preferences. Learners have different cognitive profiles: some are auditory, some visual, and others kinaesthetic. A strictly verbal, memorisation-heavy method may not be effective for everyone, especially for students with learning difficulties or attention issues. Repetitive routines without variation can lead to boredom, disengagement, and fatigue. Teenagers and university students, especially digital natives, are more responsive to interactive, visual, or technology-based learning methods.
2. Lack of interest and consistent motivation (Purbohadi et al., 2019; Rosmansyah & Rosyid, 2017; Shukri et al., 2020): Lack of sustained motivation is among the most

significant challenges in Quranic memorisation, particularly among adolescents and university students who are prone to academic and emotional pressures. Motivation is the force that stimulates, sustains, and directs learning activities. Quran memorisation is not a single task but a continuous process requiring daily, consistent effort, self-discipline, and a spiritual connection. Though most students begin with great zeal, research and field studies show that motivation gradually dwindles. The memorisation process is tiresome and monotonous, and without frequent internal or external reinforcement, students risk getting bored, skipping revision classes, experiencing memorisation exhaustion, or eventually abandoning their Hifz goals altogether. This is particularly noticeable in teens and university students, who are balancing hectic academic schedules, studies, personal responsibilities, pressure, and emotional development. One reason motivation wavers is the lack of specific goals and performance measures: without observable progress, motivation fades. In other words, without specific progress monitoring, active learning activities, or strong identification with the verses, numerous learners experience memorisation weariness and withdrawal.

3. Difficulty in memory retention (Khafidah et al., 2020; Radina et al., 2020; Shukri et al., 2020; Yusuf et al., 2019): Retention in Quranic memorisation is the ability to recall, remember, and recite memorised verses with accuracy over time. Effective *huffaz* are not so many people who memorise the Quran, but rather, and most importantly, people who maintain the accuracy and fluency of memorisation through regular revision. Most students have poor retention rates due to various factors, especially when they move to new verses too quickly without reviewing the earlier verses. Low retention remains one of the most pressing challenges in Quranic memorisation, as learners often forget previously memorised material for reasons such as:

    a. Ineffective revision practices
       Many learners focus on memorising new verses but neglect consistent Muraja'ah (a proper, systematic revision process). In addition, without planned revision cycles, older memorisation fades rapidly.
    b. Lack of feedback and self-monitoring
       Learners often do not realise they are forgetting until they are asked to recite after a long interval. Thus, without self-monitoring tools, forgetting goes unnoticed and uncorrected.
    c. Surface-level memorisation or rote learning without deep processing.
       Memorisation through mechanical repetition (rote) without understanding meaning, structure, or connections results in surface-level encoding, which fades quickly.

From the identified problems above, a solution was designed in the next stage.

Step 2: Designing and Developing a Gamified System

The second phase involves planning a prototype solution (artefact design) that uses gamification elements to help learners memorise the Quran and achieve the desired learning outcome. The gamification elements were adapted from the previous research that has been discussed in the Literature Review section. In this study, the MDA gamification model proposed by Hunicke et al. (2004) was used to plan and design the necessary game elements to be implemented. MDA is a formal approach for understanding and analysing game design into three distinct components: Mechanics, Dynamics and Aesthetics. This framework is employed as a model of the main structural components of gamification, as it has been seen as a common framework in game system design and gamification (Sezgin & Yüzer, 2020). Mechanics refers to the rules and systems that define how the game operates. Dynamics describes how players interact with the game and with each other. Aesthetics focuses on the players' emotional and sensory experiences created by the game. At this stage, several criteria for game activity must be carefully selected, and the design must align with the learning content, for example, by asking the player or memoriser to set their goals (such as the target verse to memorise within a certain period or milestone). In addition, incorporate tokens of appreciation (such as rewards and points) into the game design so that students earn points for the activities they complete. Concurrently, the meta-reward for continuous effort is represented by badges, levels, social points, and bonus points. Table 5 illustrates several aspects of gaming incorporated into the Quranic memorisation process.

When designing a game, it is important to incorporate both external and internal motivation to achieve its goals and align with players' personalities. Therefore, this research adopts the three key intrinsic motivation elements from Malone's model challenge, curiosity, and control, as well as findings from studies by Shute and Ke (2012) , Noor et al. (2018) and Noor et al. (2019) as shown in Table 6. By incorporating features such as challenges, points, levels, badges, and rankings into a Quranic memorisation gamification application, a new model can be developed to make memorisation more enjoyable and engaging.

To design the Gamification Learning for Al-Quran Memorisation (GLAM-Q) application, each game element was applied particularly during the system registration, recitation of verses, quiz evaluation and learning analysis, as stated in Table 7.

Step 3: Evaluation of Gamified System

The process involves assessing the extent to which gamification components affect the motivation and engagement of students towards memorising the Quran. Performance or measurement after each game activity should be evaluated and visualised on an interactive dashboard to track Quran memorisation progress. This formative feedback should be supportive, timely, and specific. The integration of learning analytics parameters in a

Table 5
Integration of game elements into Quranic memorisation

Game elements	Integration into Quranic memorisation
Progression systems	Levels and Stages: Divide the memorising process into levels. Completing a level unlocks the next, providing a sense of achievement. Milestones: Celebrate milestones, such as memorising a certain number of verses, completing a Surah, or mastering the ability to recall many or all verses in a particular Surah.
Points and rewards	**Earning Points**: Reward points for good recitation of the Quran verses, participating in memorisation classes, or performing tasks independently. **Virtual or Physical Rewards**: Award prizes such as certificates, medals, or small gifts for the achievement of some goals.
Badges and achievements	**Skill Badges**: Merit recognition, such as skill badges, for mastering memorising skills or answering questions quickly. **Achievement Tiers**: Create tiers such as "Beginner", "Intermediate", and "Advanced" to motivate learners to progress.
Leaderboards	**Healthy Competition**: Use leaderboards to track progress among peers and encourage friendly competition.
Challenges or Quests	**Daily/Weekly Challenges: Make tiny, attainable tasks, such as memorising a couple of verses or reciting a Surah.**
Feedback and Progress Tracking	**Visual Progress Bars**: Visualise progress, such as a progress-tracking bar for each verse memorised. **Instant Feedback**: Give prompt responses on the recitation and memorisation progress.
Storytelling and Theme	**Narrative Integration**: Apply Quran narratives to place memorisation in context, more interesting with pleasant and vivid visuals. **Thematic Memorisation:** Practise memorising verses or Surahs of a similar theme, i.e., patience, faith, or gratitude.
Time-based Elements	**Timed Challenges**: Set deadlines on memorising certain parts to create urgency and concentration. **Daily Streaks**: Promote regular practice by rewarding daily progress-tracking streaks of memorisation.
Personalisation	**Custom Goals**: Enable learners to establish personal memorisation goals and monitor progress. **Adaptive Learning**: Tune the challenge according to the learner's speed and skill.

gamified Quran memorisation application aims to measure motivation and engagement. The measurement of learning analytics for each evaluation criterion is summarised in Table 8.

The analysis of students' perceptions towards elements of gamification which affect motivation as well as engagement of memorising the Quran through gamification and learning analytics has already been explained in the Results and Discussion section.

Proposed GLAM-Q Prototype Application

The screens of the Gamification Learning for Al-Quran Memorisation (GLAM-Q) application were designed based on four main menus: register, memorisation, interactive

Table 6
The adaptation of Malone's motivational theory, Noor et al. (2018) and Noor et al. (2019) *in the proposed approach*

Elements	Characteristics	Adoption in the proposed GLAM-Q prototype
Challenge	i. Set specific and clear goals and rules	• Games should present clear overall and intermediate goals; the goal here is to master Quran memorisation according to the location (numbers/ pages) of the verses. • Games have rules that guide players on what to do and when. For example: a. Memorisers must repeat the memorisation at least 10 times before continuing with the next verses (Nafi et al., 2019). b. The formula for calculating the score is determined not only by the number of correct answers, but also by the deductions if the player asks for a clue.
	iii. Provides adaptive challenges	• Good games balance the difficulty of challenges at appropriate levels (easy, medium, complicated) to match players' abilities. • The challenges should match the player's skill level, increase as the player progresses through the game, and allow for player-centred pacing.
	iv. Uncertain outcomes	• The variety of the Quran verses makes testing difficult. • Some hidden information is provided as clues but is selectively revealed by players when needed. • Players will be asked to guess the location of the verses at random.
	v. Performance feedback	• Provides clear, frequent, constructive and positive feedback to promote competence and self-esteem relevant to memorisers. • Rapid feedback also stimulates motivation and engagement in the Quranic memorisation process.
Curiosity	i. Sensory curiosity	• Uses a combination of graphics and visual elements to stimulate the senses of players.
	ii. Cognitive sensory - incompleteness of competence	• Players cannot proceed to other levels and modules if they do not achieve the specified target.
Control	i. Contingency - responsive environment	• Provides timely information about their performance, such as score status, number of repetitions, and number of clue buttons clicked.
	ii. Choice	• Players can control the process by themselves to collect points, badges and participate in other activities. • Games should have an easy, user-friendly interface that provides tutorials or help to enable players' skill development as they progress through the game and should reward players for developing their skills.

Table 7
The application of game elements in the memorisation stages

Memorisation Stage	Game Elements Applied	Parameter of Learning Analytics
Registration	Avatar	• Date and time log into the system • The frequency of access to the system
	Progress bar	• The completion of a particular Surah • The number of verses memorised
Memorisation (recitation of the verses)	Points	• The number of repetitions for each verse contributes some points. If memorised more than ten times, addition points will be rewarded.
	Badges (or trophy)	• The completion of Surah
	Narratives (storylines of the Surah help to describe the meaning of the verses)	
Quiz Assessment	Level Challenges	• Type of Quiz including fill- in-the-blanks (easy level), sorting the verses in the correct sequence (medium level), and guessing the location number of a verse (difficult level)
	Points Scores Badges Timer	• The number of correct answers • The duration for recalling the verses • The number of corrections • The number of hints or clues clicked
	Avatar (meta-reward to change the preferred avatar)	
Learning Analysis	Leaderboard	• Ranking (based on the highest aggregate scores and the shortest duration of quiz assessment).

quiz, and learning analysis, which includes formative assessment and summative assessment. For usability testing, the prototype was developed only for Surah Al-Baqarah (verses 1 to 5), Surah Az-Zalzalah (verses 1 to 8), and Surah As-Syams (verses 1 to 15). The application was designed according to several stages, as follows:

1. Registration: Students must log in to their personal account, and they can choose an avatar for their personal profile to enhance the gamification environment. Details for the registration screen are shown in Figure 1.
2. Memorisation: The memorisation of verses must be done page by page (Figure 2), with the number of verses, the number of repetitions and the duration for recalling the verses recorded accordingly, as shown in Figure 3. This approach is considered goal-oriented chunking (deliberately under strategic control) and perceptual chunking (a more automatic and continuous process of chunking during perception). Gobet et al. (2001), in their work, suggested integrating the

Table 8
The measurement of learning analytics for motivation and engagement criteria

No.	Evaluation criteria	Gamification elements	Learning analytics (parameters)
1.	Motivation	1. Reward: Incorporate tokens of appreciation as rewards or points when designing the game. Students receive points based on the activities they perform or complete. Meanwhile, the meta-reward for continuous effort is represented by badges, levels, social points, and bonus points. 2. Score: The formula for calculating the score is determined not only by the number of correct answers, but also by deductions if the player asks for a clue or if a correction is made. 3. Achievement: Over a certain status, such as badges, medals, and/or trophies that praise the players' specific actions. 4. Progression: Allows players to determine their position in a game based on their progress. Examples and synonyms include progress bars, maps, and steps. 5. Levels: Players cannot proceed to other levels or modules if they do not achieve the determined target. 6. Stats: Visible information such as results, indicators, or data from the game presented to the user and related to their outcomes within the game. For example, it provides timely information about their performance, such as score status, number of repetitions, and number of clue buttons clicked. 7. Time pressure: Countdown, clock, or/ and/or timers	• Memorisation time • Response period • Scores / points collected
2.	Engagement	1. Number of repetitions and period in memorising verses. 2. The number of clue buttons is clicked. 3. Chance: to have a second chance to repeat the assessment, allowing players to redo or restart the action. 4. Players can control the process by themselves to collect points, badges, and other rewards for skill development. 5. Feedback from the formative and summative assessments.	• The number of times the hint is pressed. • The number of corrections is made. • Response period.

goal-oriented and perceptual forms of chunking for more comprehensive learning, which can also be used to predict typical errors in capturing details of the learning process. Memorisation will begin with the first selected verse. The number of repetitions and the duration will be recorded. A saturated red text is used to highlight the verse from left to right, following the reading flow. A clue picture will be displayed as the cursor moves over the verse. After the verse has been memorised several times, the next verse will be displayed only after the previous verse has been fully memorised. A notification message will inform the user whether the number of memorisations is sufficient. If it is fewer than 10 times, the user will be prompted to repeat the process, and the next verse (Figure 4) will not be displayed. Otherwise, the next verse will be shown.

3. Assessment (with interactive quizzes): Gamification elements were implemented in both formative and summative assessments. Formative assessment is implemented as soon as the student has finished memorising a page of Surah, while summative assessment is carried out after all the pages have been memorised. Quiz assessments wwill be provided at three levels of difficulty (easy, intermediate, and difficult) and distributed in each session, as shown in Figures 5 to 7 below. The first level (easy) is a fill-in-the-blank task in which students drag the answer options provided below. The second level (intermediate) requires students to arrange the verses in the correct sequence. Finally, the third level (difficult) asks students to identify the location number of a verse. The quizzes were designed with game elements such as levels/challenges, points/scores/badges, and a timer (Table 9). During these assessments, relevant data will be captured (Table 10), including the total points or scores collected, time spent answering the questions, the number of corrections, the number of verses successfully memorised, and the number of mistakes. Examples are shown in Figure 8.
4. Learning analysis: The progression of an individual's learning performance is analysed as a summative assessment. The data captured (Table 11) will be summarised and displayed on a dashboard. In Figure 9, the overall score obtained will determine each student's ranking and will be displayed on a scoreboard.

