

Pertanika Journal of
**SCIENCE &
TECHNOLOGY**

JST

VOL. 30 (4) OCT. 2022



PERTANIKA
JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

PERTANIKA JOURNAL OF SCIENCE & TECHNOLOGY

About the Journal

Overview

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

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

Pertanika Journal of Science & Technology is a **quarterly** (*January, April, July, and October*) periodical that considers for publication original articles as per its scope. The journal publishes in **English** and it is open for submission by authors from all over the world.

The journal is available world-wide.

Aims and scope

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

History

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

Vision

To publish a journal of international repute.

Mission

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

Quality

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

Abstracting and indexing of Pertanika

Pertanika Journal of Science & Technology is now over 27 years old; this accumulated knowledge and experience has resulted the journal being abstracted and indexed in SCOPUS (Elsevier), Clarivate Web of Science (ESCI), EBSCO, ASEAN CITATION INDEX, Microsoft Academic, Google Scholar, and MyCite.

Citing journal articles

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

Publication policy

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

Code of Ethics

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

Originality

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

All submitted manuscripts must be in the journal's acceptable similarity index range:
≤ 20% – PASS; > 20% – REJECT.

International Standard Serial Number (ISSN)

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

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

Lag time

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

Authorship

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

Manuscript preparation

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words *Introduction, Materials and Methods, Results, And Discussion*. IMRAD is simply a more 'defined' version of the "IBC" (*Introduction, Body, Conclusion*) format used for all academic writing. IMRAD indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: *Title, Authors, Keywords, Abstract, Conclusions, References, and Acknowledgement*. Additionally, some papers include *Appendices*.

The *Introduction* explains the scope and objective of the study in the light of current knowledge on the subject; the *Materials and Methods* describes how the study was conducted; the *Results* section reports what was found in the study; and the *Discussion* section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the journal's **Instruction to Authors** (http://www.pertanika.upm.edu.my/Resources/regular_issues/Regular_Issues_Instructions_to_Authors.pdf).

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.

As articles are double-blind reviewed, material that may identify authorship of the paper should be placed only on page 2 as described in the first-4-page format in *Pertanika*'s **Instruction to Authors** (http://www.pertanika.upm.edu.my/Resources/regular_issues/Regular_Issues_Instructions_to_Authors.pdf).

The journal's peer review

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

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

Operating and review process

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

1. The journal's Chief Executive Editor and the Editor-in-Chief examine the paper to determine whether it is relevance to journal needs in terms of novelty, impact, design, procedure, language as well as presentation and allow it to proceed to the reviewing process. If not appropriate, the manuscript is rejected outright and the author is informed.
2. The Chief Executive Editor sends the article-identifying information having been removed, to 2 to 3 reviewers. They are specialists in the subject matter of the article. The Chief Executive Editor requests that they complete the review within 3 weeks.

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

3. The Editor-in-Chief examines the review reports and decides whether to accept or reject the manuscript, invite the authors to revise and resubmit the manuscript, or seek additional review reports. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the authors) are forwarded to the authors. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.
4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the Chief Executive Editor along with specific information describing how they have addressed' the concerns of the reviewers and the editor, usually in a tabular form. The authors may also submit a rebuttal if there is a need especially when the authors disagree with certain comments provided by reviewers.
5. The Chief Executive Editor sends the revised manuscript out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.
6. When the reviewers have completed their work, the Editor-in-Chief examines their comments and decides whether the manuscript is ready to be published, needs another round of revisions, or should be rejected. If the decision is to accept, the Chief Executive Editor is notified.
7. The Chief Executive Editor reserves the final right to accept or reject any material for publication, if the processing of a particular manuscript is deemed not to be in compliance with the S.O.P. of *Pertanika*. An acceptance letter is sent to all the authors.

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

Pertanika Journal of

**SCIENCE
& TECHNOLOGY**

Vol. 30 (4) Oct. 2022



A scientific journal published by Universiti Putra Malaysia Press



EDITOR-IN-CHIEF

Luqman Chuah Abdullah
Chemical Engineering

CHIEF EXECUTIVE EDITOR

UNIVERSITY PUBLICATIONS COMMITTEE

CHAIRMAN

Nazamid Saari

EDITORIAL STAFF

Journal Officers:

Kanagamaral Silvarajoo, *ScholarOne*
Siti Zuhaila Abd Wahid, *ScholarOne*
Tee Syn Ying, *ScholarOne*
Ummi Fairuz Hanapi, *ScholarOne*

Editorial Assistants:

Ku Ida Mastura Ku Baharom
Siti Juridah Mat Arip
Zulinaardawati Kamarudin

English Editor:

Norhanisah Ismail

PRODUCTION STAFF

Pre-press Officers:

Nur Farrah Dila Ismail
Wong Lih Jiun

WEBMASTER

IT Officer:

Illi Najwa Mohamad Sakri

EDITORIAL OFFICE

JOURNAL DIVISION

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

Gen Enquiry
Tel. No: +603 9769 1622 | 1616
E-mail:
executive_editor.pertanika@upm.edu.my
URL: www.journals-jd.upm.edu.my

PUBLISHER

UPM Press
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia.
Tel: +603 9769 8851
E-mail: penerbit@putra.upm.edu.my
URL: <http://penerbit.upm.edu.my>



ASSOCIATE EDITOR

2021-2023

Adem Kilicman
Mathematical Sciences
Universiti Putra Malaysia, Malaysia

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

2020-2022

Abdul Latif Ahmad
Chemical Engineering
Universiti Sains Malaysia, Malaysia

Hsiu-Po Kuo
Chemical Engineering
National Taiwan University, Taiwan

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

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

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

Najafpour Darzi Ghasem
Bioprocess Technology, Chemical Engineering, Water and Wastewater Treatment Technology, Biochemical Engineering and Biotechnology, Bioethanol, Biofuel, Biohydrogen, Enzyme and Fermentation Technology
Babol Noshirvani University of Technology, Iran

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

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

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

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

Mohd Sapuan Salit
Concurrent Engineering and Composite Materials
Universiti Putra Malaysia, Malaysia

Norbahiah Misran
Communication Engineering
Universiti Kebangsaan Malaysia, Malaysia

Daud Ahmad Israf Ali
Cell Biology, Biochemical, Pharmacology
Universiti Putra Malaysia, Malaysia

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

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

Hari M. Srivastava
Mathematics and Statistics
University of Victoria, Canada

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

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

INTERNATIONAL ADVISORY BOARD

2021-2024

CHUNG, Neal Tai-Shung
Polymer Science, Composite and Materials Science
National University of Singapore, Singapore

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

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

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

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

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

The journal is indexed in SCOPUS (Elsevier), Clarivate-Emerging Sources Citation Index (ESCI), BIOSIS, National Agricultural Science (NAL), Google Scholar, MyCite, ISC. In addition, Pertanika JSSH is recipient of "CREAM" Award conferred by Ministry of Higher Education (MoHE), Malaysia.



Pertanika Journal of Science & Technology
Vol. 30 (4) Oct. 2022

Contents

Foreword <i>Chief Executive Editor</i>	i
Application of Groupwise Principal Sensitivity Components on Unbalanced Panel Data Regression Model for Gross Regional Domestic Product in Kalimantan <i>Desi Yuniarti, Dedi Rosadi and Abdurakhman</i>	2315
Evaluation of Subcritical Organic Rankine Cycle by Pure and Zeotropic of Binary and Ternary Refrigerants <i>Omid Rowshanaie, Mohd Zahirasri Mohd Tohir, Faizal Mustapha, Mohammad Effendy Ya'acob and Hooman Rowshanaie</i>	2333
Optimal Material Selection for Manufacturing Prosthetic Foot <i>Fahad Mohanad Kadhim, Samah Falah Hasan and Sadiq Emad Sadiq</i>	2363
Development of Regression Models for the Prediction of NiTi Archwire Forces in an Orthodontic Bracket System Using Response Surface Methodology <i>Sareh Aiman Hilmi Abu Seman, Ching Wei Ng, Muhammad Fauzinizam Razali and Abdus Samad Mahmud</i>	2377
Robust Linear Discriminant Rule Using Double Trimming Location Estimator with Robust Mahalanobis Squared Distance <i>Yik Siang Pang, Nor Aishah Ahad and Sharipah Soaad Syed Yahaya</i>	2393
Design Guidelines for Vegetative Approach Application to Mitigate Riverbank Landslides in Mangkang Wetan and Mangunharjo Villages, Semarang Coastal Area, Indonesia <i>Santy Paulla Dewi and Rina Kurniati</i>	2407
Quality Evaluation of Color and Texture of the Dabai Fruit (<i>Canarium odontophyllum</i> Miq.) at Different Temperatures and Times of Blanching <i>Rosnah Shamsudin, Siti Hajar Ariffin, Wan Nor Zanariah Zainol @ Abdullah, Nazatul Shima Azmi and Arinah Adila Abdul Halim</i>	2427
<i>Review Article</i> Computed Tomography and Other Imaging Modalities in Pediatric Congenital Heart Disease <i>Hongying Chen, Norafida Bahari, Suraini Mohamad Saini and Noramaliza Mohd Noor</i>	2439

Evaluation of Phytochemical and Mineral Composition of Malaysia's Purple-Flesh Sweet Potato <i>Rosnah Shamsudin, Nurfarhana Shaari, Mohd Zuhair Mohd Noor, Nazatul Shima Azmi