RESULT AND DISCUSSION (A GLAM-Q APPLICATION)

Merging gamification into usability testing paves the way for an integrated, interactive, and effective means of study. This section outlines students' perceptions of gamification elements that influence motivation and engagement in Quranic memorisation through gamification and learning analytics. All items in the questionnaire were adapted from established instruments, namely the Student Engagement Scale (SES) and the Intrinsic Motivation Inventory (IMI). Before distributing the questionnaire to respondents, ten experts with a minimum of five years of experience in teaching and/or research in



Figure 1. Registration screen interface with avatar selection

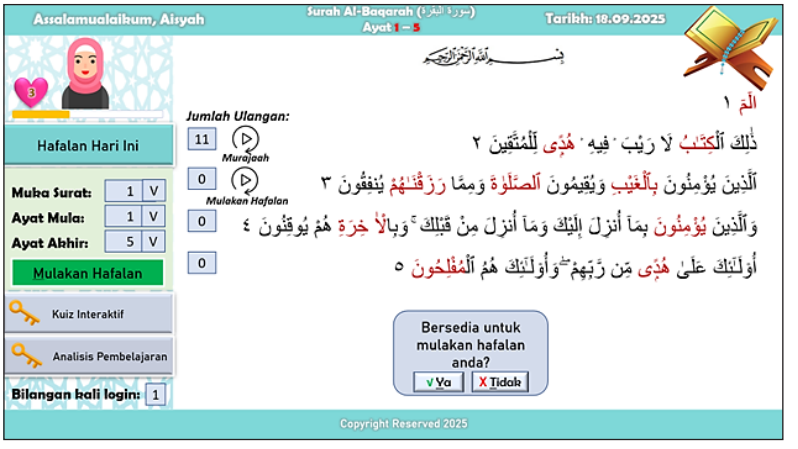


Figure 2. Memorisation screen interface

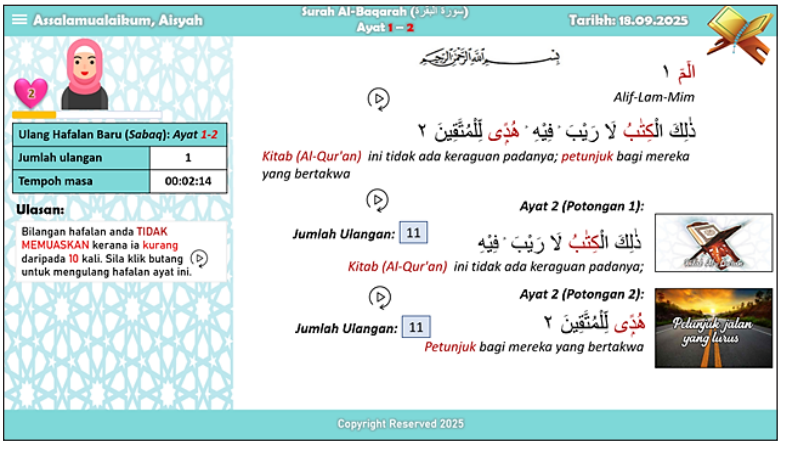


Figure 3. The verse is highlighted by segments (with certain chunks)

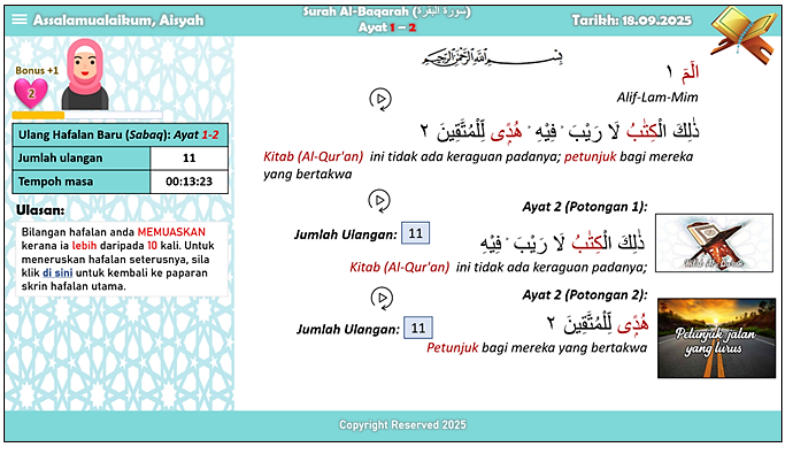


Figure 4. Example of a screen that informs the user whether the number of memorisations is enough or not

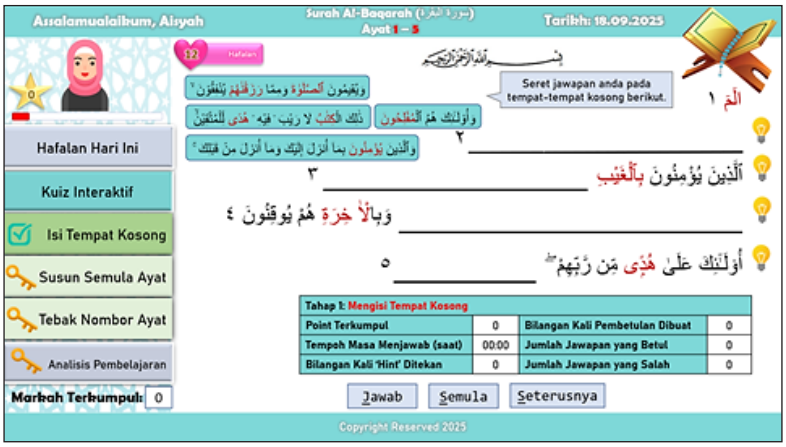


Figure 5. Screen of quiz level 1



Figure 6. Screen of quiz level 2



Figure 7. Screen of quiz level 3



Figure 8 (i-ii). Example of screens that provide feedback in student assessment



Figure 9. Dashboard screens that summarise the formative and summative memorisation performances

Computer Science, Information Technology, Multimedia, Human-Computer Interaction (including gamification or game design), Islamic Studies, Quranic memorisation, or a related discipline have reviewed the content validity of the questionnaire. The modified questionnaire consisted of 31 items, with engagement and motivation as the two major constructs. Engagement was evaluated based on three aspects: behaviour, emotion and

Table 9
Integrated game design elements in the challenge-based gamified quizzing

No.	Game elements	Description	Purpose
1.	Points/scores/ badges	Numeric measure of memoriser's performance (answer corrected).	An indicator of explicit reward.
2.	Levels/challenges	The difficulty level of quizzing provides a sense of progression.	Indicator of progression and difficulty.
3.	Timer	Numeric measure of memoriser's performance (time used).	An indicator of time pressure.

Note. Adapted from Anunpattana et al. (2021)

Table 10
Learning analysis during formative assessment

No.	Analysis data (while answering the quiz)	Analysis data (after answering the quiz)
1.	Verse number.	1. Verse number.
2.	The accumulated points.	2. The number of memorisation repetitions.
3.	The duration of answering each quiz question.	3. The duration of learning (during memorisation and answering quizzes).
4.	The number of times the 'Hint' button is pressed.	4. The number of sentences successfully memorised.
5.	The number of times corrections are made.	5. The total score accumulated.
6.	The correct number of sentences answered.	6. How many times has the system been accessed
7.	The number of wrongly answered sentences.	

Note. Formative assessment is conducted when learners answer the quizzes for each part of memorisation

Table 11
Learning analysis during summative assessment

No.	Analysis data (for summative assessment)
1.	Verse number
2.	The number of memorisation repetitions (for each sentence)
3.	The duration of memorisation repetition (for each sentence)
4.	Total learning time (during memorisation and answering quizzes)
5.	Total accumulated score (for each formative assessment)
6.	Total cumulative score (overall) – in per cent
7.	The number of times the system was accessed
8.	Ranking on the scoreboard, which is based on the highest total (overall) score and the shortest (overall) period

Note. The information collected during the summative assessment is based on a student's overall progression

cognition perspectives. Motivation was assessed using self-determination theory (SDT), namely, autonomy, competence, and relatedness, and the ARCS model, which includes attention, relevance, confidence and satisfaction.

The GLAM-Q application was tested through a descriptive analysis using a random sampling technique on purposively selected participants. A summary of responses from 46 respondents among students in higher education —male: 20 (43.5%), female: 26 (56.5%), aged 19 to 27— is provided in Table 12. The analysis of questionnaire items measuring respondents' views on Quranic memorisation in a game-based system yielded extremely high overall mean scores on a 5-point Likert scale, ranging from 4.55 to 4.91. The highest mean score of 4.91 was obtained by statements assessing *feeling a deep sense of fulfilment when memorising a new Surah*, and *increased motivation after successfully answering a quiz assessment*. The lowest mean score of 4.55 was for statements assessing *seeking additional resources*, though this comes under the category "agree to strongly agree".

Table 12
The analysis of results

No.	Question item	Mean	SD	Cronbach's Alpha
1.	I memorise the Quran because I truly enjoy learning and reciting its verses.	4.76	0.44	0.966
2.	I enjoy challenging myself to improve my recitation and memorisation skills.	4.82	0.39	0.967
3.	I seek additional resources (e.g., verse translations) to deepen my understanding of the verses I memorise.	4.55	0.62	0.969
4.	I feel a deep sense of fulfilment when I successfully memorise a new surah.	4.91	0.29	0.966
5.	I put a lot of effort into my Quranic memorisation because I believe it is important in obtaining the blessings of Allah SWT.	4.82	0.39	0.966
6.	I follow the rules for memorising the Quran using this gamified system: I repeat the recitation of verses several times before proceeding to the next.	4.82	0.39	0.97
7.	I follow the rules in this gamified system by memorising all the verses of the Surah before taking the quiz.	4.79	0.48	0.967
8.	I actively participate in Quranic recitation activities through this gamified system.	4.82	0.39	0.966
9.	I actively participate in quiz challenges through this gamified system.	4.76	0.50	0.966
10.	I consistently repeat previously memorised Quranic verses using this gamified system.	4.76	0.44	0.966
11.	I use this gamified system to improve my progress in Quranic memorisation.	4.70	0.47	0.966
12.	I persist in memorising Quran verses because I receive immediate feedback from the learning analysis in this gamified system.	4.58	0.56	0.965

Table 12 (*continued*)

No.	Question item	Mean	SD	Cronbach's Alpha
13.	I try to complete my memorisation goals using this gamified system, even when it is difficult.	4.58	0.61	0.967
14.	I use gamification features (e.g., levels, points, trophies, and leaderboards) to track my progress in Quranic memorisation.	4.67	0.54	0.965
15.	I am more likely to persist in memorising difficult or longer verses when I receive rewards (incentives) from memorisation activities through this gamified system.	4.67	0.60	0.967
16.	I am committed to my memorisation goal regardless of external rewards.	4.64	0.49	0.967
17.	I enjoy memorising the Quran when using this gamified system.	4.82	0.39	0.97
18.	I feel more excited to continue my Quranic memorisation when I receive rewards (e.g., unlocking new levels, earning points, and receiving virtual trophies) through this gamified system.	4.67	0.54	0.965
19.	I feel fun to memorise the Quran when competing with others on the leaderboard (through the challenge).	4.58	0.66	0.967
20.	I feel supported when I receive feedback or encouragement from this gamified system.	4.70	0.47	0.966
21.	I feel confident in my ability to memorise the Quran effectively using this gamified system.	4.79	0.42	0.966
22.	I believe this gamified system helps me improve my memorisation skills over time.	4.79	0.42	0.965
23.	I feel a sense of accomplishment when I unlock new milestones (e.g., completing a new Surah and/or reaching a new level in the assessment).	4.82	0.39	0.966
24.	I feel my motivation increased when I managed to answer the memorisation quiz assessment well.	4.91	0.29	0.966
25.	I think gamification features in this system (e.g., progress tracking bar) help me stay focused on my memorisation goals.	4.76	0.44	0.97
26.	I think the use of gamification elements (e.g., levels, points, and trophies) encouraged me to set specific goals for the Quran memorisation challenge.	4.79	0.42	0.966
27.	I am more likely to use memorisation techniques (e.g., chunking into short verses, repetition during recitation, visualisation, and association via text and pictures) when they are incorporated into a structured, gamified learning environment.	4.67	0.54	0.967
28.	I plan my memorisation strategy based on my progress and performance in this gamified challenge system.	4.70	0.53	0.966
29.	I actively reflect on the meaning of the verses I memorise (with pictures), not just their pronunciation, when using this gamified system.	4.76	0.44	0.966
30.	I think that multimedia elements in this gamified system, such as text, visuals (images), and/or audio, help enhance the memorisation of Quranic verses.	4.85	0.36	0.966
31.	I find it easier to memorise the Quran when interactive learning tools (e.g., audio and visual cues) are included in this gamified system.	4.85	0.36	0.965

Standard deviations (SDs) from 0.29 to 0.66 suggest a low response variation, i.e., most respondents kept high ratings for statements on the same scale side. This proves that the gamified system functions well across the group of respondents. Reliability analysis also showed very high internal consistency of all the statements, having a Cronbach's alpha of 0.965 to 0.970. The standard criterion (Taber, 2018) clarifies that a value above 0.9 indicates exceptional reliability. This provides confidence that the questionnaire items are reliable and consistently measure engagement, motivation, and support for memorisation.

Table 13 shows that the gamification learning application has a positive impact on learner engagement in three dimensions, namely, behaviour (96.83%), emotion (99.13%) and cognitive (97.55%). Behavioural analysis clearly indicates an increased level of active participation in improving Quranic memorisation, not only in the recitation task but also in the quiz assessment, as reflected in higher interaction rates, task completion rates and the amount of memorisation. In terms of emotional engagement, learners exhibited favourable affective responses toward the gamified application, such as truly enjoying and feeling excited reciting the verses when receiving rewards, feeling a deep sense of fulfilment after successfully memorising a new Surah, feeling supported when they receive feedback or encouragement from this gamified application, feeling a sense of accomplishment when they unlock new milestones and compete with others on the leaderboard, among others. Cognitive engagement also improves, as reflected in learners' sustained attention, effort, and interaction with complex tasks. The structure of the gamified activities appears to facilitate deeper processing, problem-solving, and strategic thinking rather than superficial interaction. This suggests that gamification is supportive of meaningful learning by

Table 13
The variable measurement scale results

Construct	Variable	Question Item	Result		
			Strongly Agree	**Agree**	**Total**
Engagement	Behaviour (BE)	2, 5 – 16	68.90%	27.93%	96.83%
	Emotion (EE)	1, 4, 17 – 24	73.26%	25.87%	99.13%
	Cognitive (CE)	3, 25 – 31	67.12%	30.43%	97.55%
Motivation	Autonomy (AU)	2, 5 – 13, 15 – 16, 28	68.73%	28.26%	96.99%
	Competence (CM)	13, 21 – 24	72.17%	25.65%	97.82%
	Relatedness (RE)	19 – 20, 28	66.67%	30.43%	97.10%
	Attention (AT)	25 – 26, 29 – 31	70.87%	28.70%	99.57%
	Relevance (RV)	3, 14, 27, 29 – 31	66.67%	29.71%	96.38%
	Confidence (CF)	21 – 24	76.09%	23.91%	100.00%
	Satisfaction (SA)	1, 4, 15 – 18, 23 – 24	72.83%	25.82%	98.65%

encouraging active knowledge construction among learners on the condition that it is aligned with an instructional objective.

In addition, gamification, when used in conjunction with Quranic memorisation, can significantly enhance learner motivation, particularly in terms of autonomy (96.99%), competence (97.82%), relatedness (97.10%), attention (99.57%), relevance (96.38%), confidence (100%), and satisfaction (98.65%). In terms of autonomy, learners demonstrated more self-regulation in behaviour; repeated interactions with the application were also depicted. Regarding competence, performance achievement mechanisms such as points, levels, and progress appear to improve learners' perceptions about the development and mastery of skills. This aspect of competence is important to maintaining motivation, since learners are more likely to be motivated when they see that their efforts are effective. The dimension of relatedness was reinforced through application-embedded social or comparative elements, such as leaderboards. This feature fostered a sense of connection and belonging, motivating learners to participate consistently and align their efforts with those of others. The ARCS model states that the application successfully captured learners' attention through interactive tasks, game mechanics, and dynamic feedback. In terms of relevance, learners' continuous participation in the application may indicate that the content and challenges were perceived as meaningful and aligned with their learning needs and objectives, thereby strengthening motivational commitment. Learners' confidence is built through incremental challenges and immediate feedback, which helps students monitor progress to estimate upcoming success. Finally, satisfaction is reflected in the learners' steady involvement and willingness to continue using the application, suggesting that the rewards, achievements, and overall learning experience have met or exceeded learner expectations.

Consequently, it can be inferred that GLAM-Q triumphantly created an entertaining learning environment that motivates students to use the application continuously. The findings emphasise that gamification applications have immense potential in boosting learning quality in religious content exploration using digital platforms.

CONCLUSION

Most studies treat gamification as a stable phenomenon, whereas in practice, the gamification world can be designed in almost any form. Despite researchers' efforts to assess the impact of gamification in educational contexts, several methodological shortcomings remain. In fact, the number of research studies with high methodological rigour is decreasing, and consequently so is the reliability of the outcomes. In this study, we identified the fundamental elements that define the methodological approach to implementing gamification in Quranic memorisation. As gamification can incorporate game design components in various ways and take many distinct forms, it is inappropriate to study the motivational effects of gamification as a generic construct. This is because gamified

learning will have a substantial impact on students' subject knowledge and attitudes towards continuous learning if they are highly motivated to participate.

This article reviews previous research on current methods of teaching and learning that incorporate gamification into Quranic memorisation. This study has also identified gamification features to consider when determining whether gamified activities and applications are appropriate for use in existing Quranic memorisation environments. To keep students motivated to memorise Quranic verses consistently, several gamification elements were incorporated to enhance engagement throughout lessons and assessments. We have access to data that directly reflects engagement, such as total learning time and the number of system accesses. The general consensus is that a user who connects more frequently and for a longer period is considered more engaged in the activity than one who connects less frequently or for a shorter period. However, this does not allow us to compare engagement levels at two different times; it only allows us to compare the overall level of engagement. Therefore, we require a different method to monitor engagement in real time. To gauge engagement, we must monitor how the user interacts with both the gamification layer and the learning environment by capturing the duration of learning, specifically while memorising and answering the quizzes.

The main contribution of this study is the suggestion that future gamification designs should incorporate components that may sustain students' motivation for longer by addressing issues that could cause them to lose enthusiasm for memorising the Quran. In addition, it is important to carefully apply gamification features such as points, progression, levels, acknowledgement, sensation, and others, so that students remain focused on learning rather than merely accumulating rewards. Future research on gamification should also examine whether applying gameplay activities would be beneficial, as it may otherwise leave students bored and waste a significant amount of their time. Future gamification programmes should also be made with utmost consideration for demographic characteristics so that the gamification aspects can properly enhance students' motivation and interest in memorising the Quran. Further research on the implementation of gamification and psychological concepts can be conducted, integrating future developments and innovations, based on the positive impact observed after the implementation of such innovations. Additionally, it is believed that this study has helped the students memorise the Quran better and aided teachers in evaluating their students more conveniently.

ACKNOWLEDGEMENT

The authors want to express their gratitude to the experts who have participated in the study with their valuable comments. This work was conducted without any funding from public, commercial, or not-for-profit agencies.

REFERENCES

Rahman, M. H. A., Panessai, I. Y., Noor, N. A. Z. M., & Salleh, N. S. M. (2018). Gamification elements and their impacts on teaching and learning - A review. *The International Journal of Multimedia & Its Applications*, *10*(06), 37-46. https://doi.org/10.5121/ijma.2018.10604

Abdullah, N. M. S. A. N., Sabbri, F. S. M., & Isa, R. A. M. (2021). Tahfiz students' experiences in memorising the Qur'an: Unveiling their motivating factors and challenges. *IIUM Journal of Educational Studies, 9*(2), 42-63. https://doi.org/10.31436/ijes.v9i2.248

Adhoni, Z. A., Hamad, H. Al, Siddiqi, A. A., & Adhoni, Z. A. (2014). An instant-based Qur'an memoriser application interface. In *Proceedings of the 2013 International Conference on Information Science and Cloud Computing Companion (ISCC-C 2013)* (pp. 382-387). https://doi.org/10.1109/ISCC-C.2013.14

Ahmad, A. M., Rahman, A. A., Saleh, M. H., & Daud, Z. (2024). Diversifying Quranic revision methods using gamification-based teaching material for Tahfiz education. *International Journal of Evaluation and Research in Education, 13*(2), 987-996. https://doi.org/10.11591/ijere.v13i2.26218

Akbari, M. Y. A., Kusaeri, Bakar, M. Y. A., & Rusydiyah, E. F. (2024). Typology and effectiveness of Al-Quran memorisation applications. *Tafkir: Interdisciplinary Journal of Islamic Education, 5*(4), 598-611. https://doi.org/10.31538/tijie.v5i4.1158

Al Rumaisa, A., Mudrikah, F., & Rakin, M. (2025). The effectiveness of gamification-based Islamic religious education learning models on digital Al-Qur'an literacy. *Journal on Islamic Studies, 2*(1), 1-11. https://doi.org/10.35335/b6vgar24

Almosallam, E. A. (2015). Towards improving Quran memorisation using mind maps. In *Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences (NOORIC 2013)* (pp. 128-132). https://doi.org/10.1109/NOORIC.2013.36

Almosallam, E., Alawadh, M. M., Alhasani, R. S., Almansour, S. M., Altamimi, W. A., & Altujjar, Y. R. (2016). ITQAN: A mobile-based assistant for mastering Quran memorisation. In *Proceedings of the 2015 5th International Conference on e-Learning (ECONF 2015)* (pp. 349-352).https://doi.org/10.1109/ECONF.2015.42

Alomair, Y., & Hammami, S. (2020). A review of methods for adaptive gamified learning environments. In *Proceedings of the 3rd International Conference on Computer Applications and Information Security (ICCAIS 2020)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCAIS48893.2020.9096871

Ambo, N. F., & Mokhsein, S. E. (2019). Trend and issue in learning strategy of Tahfiz Model Ulul Albab (TMUA). *International Journal of Academic Research in Business and Social Sciences, 9*(7), 1418-1426. https://doi.org/10.6007/IJARBSS/v9-i7/6789

Anunpattana, P., Nor, M., Khalid, A., Iida, H., & Inchamnan, W. (2021). Capturing potential impact of challenge-based gamification on gamified quizzing in the classroom. *Heliyon, 7*(19), Article e08637.https://doi.org/10.1016/j.heliyon.2021.e08637

Anwar, M. A., & Inayati, N. L. (2025). Gamification of Islamic education: Exploring the role of Wordwalls in increasing student participation in learning the Qur'an and Hadith. *Fikroh: Jurnal Pemikiran Dan Pendidikan Islam, 18*(2), 281-291. https://doi.org/10.37812/fikroh.v18i2.1898

Aziz, M. M., Abdullah, W. M., Ahmad, A. M., Mushim, M. A. A., & Shahrudin, M. S. (2019). Comparison between conventional method and modern technology in Al-Qur'an memorisation. *International Journal of Recent Technology and Engineering, 8*(1), 289-294. https://doi.org/10.35940/ijrte.B1373.0882S819

Fanani, L., Brata, K. C., Brata, A. H., & Fauzi, M. A. (2019). Usability evaluation of mobile-based application for Al-Quran writing learning with gamification. In *Proceedings of the 2019 International Conference on Sustainable Information Engineering and Technology (SIET 2019)* (pp. 276-279). IEEE. https://doi.org/10.1109/SIET48054.2019.8986123

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243. https://doi.org/10.1016/S1364-6613(00)01662-4

Haryono, K., Rajagede, R. A., & Negara, M. U. A. S. (2022). Quran memorisation technologies and methods: Literature review. *International Journal on Informatics for Development, 11*(1), 192-201. https://doi.org/10.14421/ijid.2022.3746

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, No. 1, p. 1722).https://doi.org/10.1016/j.cgh.2017.08.026

Ishak, I., Najwa, N., Rahim, A., Salim, N. I., Ulfah, C., Sholeha, N., Husaini, T., Jamaludin, I., Khalid, H. M., Adeena, N., Lutfi, A., Afzaruddin, S. S., Ghazali, A. R., Fitri, A., Ludin, M., Malia, N., Warif, A., Ibrahim, F. W., Othman, F. A., & Yahya, H. M. (2022). Development and validation of Huffaz ProHealth 1.0: A module to improve the well-being of Tahfiz students in Selangor, Malaysia. *International Journal of Environmental Research and Public Health, 19*, Article 7718.https://doi.org/10.3390/ijerph19137718

Karamert, Ö., & Kuyumcu Vardar, A. (2021). The effect of gamification on young mathematics learners' achievements and attitudes. *Journal of Educational Technology and Online Learning, 4*(2), 96-114. https://doi.org/10.31681/jetol.904704

Khafidah, W., Wildanizar, Tabrani, Nurhayati, & Raden, Z. (2020). The application of Wahdah method in memorising the Qur'an for students of SMPN 1 Unggul Sukamakmur. *International Journal of Islamic Educational Psychology, 1*(1), 37-49. https://doi.org/10.18196/ijiep.1104

Khaleel, F. L., Ashaari, N. S., & Wook, T. S. M. T. (2020). The impact of gamification on students' learning engagement. *International Journal of Electrical and Computer Engineering, 10*(5), 4965-4972. https://doi.org/10.11591/ijece.v10i5.pp4965-4972

Misnan, N., Saleh, W. A., Manaf, S. M. A., & Mohamad, M. T. (2021). I-Tasmi' Tahfiz: Innovation for memorising Al-Quran. *E-Journal of Islamic Thought and Understanding, 4*(2), 14-25. https://doi.org/10.24191/ejitu.v4i2.6078

Sabbri, F. S. M. (2016). *Exploring student motivation and challenges in Tahfiz class in selected private Islamic secondary schools in Selangor: A case study* [Master's thesis, International Islamic University Malaysia].

Moulana, S. J. (2017). *Synergy: Game design + Qur'an memorisation* [Master's thesis, Virginia Commonwealth University]. https://doi.org/10.25772/JT9P-WA53

Musa, M. G., Yusop, M. H. N., & Sopee, M. M. M. (2018). i-Tasmik mobile platform – Enabling Tahfiz students to memorise Al-Quran independently. In *Proceedings of the 2018 International Conference on*

Information and Communication Technology for the Muslim World (ICT4M 2018) (pp. 24-29). https://doi.org/10.1109/ICT4M.2018.00014

Mustafa, N. M., & Basri, M. (2014). A preliminary study on mobile Quranic memorisation for remote education learning using RFID technology: KUIS as a case study. In *Proceedings of the Global Conference on Language Practice & Information Technology (GLIT 2014)* (pp. 160-165).

Mustafa, N. M., Mohd Zaki, Z., Mohamad, K. A., Basri, M., & Ariffin, S. (2021). Development and alpha testing of EzHifz application: Al-Quran memorisation tool. *Advances in Human-Computer Interaction, 2021*, Article 5567001. https://doi.org/10.1155/2021/5567001

Mustafa, N. M., Zaki, Z. M., Mohamad, K. A., Basri, M., & Ariffin, S. (2019). The design of Quran memorisation tool using low-fidelity prototype. In *Advancing Technology Industrialisation through Intelligent Software Methodologies, Tools and Techniques* (Vol. 318, pp. 430-443). https://doi.org/10.3233/FAIA190069

Nafi, N. M., Mokhtar, W. K. A. W., & Imas, M. M. (2019). The Holy Quran memorisation in the globalisation era. *International Journal of Academic Research in Business and Social Sciences, 9*(11).https://doi.org/10.6007/IJARBSS/v9-i11/6579

Ni'mah, M. K., Muhtarom, A., & Marzuki, A. (2025). Implementation of the use of gamification learning media in increasing motivation for learning Al-Qur'an Hadis at MA Ma'arif Sukorejo. *Al-Ulum: Jurnal Pendidikan Islam, 6*(2), 373-387. https://doi.org/10.56114/al-ulum.v6i2.12260

Noor, N. M., Ismail, M., Yussof, R. L., & Yusoff, F. H. (2019). Measuring Tajweed augmented reality-based gamification learning model (TARGaLM) implementation for children in Tajweed learning. *Pertanika Journal of Science and Technology, 27*(4), 1821-1840.

Noor, N. M., Yussof, R. L., Yusoff, F. H., & Ismail, M. (2018). Gamification and augmented reality utilisation for Islamic content learning: The design and development of Tajweed learning. In *Proceedings of the International Conference on User Science and Engineering* (pp. 163-174). Springer. https://doi.org/10.1007/978-981-13-1628-9

Noradzan, H., Zaini, A. A., Rahman, R. A., Hussin, A., & Aziz, M. N. A. (2019). Embedding gamification approach in education. *International Journal on E-Learning and Higher Education, 10*(1), 29-44.

Nurtanto, M., Kholifah, N., Ahdhianto, E., Samsudin, A., & Isnantyo, F. D. (2021). A review of gamification impact on student behavioural and learning outcomes. *International Journal of Interactive Mobile Technologies, 15*(21), Article 22. https://doi.org/10.3991/ijim.v15i21.24381

Pranata, S., Suhada, H., Aini, Q., Rahardja, U., & Amrikhasanah, O. G. (2020). Learning management system using Shari'a gamification method for Qur'an Hafiz. *Journal of Advanced Research in Dynamical and Control Systems, 12*(2), 2532-2540.

Purbohadi, D., Rahmawati, B. R. N., & Setiyawan, H. (2019). Development of Qur'an memorisation learning model based on mobile learning. *Journal of Physics: Conference Series, 1381*(1), Article 012029. https://doi.org/10.1088/1742-6596/1381/1/012029

Radina, Husin, & Zubaidillah, M. H. (2020). Study of analysis of difficulties of elementary school children in memorising the Koran (case study MI Ummul Qura Amuntai). In *Proceedings of the 1st International Conference on Islamic and Social Education Interdisciplinary: Transforming multidimensional aspects via Islamic and social education* (Vol. 1, No. 1, pp. 71-88).

Rahman, A. A., Shakor, M. F. A., Sukari, A., Nozlan, N. N., & Lokman, N. W. (2022). The effectiveness of Global Tahfiz Game (GTG) gamification in teaching and learning in Malaysia. *Journal of Fiqhiyyat, 3*(1), 30-40. https://doi.org/10.47405/mjssh.v7i10.1785

Rosmansyah, Y., & Rosyid, M. R. (2017). Mobile learning with gamification for Al-Quran memorisation. In *Proceedings of the 2017 International Conference on Information Technology Systems and Innovation (ICITSI 2017)* (pp. 378-383). IEEE. https://doi.org/10.1109/ICITSI.2017.8267974

Saleem, A. (2015). *Does memorisation without comprehension result in language learning?* [Master's thesis, Cardiff University].

Salehuddin, K., Shahimim, M. M., Sulaiman, M. Z., & Zolkapli, R. B. M. (2019). Heat maps and scan paths: Qualitative eye-tracking evidence on how the Qur'an is memorised through reading. *Journal of Nusantara Studies, 4*(2), 318-334. https://doi.org/10.24200/jonus.vol4iss2pp318-334

Schunk, D. H. (2012). *Learning theories: An educational perspective* (6th ed.). Pearson.

Senan, N., Wan Ab Aziz, W. A., Othman, M. F., & Suparjoh, S. (2017). Embedding repetition (Takrir) technique in developing Al-Quran memorising mobile application for autism children. *MATEC Web of Conferences, 135*, Article 00076. https://doi.org/10.1051/matecconf/201713500076

Sevutra, R., & Erlinda, S. (2019). Application of Android-based gamification of memorising the Qur'an and Hadith using the Scott method (Penerapan gamification hafalan Alquran dan Hadis berbasis Android menggunakan metode Scott). *Jurnal Ilmu Komputer Dan Bisnis, 10*(2), 2276-2282. https://doi.org/10.47927/jikb.v10i2.175

Sezgin, S., & Yüzer, T. V. (2020). Analysing adaptive gamification design principles for online courses. *Behaviour & Information Technology*, 1-17. https://doi.org/10.1080/0144929X.2020.1817559

Shamsuddin, S. N. W., Bakar, N. F. A., Makhtar, M., Isa, W. M. W., Rozaimee, A., & Yusof, N. (2016). A framework for designing mobile Quranic memorisation tool using multimedia interactive learning method for children. *Journal of Theoretical and Applied Information Technology, 92*(1), 20-27.

Shukri, N. H. A., Nasir, M. K. M., & Razak, K. A. (2020). Educational strategies on memorising the Quran: A review of literature. *International Journal of Academic Research in Progressive Education and Development, 9*(2), 632-648. https://doi.org/10.6007/IJARPED/v9-i2/7649

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives*. Springer. https://doi.org/10.1007/978-1-4614-3546-4

Siregar, F. R., Sukmana, Y., & Rosmansyah, Y. (2023). The influence of gamification on Quran reading learning. In *Proceedings of the 10th International Conference on ICT for Smart Society (ICISS 2023)*.https://doi.org/10.1109/ICISS59129.2023.10291695

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment, 3*(3), 1-24.https://doi.org/10.1007/978-3-319-66715-7

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Umar, I. N., & Aziz, Z. A. (2015). The effects of multimedia with different modes of presentation on recitation skills among students with different self-regulated learning levels. *Procedia- Social and Behavioural Sciences, 197*, 1962-1968. https://doi.org/10.1016/j.sbspro.2015.07.584

Urh, M., Vukovic, G., Jereb, E., & Pintar, R. (2015). The model for introduction of gamification into e-learning in higher education. *Procedia -Social and Behavioural Sciences, 197*, 388-397. https://doi.org/10.1016/j.sbspro.2015.07.154

Yahya, S. A., & Salamun, H. (2021). Self-leadership and gamification as Madani society's human resource educational elements. In *Proceedings of the First International Conference on Science, Technology, Engineering and Industrial Revolution (ICSTEIR 2020)* (Vol. 536, pp. 431-437). Atlantis Press. https://doi.org/10.2991/assehr.k.210312.071

Yildiz, İ., Topçu, E., & Kaymakci, S. (2021). The effect of gamification on motivation in the education of pre-service social studies teachers. *Thinking Skills and Creativity, 42*, Article 100907.https://doi.org/10.1016/j.tsc.2021.100907

Yushaa, S. M., Mazlina, S., Khuzzan, S., & Hanid, M. (2021). Gamification elements and its impact on students. *Journal of Technology and Operations Management, 16*(2), 62-75. https://doi.org/10.32890/jtom2021.16.2.6

Yusuf, S. A. M., Noh, M. A. C., & Razak, K. A. (2019). Tahfiz teachers' ability in teaching the subject Hifz Quran in implementing the integrated curriculum Tahfiz (KBT) in secondary schools. *International Journal of Academic Research in Business and Social Sciences, 9*(5), 836-850. https://doi.org/10.6007/IJARBSS/v9-i5/6010

Zein, U. N., & Syahrudin. (2025). Implementation of gamification in Al-Quran learning at Madrasah Diniyah Al-Iman Kotabumi. *JISEI: Journal of Islamic Studies and Educational Innovation, 1*(3), 369-379. https://doi.org/10.61094/jisei.v1i3.60

Zulkefli, M., & Jamil, A. I. (2025). The development and implementation of game-based learning for Tasawwur Islam subjects: Insights from teachers and classroom integration. *Journal of Islamic Educational Research, 11*(2), 57-70. https://doi.org/10.22452/jier.vol11no2.6

Pertanika J. Sci. & Technol. 34 (2): 1265 - 1285 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Development of an Indoor Testing Facility to Standardise the Analysis of UAV Spraying Efficacy

Siti Amni Ismail1[1]*, Azmi Yahya[2], Ahmad Suhaizi Mat Su[3], Anas Mustafah[2], Norhayu Asib[4], and Emmanuel Otache Adanu[5]

[1]Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA, Jasin Campus, 77300 Jasin, Melaka, Malaysia

[2]Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[3]Department of Agricultural Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[4]Department of Crop Protection, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[5]Department of Agriculture Education, Federal College of Education (Technical), P.M.B. 60, Gombe, Nigeria

ABSTRACT

Workforce shortages, growing input prices, and shortcomings in pesticide usage have caused increasing challenges to the rising production of rice in Southeast Asia. Traditional knapsack and mist-blower spraying techniques are laborious, present dangers to workers, and lead to decreased field capacity with irregular pesticide deposition. As Unmanned Aerial Vehicles provide enhanced operating efficiency and accuracy (UAVs), it is now recognised as an acceptable choice. Nevertheless, UAV spraying possesses distinct aerodynamic challenges, which include rotor-induced downwash airflow, droplet drift, and irregular deposition patterns. Such challenges are not well-addressed by the present testing approaches. The present wind tunnel facilities and computational models were mainly developed for aircraft that are manned at high altitude. Henceforth, this article stresses producing a testing facility that can be operated indoors, particularly intended to assess the performance of UAVs spraying within typical and replicable settings. The facility allows for meticulous monitoring of downwash airflow, droplet formation, and deposition distribution, which offer valuable insights into spray patterns affected by UAV design, payload, flight speed, and nozzle configurations. To ensure that the experiments are repeatable, standard operating procedures were created. Earlier tests showed that the facility can recreate real-like conditions for UAV spraying and allows for systematic adjustments of varied settings.

ARTICLE INFO

Article history:
Received: 07 October 2025
Accepted: 08 April 2026
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.27

E-mail addresses:
amni@uitm.edu.my (Siti Amni Ismail)
azmiy@upm.edu.my (Azmi Yahya)
asuhaizi@upm.edu.my (Ahmad Suhaizi Mat Su)
anas_mustafah@upm.edu.my (Anas Mustafah)
norhayuasib@upm.edu.my (Norhayu Asib)
engr.adanu@gmail.com (Emmanuel Otache Adanu)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

The testing setup signifies a new benchmark procedure for studying UAV pesticide application. This facility not only presents a way to make spraying safer and more effective, but also supports ecologically sustainable practices in rice agriculture.

Keywords: Downwash airflow measurement, droplet distribution, indoor testing facility, spray deposition analysis, unmanned aerial vehicle (UAV) spraying

INTRODUCTION

Even though rice production frequently does not keep up with its consumption, rice continues to be the main staple food in Asia. In Malaysia, the demand for rice rose from 2.7 million metric tonnes in 2016 to an expected 3.2 million metric tonnes by 2027. Simultaneously, the amount of rice produced rose a little, from 1.75 to 1.9 million metric tonnes (Omar et al., 2019). The nation relies greatly on imported rice, mostly from Vietnam and Thailand, and it is about 70% self-sufficient. The growing disparity shows the need to improve the industry via mechanisation. Thus, the Eleventh Malaysia Plan (11MP) and Third and Fourth National Agricultural Policies (NAP3, NAP4) are among national plans that highlight this as a priority (Dardak, 2015).

Even though there is a considerable mechanisation in preparing the land and harvesting rice, the use of pesticides is still one of the least mechanised phases in cultivating rice. Although more than 90% of spraying is conducted with machines, the index of mechanisation is just 19.3%. This reveals that we still rely a lot on manual work (Muazu et al., 2014; Omar et al., 2019). Spraying is usually carried out with knapsack sprayers or mist blowers that are strenuous on the body, make workers exposed to dangerous chemicals, and do not provide good coverage. Field checks found that a large number of operators do not follow the right spraying amount and timing. This leads to inconsistent use of pesticides, less effectiveness, and more chemical waste (Ismail et al., 2019; Mairghany et al., 2018). This lack of efficiency results in cost increase, resistant pests, and health deterioration in both humans and the environment.

Unmanned Aerial Vehicles (UAVs) have emerged as an alternative technology for spraying. UAVs can operate in the fields without soil compaction, apply chemicals precisely and reduce operator exposure to pesticides. Studies report that the field capacities of using UAVs are 2.0-4.5 ha h^{-1} at application rates of 14-39 L ha^{-1} and outperforming knapsack sprayers (Giles & Billing, 2014). The adoption of UAVs has expanded to Malaysia's major rice schemes since 2018, due to its benefits, including higher efficiency and reduced labour demand (Faiçal et al., 2017; Morley et al., 2017).

Although the usage of UAVs brings many benefits, UAV spraying introduces new challenges. Rotor-induced downwash strongly affects droplet atomisation and transport. Fine droplets (<150 μm) are prone to drift, while coarse droplets limit canopy penetration

(Lan et al., 2017; Vernay et al., 2016). Non-uniform deposition across swaths has been widely reported in rice and citrus trials (Qin et al., 2016; Wang et al., 2017; Zhang et al., 2016), undermining pest control and increasing off-target risks (Kruger et al., 2019). The way rotor airflow works, along with varied payloads and flight conditions, makes spray dynamics complicated, rendering challenging optimisation.

Efforts to make simulations of UAV spraying have depended on computational fluid dynamics (CFD) and modifications of manned aircraft spray models like AGDISP (Agricultural Dispersal) and FSCBG (Forest Service Cramer Barry Grim). While CFD can visualise airflow and droplet interactions (Yang et al., 2018; Zheng et al., 2018), its accuracy depends on empirical validation. Legacy models, however, were developed for high-altitude fixed-wing aircraft and cannot fully represent low-altitude, multi-rotor UAV conditions (Teske & Barry, 1993; Yang et al., 2017). Experimental efforts using pitot tubes, anemometers, and strain gauges (Chen et al., 2017; Huang et al., 2018; Wu et al., 2019) provide insights but are fragmented, costly, and lack standardisation. Most focus on airflow measurement alone, with little integration of droplet deposition analysis.

Existing wind tunnel facilities, designed for manned aircraft, also fall short for UAV applications. They are unable to replicate the simultaneous vertical downwash and forward propulsion of UAVs when delineating droplet spectra at elevated horizontal airspeeds. Field experiments are crucial yet limited by unforeseen environmental changes. Thus, the optimisation of UAV spraying, which involves factors like flight altitude, velocity, nozzle type, and formulation, remains predominantly empirical, platform-dependent, and challenging to standardise. A controlled indoor testing facility can address these shortcomings by offering consistent and reproducible settings for UAV spraying studies. This would enable systematic assessment of downwash airflow, droplet dimensions, and deposition patterns under varied parameters. Significantly, it would facilitate the development of UAV-specific spray compositions and technologies aimed at reducing drift (de Oliveira et al., 2015; Wang et al., 2018).

This study presents the development of a dedicated indoor testing facility for evaluating UAV spraying in rice cultivation. The facility simulates field-relevant operating conditions within a controlled environment, allowing experiments to be conducted in a consistent and repeatable manner while enabling reliable measurement of results. To ensure consistency during testing, standard operating procedures were established, and systems for airflow and droplet measurement were incorporated for performance evaluation. Overall, the facility offers a controlled and configurable framework for investigating UAV spray behaviour under defined conditions and can provide a foundation for more systematic analysis of UAV spraying performance and contributes towards improving the efficiency and sustainability of UAV-based applications in rice cultivation.

METHODOLOGY

General Concept of the Development

The indoor testing facility was built in the Faculty of Engineering, Universiti Putra Malaysia. The functional floor area was roughly 23 m X 15 m, with a total ceiling height of 5 m, as shown in Figure 1. Internal partition walls were built to a height of 3 m to establish a confined testing area. The height was the maximum permitted in accordance with the present building structure. Even though the upper section of the area is left unobstructed, this arrangement permitted sufficient flow of air following every spraying session. This reduced the likelihood of excessive moisture accumulation or residual chemical concentration. This open upper clearance also served as a safety measure to avoid prolonged retention of pesticide atomisers. Access to the testing area was provided through two doors, which remained closed during experimental runs. All rig operations were monitored and controlled from an adjacent room separated by a shutter door, allowing tests to be conducted without exposing personnel to sprayed chemicals.

The indoor UAV testing facility was developed to evaluate the airflow profile of the UAV and to quantify the spraying deposition under controlled conditions, while allowing simulation of selected operational parameters observed in field applications. To get the same airflow profile and deposition effects as in the field, the facility was developed to emulate the actual operation conditions of the tested UAV in the field. An indoor facility was chosen to allow better control of the travelling operation of the tested UAV and, at the same time, to overcome the environmental disturbances that may affect the UAV spray deposition during the operation. The achieved overall length and height of the tested UAV were approximately 23 m and 4.5 m, respectively. The schematic diagram of the facility is shown in Figure 2.

The testing facility has two main sections. The first section consists of a rail support structure and its driving unit. Attached to the beam of the rail support structure was a moving carriage that was used to mount the tested UAV. The height of the mounting frame could be adjusted at three tested height levels, while the forward speed of the moving carriage can be regulated at three levels of flight speed. The required test flight heights and speeds were set accordingly before the evaluation test. The frame structure was used to secure the tested UAV onto the moving carriage.

The second section was a sampling platform structure and its adjusting height unit to cover the total sampling area for spray deposition. The sampling platform was constructed to have a total of 49 sampling points, and at each of these points, a sample holder stand was placed for holding a pressure sensor or water-sensitive paper in accordance with the mode of the test runs conducted.

This facility was also equipped with the respective measurement sensors to measure the downwash pressure produced by UAV rotational blades, speed of the carriage, ambient air,



Figure 1. Indoor laboratory layout

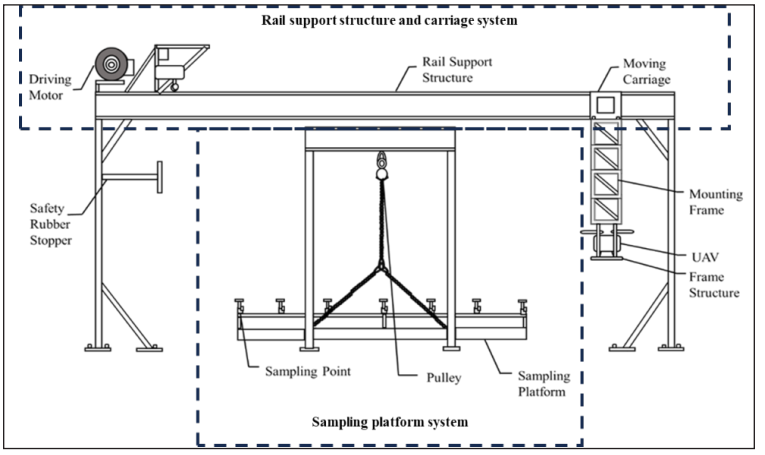


Figure 2. Schematic diagram of the indoor UAV testing facility showing (i) rail support and carriage system and (ii) sampling platform system

ambient humidity, and temperature of the test area. All of these measuring sensors were interfaced with a dedicated data acquisition system that could receive the measured signals in real time, control the information, display the information on a monitor screen and finally record the information into a storage medium.

Rail Support Structure Assembly

The railing beam and its two-end column supports were built from 305 mm x 102 mm x 33 kg/m mild steel I beam sections (Figure 3(a)). Attached to the beam was the moving carriage that was used to carry the tested UAV (Figure 3(b)). The moving carriage driving unit was driven by a 3-phase induction motor, brand TECO (model: AEEBAN) with a power size of 5.5kW. A pulley used to lift and hold the UAV during set-up was installed below the driving unit.

The maximum designed weight limit for moving carriage was 100 kg, and the unit could travel forward or reverse condition on rollers within an enclosed frame on the railing beam. The forward speed of the moving carriage was designed to be regulated up to the maximum linear speed of 10 m/s. This could be done by setting the required speed at the control console of the main driving unit. The moving carriage was used to hold the UAV and its mounting frame at a height of 3.5 m from the ground (Figure 3(c)). Changes to the height of the UAV from the ground could be adjusted by attaching the relevant mounting frame that could allow the carriage to hold the UAV at 1.5 m and 2.5 m from the ground (Figure 3 (d)). These three parameters of height and speed of the carriage would need to be set accordingly before the test evaluations were performed on the UAV.

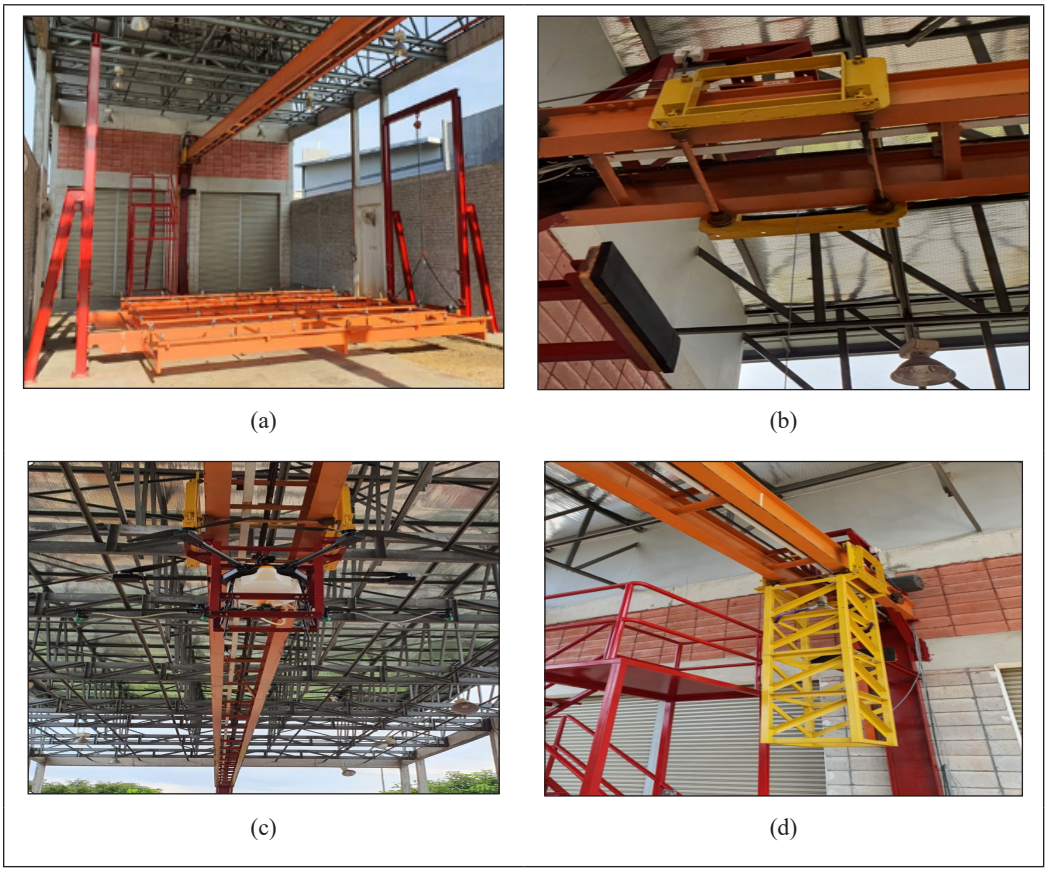


Figure 3. UAV testing facility main components: (a) rail support structure and sampling platform; (b) moving carriage at the highest level; (c) UAV mounted on moving carriage at the highest level; (d) moving carriage at the lowest level

Sampling Platform Structure Assembly

The sampling platform structure with a total sampling area of 6 m X 6 m was built from rectangular hollow steel located underneath the railing support structure. The entire sampling area was equipped with an adjustable lifting pulley at both sides that could raise the entire sampling platform up to a maximum of 3 m at 0.5 m increments. The sampling area was constructed with a total of 49 sampling points, where the location of the points could be adjusted to 1 m X 1 m (Figure 4(a)) or, 0.5 m X 0.5 m (Figure 4(b)) grid coordinates in accordance with the spray boom size of the UAV tested. The sampling holder stands were specifically made to fit the pressure transducer (Figure 5(a)) or the water-sensitive paper in their respective tests (Figure 5(b)). Each sampling holder stand could be rotated and adjusted accordingly to meet the desired setting.

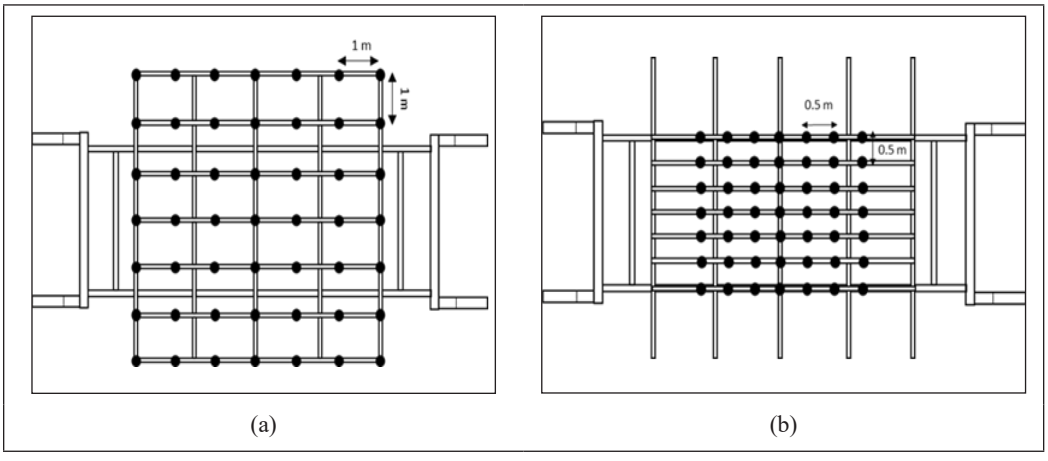


Figure 4. Sampling platform arrangement at (a) 1 m X 1 m grid and (b) 0.5 m X 0.5 m grid

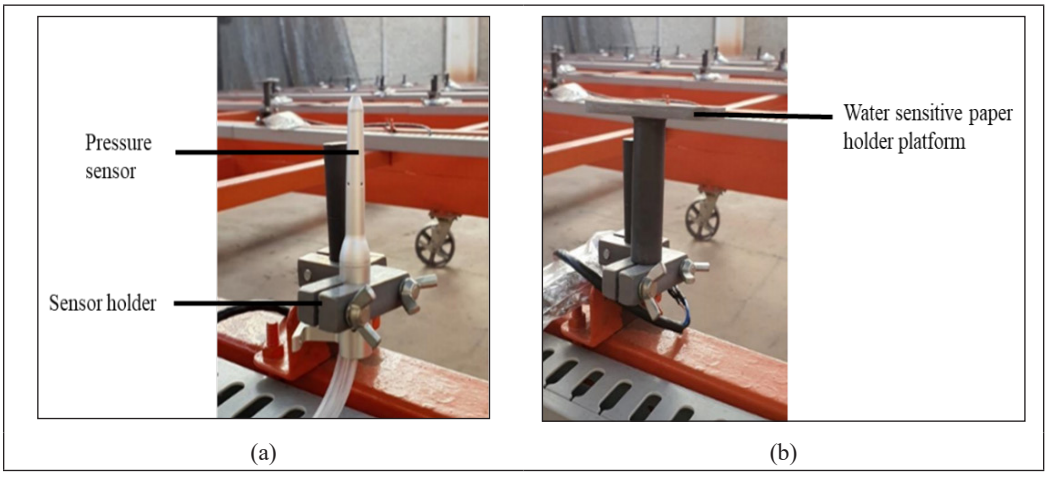


Figure 5. (a) Pressure sensor and (b) water-sensitive paper platform fixed at the sampling point

Data Acquisition System (DAQ)

Measuring Sensors

Forty-nine PT60 - MPXV7002 differential pressure transducers were fixed at every 49 standpoints on the sampling platform before the test commenced. The transducers were used to measure the downwash pressure produced by the UAV's rotating blades. It allows the measurement of the pressure of -2 to 2 kPa (-0.3 - 0.3 psi) through each port for pressure sensing and vacuum sensing, and has 0.5 to 4.5 V Output. The transducers were connected to the DAQ and recorded the differential pressure of the downwash during the test runs.

One inductive proximity encoder was installed at the main drive shaft and used in the triggering process. The encoder counts the rotation of the drive shaft as the carriage moves forward. The triggering started at the 10th rotation. Once triggered, all the pressure transducers began taking data and stopped when triggered again when the carriage was at the 21st rotation. In the interfacing software, the user would be able to set at which rotation the proximity encoder would start and stop to trigger the pressure sensors. Four pulse signal wind sensor digital anemometers, four units of LM35 temperature sensors and four units of HPP809A031-ND humidity sensors that were located at four different locations in the indoor facility were used to measure the ambient wind, temperature and humidity of the laboratory space. The height of all the transducers was 2 m from the ground and had a distance of 0.7 m from the wall. The readings from the four sensors were sent to the DAQ data acquisition to be recorded.

DAQ Hardware and Software Systems

The data acquisition system (DAQ) for acquiring data comprises the following: (i) transducers for acquiring the measured data, (ii) DAQ hardware for interfacing the sensors to a computer, (iii) computer, (iv) DAQ device drivers, and (v) DAQ software to process, visualise, and store the measurement data. The block diagram of the data acquisition system with associated transducers is shown in Figure 6.

The DAQ hardware and software systems are comprised of two main components. One unit of National Instrument (NI) cRIO- 9073 controller, 5 units of NI- 9205, one unit of NI-9422, one unit of NI-9411 modules, and one unit of desktop computer were used as hardware to acquire data from all the transducers. The modules were selected to fulfil their compatibility with the available transducers. All data processing from the signal acquired was done on the computer.

The NI-DAQmx driver from the National Instruments was installed on the desktop for communications to be made to all available devices in the system. Laboratory Virtual Instrument Engineering Workbench (LabVIEW) 2019 from the National Instruments was used as a system-design platform and development environment for developing the visual programme for data monitoring and acquisition.

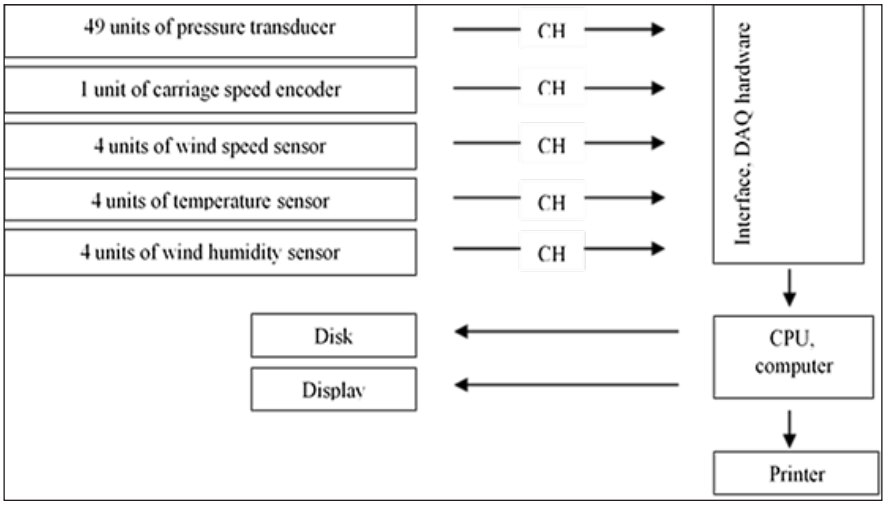


Figure 6. Block diagram of a data acquisition system (DAQ)

The display of data was made available to the user in real time on the monitor screen, and the data was permanently stored in a defined file in the storage medium of the computer. Once the data has been recorded in the designated files, it is readily extractable for further analysis. The types of data that were recorded were i) ambient data from all temperature, humidity and wind speed sensors, ii) average values of the temperature, humidity and wind speed sensors, iii) carriage speed, and lastly iv) data captured by pressure sensors.

Determining the Sampling Area

The sampling area was the area where the speed of the carriage would remain stable at the rated speed set by the user. The sampling area is important to be determined as a guideline to allocate the pressure sensors for the downwash data and water-sensitive paper for deposition verifications while the UAV is stable at the rated speed. The sampling area was defined as the section of the rig where the carriage maintained a stable velocity, ensuring accurate placement of pressure sensors and water-sensitive papers for downwash and deposition measurements. Stability of speed was crucial for reliable and repeatable experiments.

To determine the sampling area, the carriage movement was calibrated along the rail. Every sprocket turn was marked on the floor, with 27 stops reported (S1-S27). Speed was measured at various set speeds (2-7 m/s), making use of data from an inductive proximity encoder and timing software. This helped the plotting of the velocity profile over the distance travelled, allowing the determination of the areas where the motion was constant. The sampling area was determined by analysing the carriage speed profile along the rail system to identify regions of stable motion suitable for measurement.

Development of Testing Rig SOP

An SOP was developed for facility usage to allow consistent adherence to technical requirements and to support data quality. The SOP's developed for the testing procedure are described below.

Selection of UAV

Any UAV that is commonly used for aerial chemical spraying in wetland rice cultivation could be selected to undergo the test. As the UAV configurations are different from each other, before the test, the UAV used in this study needs to be measured for the purpose of building the mounting frame. Figure 7 shows an example of the UAV tested and the mounting frame built to mount the test UAV on the carriage.



Figure 7. UAV is placed in a custom-designed mounting frame before mounting onto the testing rig

Setting the Height, Speed, and Location of the Moving Carriage

The height and speed of the moving carriage of the rail support structure need to be set before executing the spraying evaluation by the tested UAV. Adjustments to the height of the moving carriage can be 1.5 m, 2.5 m, or 3.5 m from the ground. To perform the tests, a carriage mounting frame extender (Figure 8(a)) was used to hang the tested UAV to achieve the required heights from the ground (Figure 8(b)). The moving speed of the carriage could be set to a maximum speed of 10 m/s by setting the frequency at the Shihlin AC driver parameters (Shihlin Electric, SF-G Series AC Drive). Based on the available parameters, three different speeds (i.e., low, medium, and high) were set. In addition, the moving carriage had to be in the middle of the sampling area for the spraying or downwash evaluation under hovering conditions. For the evaluation of spraying under moving conditions, the moving carriage with the mounted UAV was set to move from the starting point location until the endpoint of the railing beam.

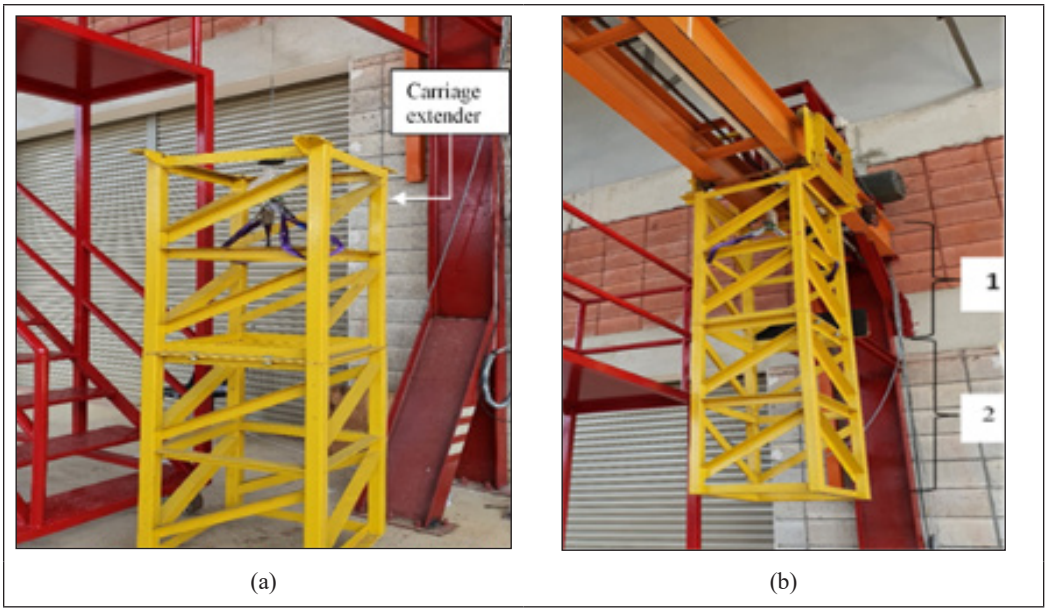


Figure 8. (a) A carriage extender used to create the height of the UAV; (b) Extenders 1 and 2 were used to create heights of 2.5 m and 1.5 m from the ground, respectively

Deciding the Area of the Sampling Point

The sampling point area had to be determined according to the spray boom size of the tested UAV. As such, the evaluation of the sampling points on the sampling platform structure was determined by fixing them at grids of 0.5 m X 0.5 m and 1 m X 1 m sampling points.

Deciding the Type of Evaluation

The indoor testing facility allows two main forms of assessment to be conducted. The first involves characterising the rotor-induced downwash using pitot tube sensors positioned at the designated grid points on the sampling platform. The second involves evaluating spray deposition, in which water-sensitive papers are placed at the same grid locations to capture droplet distribution from the UAV sprayer boom. When switching from airflow measurement to droplet deposition testing, the pressure sensors on the platform are removed and replaced with water-sensitive papers to ensure full coverage of the sampling grid.

Setting the Rotational Speed of the Blades

This was used to obtain accurate downwash and deposition data of the UAV using the developed testing facility; the UAV's blade rotational speed was set to be like the rotational speed of the UAV blades during actual field flight operation. To do so, a field test was carried out beforehand to obtain the rotational speed data of each UAV blade.

In this field test, the rotational speed of each blade on the UAV was recorded under various conditions of different payloads and flight speeds. After the field test, these recorded rotational speeds were accessed from the UAV flight controller console using the interfacing software. The rotational speed data was used as a reference in setting up the rotational speed of the individual blades during the evaluation test in the indoor testing facility. However, while the system allows consistent setting of rotational speed under defined payload and carriage movement conditions, detailed quantitative monitoring of rotor speed fluctuations during operation was not conducted in this study.

Based on the test that had been done using the testing rig, there are two methods in setting up the rotational speed of the blades to emulate the actual rotational speeds in the field. The first method was executed on Advansia UAV (A1) by bypassing the flight controller and instead connecting to the ESC of each motor directly to the receiver and external power source. In this study, 2 motors were connected to the servo channels of one receiver, where one transmitter controlled the speed of two motors. Thus, for the hexacopter, 3 transmitters were used. Throughout the tests, the spray pump connection has stayed connected to the flight controller and was activated or stopped using the UAV controller shown in Figure 9.

To set the speed of the blade at different combinations of flight speeds and payloads, the data collected from the field was taken as a guideline to determine the test variables. While setting the speed of the blade using a transmitter, the rotational speed of the blade could be checked or revised using a wireless tachometer that was fixed near the blade. The wireless tachometer of the Arduino IR sensor can read at a maximum of 10,000 rpm, and it is located close to the rotating motor. The reading from the sensor was sent to the receiver at every 0.5 s via radio frequency (RF), which has a limitation of 50 m distance of coverage. The receiver, on the other hand, was connected to the computer via a USB port, so that the reading of the sent rotational speed by the sensor was displayed in the computer using the Arduino IDE software (serial monitor). The process of receiving and setting the rotational speed of the blades is shown in Figure 10.

Another method of setting the rotational speed was executed on UAV Oryctes by Poladrone. The same procedure in collecting actual rotational speeds from the field was carried out with different combinations of payloads and flight speeds before the rotational speed setting process of the UAV blades and was later uploaded to the rotational speed setting system. The rotational speed setting process for this method was done by bypassing the flight controller and was instead controlled by the Raspberry Pi 3 B+. Raspberry Pi is a series of small single-board computers that are widely used in robotic and other computing projects. Python was used as the programming language. The Raspberry card was assembled on the UAV and remotely accessed by VNC® Connect software from a Windows PC via the wireless local area network (WLAN) shown in Figure 11.

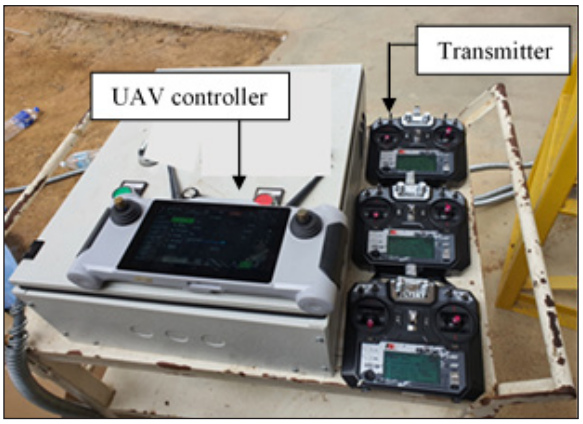


Figure 9. 3 Transmitter were used to regulate the rotational speeds of the 6 blades and 1 UAV controller to activate/stop the spray pump

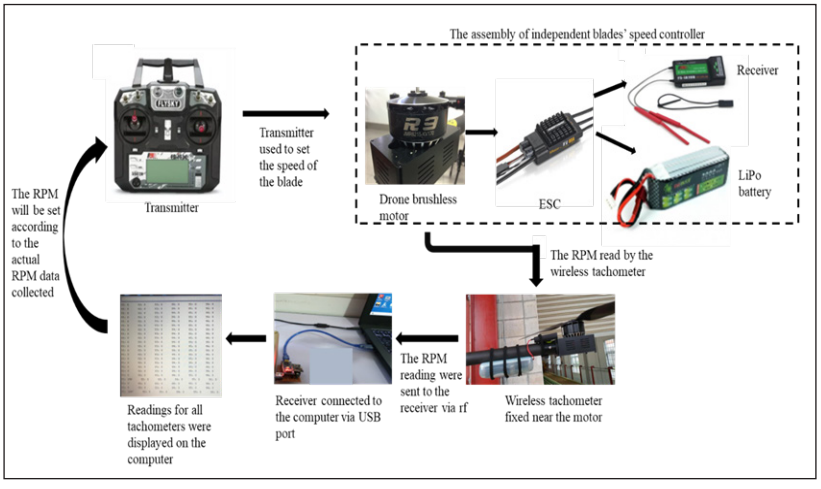


Figure 10. The process of receiving and setting the rotational speed of the blades

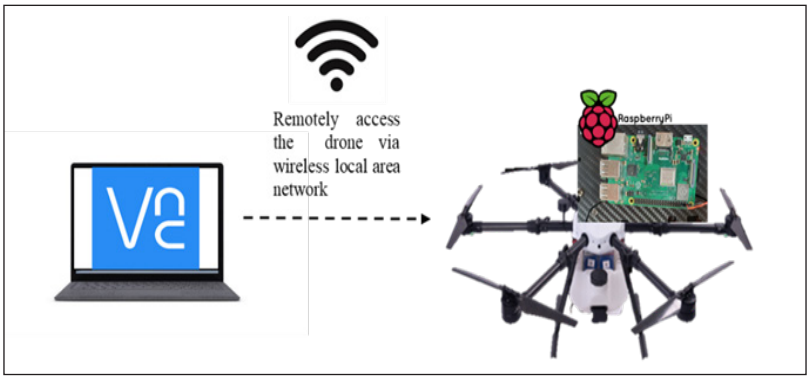


Figure 11. VNC viewer was used in remote communication with the Raspberry Pi card via WLAN

Acquiring Downwash Flow Field

The downwash airflow was assessed under two operating conditions: hovering and forward motion. During the hovering tests, the UAV was positioned at a specified height in the centre of the sampling area, and its rotor speed was set according to the selected operating parameter. The pressure sensors installed across the platform were used to record the resulting downwash. To generate a full three-dimensional profile of the pressure distribution, the measurements were repeated at several platform heights.

For the forward-motion condition, the blade rotational speed was adjusted to correspond with the intended flight velocity and payload. The vertical position and travelling speed of the carriage were then set to match these operating parameters. While the UAV traversed the rail system, the pressure sensors kept taking readings of the downwash, and these readings were being recorded at the same time via the data collection unit. The process was repeated at various platform height points to achieve the 3D pressure field during forward motion. MATLAB was later used to analyse and visualise all gathered pressure data (The MathWorks, Inc.).

Verifying the Droplet Deposition

To measure the distribution of droplet deposition, water-sensitive papers were fixed at sampling points on the sampling platform. The application parameters, such as blade rotational speed, height and speed of the carriage, were to be adjusted according to parameters determined in previous sections. The UAV spray tank was filled with the spray chemicals, and the spray system was set to run for at least 5 s in a stagnant position before the start of the test to ensure that the flow was stable. All water-sensitive papers were collected immediately upon the completion of the UAV test run. Each of the collected water-sensitive papers was scanned using an EPSON L360 scanner. The images were saved at a resolution of 8-bit, 600dpi as TIF. image format. The saved images were then further analysed using the DepositScan droplet deposition imaging system (USDA, USA) (Zhu et al., 2011). The deposition that was gathered from this test included a spatial droplet deposition pattern.

Loading and Unloading Pesticides

Spray chemicals were pre-filled into the spray tank earlier, while the UAV was still on the floor and were refilled while the UAV was hung on the rig using the stairs provided. For better accuracy of results, the remaining spray chemical in the tank was flushed before replacing it with another spray chemical.

Safety Precautions

Safety precautions implemented during the experimental procedures are summarised in Table 1.

Table 1
Safety precautions implemented during UAV spraying experiments

No.	Safety Aspect	Description
1	Handling of chemicals	All pesticide handling, loading, and disposal were carried out in accordance with manufacturer and regulatory guidelines
2	Equipment cleaning	Spraying equipment was cleaned before each test, and neutralising agents were applied when necessary
3	Personnel briefing	All personnel were briefed on test procedures before the operation
4	Restricted access	Personnel were kept clear of the UAV operation and spraying area during testing
5	Personal protective equipment (PPE)	Appropriate PPE was worn to minimise chemical exposure
6	Exposure awareness	Potential exposure routes (inhalation, ingestion, and dermal contact) were identified and controlled

Preparing Report

The test report was prepared to include detail documentation on the technical specification of the test UAV and related picture or technical drawing of the test UAV, spraying application parameters (i.e. flight height and flight speed), spray system technical specifications (i.e., spray pressure, flow rate, application rate, nozzle type, nozzle distance, and the number of nozzle), spraying material specifications (i.e., temperature of spraying materials and its physical properties), ambient conditions (i.e., air temperature, relative humidity and ambient air) and the details results from the evaluation. From the environmental data collected, the ambient temperature, wind and humidity of the laboratory space between 7 a.m. and 10 a.m. were recorded to be at around 24-28 ^{o}C, 0-0.83 m/s of ambient wind and around 44 %- 57% of humidity.

RESULTS AND DISCUSSION

Carriage Motion Stability and Sampling Area

The stability of carriage motion is a critical factor in ensuring reliable measurement of both downwash and droplet deposition. Stable speed regions were identified between S8–S20, with the most suitable sampling zone located between S11-S19. This ensured that sensors and water-sensitive papers were positioned within a region of uniform carriage velocity, thereby improving the consistency of experimental measurement. Figure 12 illustrates the final location of the sampling platform, selected based on overlapping stable regions across multiple test speeds.

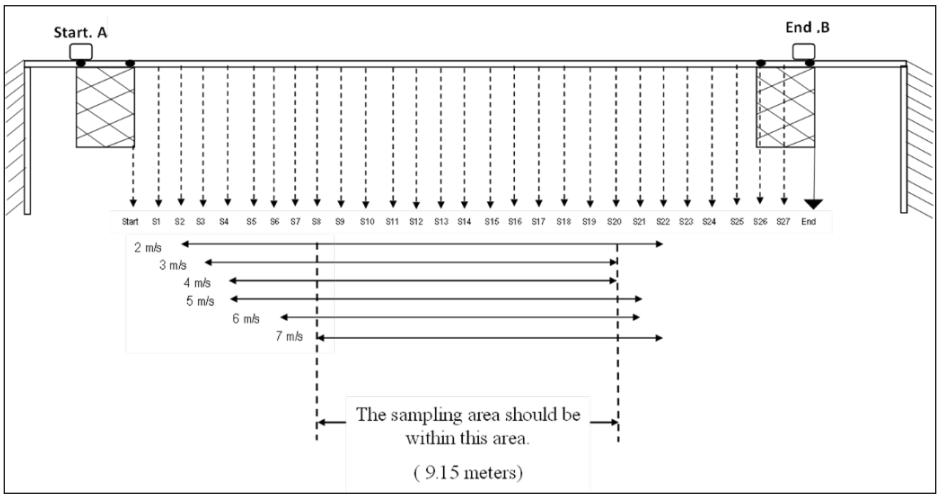


Figure 12. The location of the sampling area

Facility Performance and Downwash Measurement

The indoor testing facility was developed to provide a stable and repeatable environment for examining UAV spraying behaviour. Although the wall height was restricted to 3 m, preliminary observations indicated that this level of containment was sufficient to suppress unwanted air circulation and allow the spray cloud to settle before reaching the sampling grid. The adjustable flight height, carriage travel speed, and rotor speed controls enabled the UAV to be operated in conditions that approximate typical field spraying while maintaining experimental control.

Compared to wind tunnel-based methods, a more representative approach for analysing rotor-induced downwash effects on spraying in real situations is provided by this facility. Wind tunnel research usually keeps the UAV still or hovering (Grant et al., 2022; Wen et al., 2019; Yang et al., 2018). This can make it hard to see how the rotor's airflow interacts with the spray during actual forward movement. Likewise, tests performed at a set blade RPM (Wu et al., 2019) provide valuable aerodynamic data but fail to represent the ongoing changes that happen during real spraying. The existing facility overcomes these constraints by enabling regulated movement and configurable settings, which means the spray deposition patterns it creates more closely match what occurs in practical spraying conditions.

The facility was built to assess UAV downwash airflow and droplet deposition within a single experimental framework. However, while droplet deposition measurements were successfully demonstrated, the downwash pressure measurement subsystem did not produce a complete dataset due to intermittent sensor noise and signal instability. Nonetheless, the experimental configuration and data collection framework were completely developed.

The absence of comprehensive airflow information does not indicate a design defect but rather a hardware restriction. As a result, future optimisation efforts should emphasise sensor calibration and electrical shielding to enhance the reliability of the measurements. It is hoped that this subsystem is expected to provide quantitative datasets for the validation of computational fluid dynamics (CFD) simulations.

Droplet Deposition Characteristics

Verification of droplet deposition using water-sensitive papers revealed consistent spatial patterns across the sampling area. A higher droplet density was observed beneath the UAV centreline, with a gradual reduction towards the outer regions. This pattern matches what was documented in past studies about how UAV rotors create downwash during spraying (Lan et al., 2017; Wu et al., 2019). The ability of the facility to capture distinct and repeatable deposition patterns demonstrates its effectiveness as a controlled platform for analysing spray distribution under defined operating conditions.

The table of droplet deposition shown below (Table 2) represents deposition data collected using this testing rig. The selected flight heights and carriage speeds were chosen to represent typical operating conditions observed in UAV-based spraying and allow spray behaviour to be evaluated under controlled settings. Although the dataset is intended as a preliminary demonstration rather than a full validation, it shows that the facility can produce consistent and interpretable deposition data for comparison.

Table 2
Deposition image at every sampling point in a line

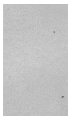
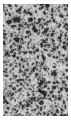
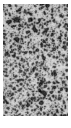
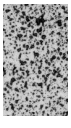
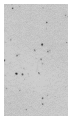
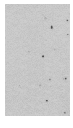
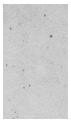
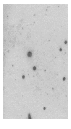
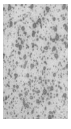
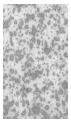
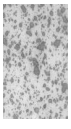
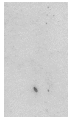
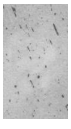
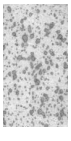
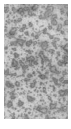
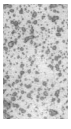
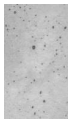
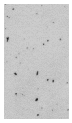
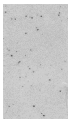
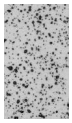
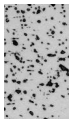
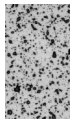
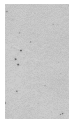
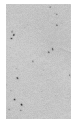
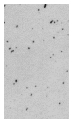
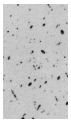
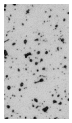
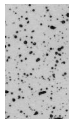
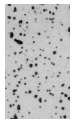
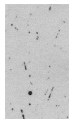
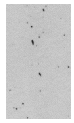
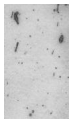
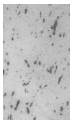
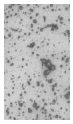
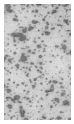
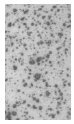
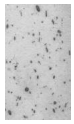
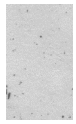
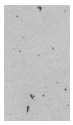
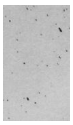
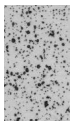
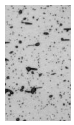
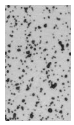
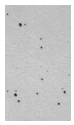
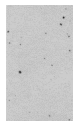
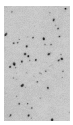
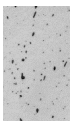
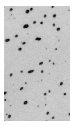
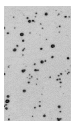
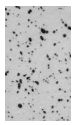
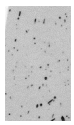
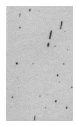
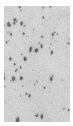
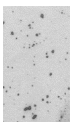
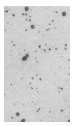
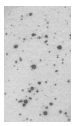
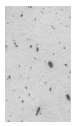
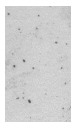
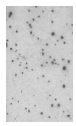
Parameter		**Deposition Image at every Sampling Point in a Line**						
Speed (m/s)	**Height (m)**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
2	1.5							
2	2.5							
2	3.5							

Table 2 (*continued*)

Parameter		Deposition Image at every Sampling Point in a Line						
Speed (m/s)	**Height (m)**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
4	1.5							
4	2.5							
4	3.5							
6	1.5							
6	2.5							
6	3.5							

Overall, the developed testing facility provides a structured experimental framework for investigating UAV spray behaviour under controlled conditions. Although full aerodynamic characterisation remains under development, the current system successfully demonstrates its capability for droplet deposition analysis and establishes a foundation for future studies integrating airflow measurement, spray optimisation, and drift reduction strategies.

CONCLUSION

This study presented an indoor setup for a testing facility to assess UAV spraying behaviour in a controlled and replicable setting. The setup enables modifying operational factors to observe how they impact droplet deposition. It offers a viable substitute for outdoor experiments where environmental variations may affect actual outcomes. The droplet deposition trials verified that the setup can generate consistent and clear spray patterns, demonstrating

its suitability for studies that compare and optimise various methods. The downwash pressure measuring subsystem has yet to produce dependable airflow data. However, including it in the facility links aerodynamic characterisation with deposition results. The facility, hence, provides a solid base for better studies on using UAV pesticide spraying. In the future, endeavours will concentrate on making the airflow sensing system stabilise, broadening the testing settings, and integrating computational modelling to predict how well the system works. These advancements are anticipated to make UAV spraying in agricultural applications more efficient and diminish the number of pesticides that miss the target.

ACKNOWLEDGEMENT

The authors would like to acknowledge the technical support provided by Advansia Sdn. Bhd and Poladrone (Aonic) Sdn. Bhd. Thank you to SEARCA for providing a scholarship under the SEARCA Postgraduate Scholarship Programme.

REFERENCES

Chen, S., Lan, Y., Li, J., Xu, X., Wang, Z., & Peng, B. (2017). Evaluation and test of effective spraying width of aerial spraying on plant protection UAV. *Transactions of the Chinese Society of Agricultural Engineering, 33*(7), 82-90. https://doi.org/10.11975/j.issn.1002-6819.2017.07.011

Dardak, R. (2015). *Transformation of agricultural sector in Malaysia through agricultural policy*. FFTC Agricultural Policy Platform. https://ap.fftc.org.tw/article/818

de Oliveira, R. B., Antuniassi, U. R., & Gandolfo, M. A. (2015). Spray adjuvant characteristics affecting agricultural spraying drift. *Engenharia Agricola, 35*(1), 109-116. https://doi.org/10.1590/1809-4430-Eng.Agric.v35n1p109-116/2015

Faiçal, B. S., Freitas, H., Gomes, P. H., Mano, L. Y., Pessin, G., de Carvalho, A. C. P. L. F., Krishnamachari, B., & Ueyama, J. (2017). An adaptive approach for UAV-based pesticide spraying in dynamic environments. *Computers and Electronics in Agriculture, 138*, 210-223.https://doi.org/10.1016/j.compag.2017.04.011

Giles, D. K., & Billing, R. (2014). Unmanned aerial platforms for spraying: Deployment and performance. *Aspects of Applied Biology, 122*, 63-69.

Grant, S., Perine, J., Abi-Akar, F., Lane, T., Kent, B., Mohler, C., Scott, C., … (2022). A wind-tunnel assessment of parameters that may impact spray drift during UAV pesticide application. *Drones, 6*(8), Article 204. https://doi.org/10.3390/drones6080204

Huang, C., Li, J., Lan, Y., Ye, W., Li, S., & Xing, H. (2018). Design and experiment of a wind speed detection equipment with wind pressure conversion near ground. *Journal of South China Agricultural University, 39*(1), 105-111. https://doi.org/10.7671/j.issn.1001-411X.2018.01.017 Ismail, S. A., Yahya, A., Mat Su, A. S., Asib, N., & Mustafah, A. M. (2019, July 21-24). Chemical spraying using an unmanned aerial system (UAS) in wetland rice cultivation in Malaysia. In *Proceedings of the 2019 ASABE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers. https://doi.org/10.13031/aim.201901854

Kruger, G. R., Klein, R. N., Ogg, C. L., & Vieira, B. C. (2019). *Spray drift of pesticides*. University of Nebraska Extension Publications.

Lan, Y., Chen, S., & Fritz, B. K. (2017). Current status and future trends of precision agricultural aviation technologies. *International Journal of Agricultural and Biological Engineering, 10*(3), 1-17.

Mairghany, M., Yahya, A., Adam, N. M., Mat Su, A. H. M., & Elsoragaby, E. (2018, February 7-8). Sustainability practices for pesticide spraying for wet paddy cultivation in Malaysia. In *Proceedings of the MSAE Conference 2018* (pp. 1-6).

Morley, C., Bradley, J., Hartley, R., Herries, D., & MacMorran, D. (2017). The potential of using unmanned aerial vehicles (UAVs) for precision pest control of possums (*Trichosurus vulpecula*). *Rethinking Ecology, 2*, 27-39. https://doi.org/10.3897/rethinkingecology.2.14821

Muazu, A., Yahya, A., Ishak, W. I. W., & Khairunniza-Bejo, S. (2014). Machinery utilisation and production cost of wetland, direct seeding paddy cultivation in Malaysia. *Agriculture and Agricultural Science Procedia, 2*, 361-369. https://doi.org/10.1016/j.aaspro.2014.11.050

Omar, S. C., Shaharudin, A., & Tumin, S. A. (2019). *The status of the paddy and rice industry in Malaysia*. Khazanah Research Institute.

Qin, W. C., Qiu, B. J., Xue, X. Y., Chen, C., Xu, Z. F., & Zhou, Q. Q. (2016). Droplet deposition and control effect of insecticides sprayed with an unmanned aerial vehicle against plant hoppers. *Crop Protection, 85*, 79-88. https://doi.org/10.1016/j.cropro.2016.03.018

Teske, M. E., & Barry, J. W. (1993). Parametric sensitivity in aerial application. *Transactions of the ASAE, 36*(1), 27-33. https://doi.org/10.13031/2013.28310

Vernay, C., Ramos, L., Douzals, J. P., Goyal, R., Castaing, J. C., & Ligoure, C. (2016). Drop impact experiment as a model experiment to investigate the role of oil-in-water emulsions in controlling the drop size distribution of an agricultural spray. *Atomization and Sprays, 26*(8), 827-851.https://doi.org/10.1615/AtomizSpr.2015013630

Wang, S. L., He, X. K., Song, J. L., Wang, S. S., Jia, X. M., & Ling, Y. (2018). Effects of xanthan gum on atomisation and deposition characteristics in water and silwet 408 aqueous solution. *International Journal of Agricultural and Biological Engineering, 11*(3), 29-34. https://doi.org/10.25165/j.ijabe.20181103.3802

Wang, S. L., Song, J. L., He, X. K., Song, L., Wang, X. N., Wang, C. L., Wang, Z. C., & Ling, Y. (2017). Performance evaluation of four typical unmanned aerial vehicles used for pesticide application in China. *International Journal of Agricultural and Biological Engineering, 10*(4), 22-31.https://doi.org/10.25165/j.ijabe.20171004.3219

Wen, S., Han, J., Ning, Z., Lan, Y., Yin, X., Zhang, J., & Ge, Y. (2019). Numerical analysis and validation of spray distributions disturbed by quad-rotor drone wake at different flight speeds. *Computers and Electronics in Agriculture, 166*, Article 105036. https://doi.org/10.1016/j.compag.2019.105036

Wu, Y., Qi, L., Zhang, H., Musiu, E. M., Yang, Z., & Wang, P. (2019). Design of UAV downwash airflow field detection system based on strain effect principle. *Sensors, 19*(11), Article 2630.https://doi.org/10.3390/s19112630

Yang, F. B., Xue, X. Y., Zhang, L., & Sun, Z. (2017). Numerical simulation and experimental verification on downwash air flow of six-rotor agricultural unmanned aerial vehicle in hover. *International Journal of Agricultural and Biological Engineering, 10*(4), 41-53. https://doi.org/10.25165/j.ijabe.20171004.3077

Yang, F., Xue, X., Cai, C., Sun, Z., & Zhou, Q. (2018). Numerical simulation and analysis on spray drift movement of multirotor plant protection unmanned aerial vehicle. *Energies, 11*(9), Article 2399.https://doi.org/10.3390/en11092399

Zhang, P., Deng, L., Lyu, Q., He, S. L., Yi, S. L., de Liu, Y., Yu, Y. X., & Pan, H. Y. (2016). Effects of citrus tree-shape and spraying height of small unmanned aerial vehicle on droplet distribution. *International Journal of Agricultural and Biological Engineering, 9*(4), 45-52. https://doi.org/10.3965/j.ijabe.20160904.2178

Zheng, Y., Yang, S., Liu, X., Wang, J., Norton, T., Chen, J., & Tan, Y. (2018). The computational fluid dynamic modeling of downwash flow field for a six-rotor UAV. *Frontiers of Agricultural Science and Engineering, 5*(2), 159-167. https://doi.org/10.15302/J-FASE-2018216

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

REFEREES FOR THE PERTANIKA

JOURNAL OF SCIENCE & TECHNOLOGY

Vol. 34 (2) Apr. 2026

The Editorial Board of the Pertanika Journal of Science and Technology wishes to thank the following:

A. K. M. Moyeenul Huq
(UAP, Bangladesh)

A.Muthusamy
(KEC, India)

Abdelkader Baghdad Bey
(UMSM, Algeria)

Abdul Nasir Aimi Salihah
(UniMAP, Malaysia)

Abinash Mahapatro
(SOA, India)

Adriano Michael Bernardin
(UNESC, Brazil)

Agus Indra Gunawan
(PENS, Indonesia)

Alaa Abdulhady Jaber
(UOT, Iraq)

Anam Naz
(UOS, Pakistan)

Arun Kumar MN
(FISAT, India)

Azyan Yusra Kapi @ Kahbi
(UiTM, Malaysia)

Balasubramanian Malaikozhundan
(GRI, India)

Basilis Boutsinas
(UPatras, Greece)

Beno Wincy Winsly
(CIT, India)

Bikash Sadhukhan
(TINT, India)

Biswarup Neogi
(JISCE, India)

Chua Stephanie
(UNIMAS, Malaysia)

Dulani Meedeniya
(UoM, Sri Lanka)

Elsadig Ahmed Mohamed Babiker
(PSAU, Saudi Arabia)

Farzaneh Mohamadpour
(USB,Iran)

Gan Chee Yuen
(USM, Malaysia)

Georgios Fotis
(AU, Denmark)

Goh Eg Su
(UTM, Malaysia)

Gopalakrishnan Velliyur Kanniappan
(BHU, Ethiopia)

Hasanudin
(UNSRI, Indonesia)

Hasnuri Mat Hassan
(USM, Malaysia)

Hazrul Abdul Hamid
(USM,Malaysia)

Hidefumi Yoshii
(SU, Japan)

Hong Ky Le
(VLUTE, Vietnam)

Huy Quy Trieu
(UNETI, Vietnam)

Irmawati Ramli
(UPM, Malaysia)

Ishak Mohd Yusoff
(UPM, Malaysia)

Ishak Pacal
(IGDIR, Turkey)

Ivan Ho Chun Wai
(UCSI, Malaysia)

Jayant Raut
(YCCE, India)

Keng Hoong Ng
(UTAR, Malaysia)

Khairun Nissa Mat Said
(UTM, Malaysia)

Laili Farhana Md Ibharim
(UPSI, Malaysia)

Low Foo Wah
(UTAR, Malaysia)

Luqman Chuah Abdullah
(UPM, Malaysia)

Manoj Kumar Mahawar
(CIRCOT, India)

Mehdi Jabbari Nooghabi
(FUM, Iran)

Mohamed Thariq Hameed Sultan
(UPM, Malaysia)

Murman Dwi Prasetio
(Tel-U, Indonesia)

Nadeem Sarwar
(BU, Pakistan)

Prasanna Kottapalle
(MCET, India)

Prashant Pande
(YCCE, India)

Qiao Zhang
(UPC, China)

Rajasekhar Nuvvusetty
(GRIET, India)

Raja Karmakar
(HIT, India)

Rashidah Ruslan
(UniMAP, Malaysia)

Riantini Vitriana
(ITB, Indonesia)

S.R.Mugunthan
(SRMIST, India)

Sheikh Ahmad Izaddin Sheikh Mohd Ghazali
(UiTM, Malaysia)

Shamshuritawati Sharif
(UUM, Malaysia)

Siva Kumar K
(MBU, India)

Suleiman Babani
(BUK, Nigeria)

Yu Fenghua
(SYAU, China)

Yazid Basthomi
(UM, Indonesia)

AU - *Aarhus Universitet*
BHU - *Bahir Dar University*
BU - *Bahria University*
BUK - *Bayero University Kano*
CIRCOT - *Central Institute for Research on Cotton Technology*
CIT - *Chennai Institute of Technology*
FISAT - *Federal Institute of Science and Technology*
FUM - *Ferdowsi University of Mashhad*
GRI - *Gandhigram Rural Institute (Deemed to be University)*
GRIET - *Gokaraju Rangaraju Institute of Engineering and Technology*
HIT - *Haldia Institute of Technology*
IGDIR - *Iğdır Üniversitesi*
ITB - *Institut Teknologi Bandung*
JISCE - *JIS College of Engineering*
KEC - *Kongu Engineering College*
MBU - *Mohan Babu University*
MCET - *Methodist College of Engineering and Technology*
PENS - *Politeknik Elektronika Negeri Surabaya*
PSAU - *Prince Sattam bin Abdulaziz University*
SOA - *Siksha 'O' Anusandhan (Deemed to be University)*
SRMIST - *SRM Institute of Science and Technology*
SU - *Setsunan University*
SYAU - *Shenyang Agricultural University*
Tel-U - *Telkom University*
TINT - *Techno India University*
UAP - *University of Asia Pacific*
UCSI - *UCSI University*
UiTM - *Universiti Teknologi MARA*
UM - *Universitas Negeri Malang*
UMSM - *University Mustapha Stambouli of Mascara*
UNESC - *University of the Extreme South of Santa Catarina*
UNETI - *University of Economics Technology for Industries*
UniMAP - *Universiti Malaysia Perlis*
UNIMAS - *Universiti Malaysia Sarawak*
UNSRI - *Universitas Sriwijaya*
UoM - *University of Moratuwa*
UOS - *University of Sargodha*
UOT - *University of Technology, Iraq*
UPatras - *University of Patras*
UPC - *China University of Petroleum (East China)*
UPM - *Universiti Putra Malaysia*
UPSI - *Universiti Pendidikan Sultan Idris*
USM - *Universiti Sains Malaysia*
USB - *University of Sistan and Baluchestan*
UTAR - *Universiti Tunku Abdul Rahman*
UTM - *Universiti Teknologi Malaysia*
UUM - *Universiti Utara Malaysia*
VLUTE - *Vinh Long University of Technology Education*
YCCE - *Yashwantrao Chavan College of Engineering*

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.

Design of an Integrated Entropy-based Framework for Dynamic Risk Weight Allocation and Mitigation Strategy Prioritisation in High-speed Rail Projects 1053
Yogesh P. Kherde, Uday P. Waghe, Radhika S. Thakre, Rajesh M. Bhagat, Sanskruti M. Padmawar, Yogeshwari D. Waje, Rishi S. Lohia, Anup K. Chitkeshwar, and Vaibhav Dhawale

TF-IDF Versus Linguistic Features: Evaluating Feature Sets for MBTI Personality Type Classification 1077
Janice Allison Anak Sabang, Stephanie Chua, and Puteri Nor Ellyza Nohuddin

Comparative Air Gasification of Agricultural Residues in an Indigenously Designed Fluidised Bed Gasifier 1107
Shahbaz Hussain, Abdul Nasir, Ch. Arslan, Shakeel Ahmad Anjum, and Sibel Irmak

Spectral Acoustic Characterisation for Bearing Fault Detection in Marine Propulsion Motors 1139
Iradiratu Diah Prahmana Karyatanti, Istiyo Winarno, Muhammad Taufiqurrohman, Dwisetiono, and Ardik Wijayanto

Quantification of Some Physical Properties of Matoa (*Pometia pinnata*) Fruit: Mass Modelling for Machine Designing and Postharvest Processing 1163
Ubaidillah Ubaidillah, Risty Dwi Wulandari Silalahi, Dimas Firmanda Al Riza, and Bambang Dwi Argo

FCA_Unet: An Enhanced UNet with Feature Coordination Attention for Wheat Leaf Disease Severity Assessment 1181
Hongyan Zang, Annie Anak Joseph, Kuryati binti Kipli, David Bong Boon Liang, Wanzhen Wang, and Rong Liu

An Interpretable Random Forest with SHAP Explanations for Multiclass Skill Level Classification Model in Malaysia's Labour Market 1203
Rabi'atul'adawiah Shabli, Ahmad Zia Ul-Saufie, Arief Gusnanto, and Nurain Ibrahim

How Gamification Motivates in Quranic Memorisation: A Method for Design and Suggestions for Future Research 1231
Siti Hasrinafasya Che Hassan, Syadiah Nor Wan Shamsuddin, and Nor Hafizi Yusof

Development of an Indoor Testing Facility to Standardise the Analysis of UAV Spraying Efficacy 1265
Siti Amni Ismail, Azmi Yahya, Ahmad Suhaizi Mat Su, Anas Mustafah, Norhayu Asib, and Emmanuel Otache Adanu

Green Synthesis of Carbon Dots from Empty Fruit Bunch Using Hydrothermal for Tetracycline Removal 823
Wan Nuraishah Wan Ishak, Ying Pei Lim, Huey Ling Tan, Noor Fitrah Abu Bakar, Ng Law Yong, and Chun Yang Yin

Effect of β-cyclodextrin Concentration on the Antioxidant Stability in an Avocado Oil Emulsion-based System 845
Carissa Eilin Ardana, Anggun Puspitarini Siswanto, Mohamad Endy Yulianto, Arva Rafif Adi Prayoga, Elsan Febiyanti, and Hermawan Dwi Ariyanto

Response Surface Optimisation of a *Strobilanthes crispus*-Loaded Red Palm Oil Nanoemulsion with Enhanced Physicochemical Stability 861
Ee Lynn Fong, Chui Fung Loke, Chin Ping Tan, Tze Ven Poh, Bee Hiong Tan, and Oi Ming Lai

OD-SAM: Automated Zero-shot Segment Anything Model for Optic Disc Segmentation 889
Bhuvaneswari S and Subashini P

Review Article 913
A Critical Review of User-Centric QoS Provisioning Frameworks in 5G Networks
Wai Leong Pang, Wen Hao Anselm Chow, Hui Hwang Goh, Swee King Phang, and Kah Yoong Chan

Improved Speech Enhancement using Parallel MVDR Beamforming and Coherent-to-Diffuse Power Ratio Post-filtering 951
Sureshkumar Natarajan, Mohd Khair Hassan, Raja Kamil, Faisul Arif Ahmad, Syaril Azrad, June Francis Macleans, Hussna Elnoor M. Abdalla, Nurbek Saparkhojayev, and Syed Abdul Rahman Al-Haddad

A Genetic Algorithm-optimised Hybrid Framework Integrating Statistical Forecasting Models and Extreme Learning Machine for Dengue Case Forecasting in the United States of America 975
Dipankar Das, Arijit Chakraborty, Sajal Mitra, and Soumyabrata Nag

Review Article 1005
Visualising Global Trends in Soil Infiltration Study by using Bibliometric Analysis
Nor Farah Atiqah Ahmad, Siti Nooraiin Mohd Razali, Suhaila Sahat, and Muhammad Azraie Abdul Kadir

Integrating Time Lag Effects in Predictive Modelling Using Random Forest for Early Detection of Bagworm Outbreaks 1025
Yi Peng Wang, Nurul Hawani Idris, Norhayu Asib, Mohamad Hafis Izran Ishak, and Alvin Lau Meng Shin

Pertanika Journal of Science & Technology
Vol. 34 (2) Apr. 2026

Content

Foreword i
Mohamed Thariq Hameed Sultan

Multi-classification of Chest X-ray Images Using an Enhanced Deep Learning Approach 601
Revathi A and Savadam Balaji

Skin Cancer Classification from Dermoscopic Images based on Convolutional Neural Network and Vision Transformer 625
Gan Mengyun, Hazlina Hamdan, Mahmud Dwi Sulistiyo, and Abdulrahman Dira Khalaf

Design, Manufacturing, and Testing a Versatile Test Stand for MP100S Integration 647
Huu Danh Tran

Measuring the Performance of Blockchain Systems with an Improved Consensus Mechanism using Scalability Metrics 671
Kolli Lalitha Kumari and Lalitha Surya Kumari

Association Rule Mining to Detect Significant Factors of Autism Spectrum Disorder at Early Stage 695
Tania Akter, Farzana Tasnim, Shahriar Hassan, Mohammad Hanif Ali, and Mohammad Shorif Uddin

JaNarX: Smart Soil Moisture Prediction using Jackal Optimiser Enabled Machine Learning Model in IoT Coupled Agricultural Applications 721
Seema Jitendra Patil and B. Ankayarkanni

Interactive Location-based Augmented Reality Virtual Chatbot 743
Lai Yee Jen, Goh Eg Su, and Johanna Ahmad

Ensemble Docking and Molecular Dynamics Pattern Similarity Analysis for the Discovery of Dengue Virus RNA-Dependent RNA Polymerase Inhibitor 763
Amaal Mohammed Salih Nasr, Mohd Basyaruddin Abdul Rahman, Muhammad Alif Mohammad Latif, Faridah Abas, Siti Munirah Mohd Faudzi, and Bimo A. Tejo

Optimisation of Silver Nanoparticle Synthesis Using Aqueous Extract of *Passiflora foetida* l. via Response Surface Methodology 787
Silvia Handayani, Alyza A. Azmi, R. Rudiyanto, Wan Iryani Wan Ismail, Kustiariyah Tarman, Faridah Abas, Wan M. Khairul, and M. Maulidiani

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

http://penerbit.upm.edu.my
Email: penerbit@upm.edu.my
Tel. No.: +603- 9769 3606

e-ISSN 2231-8526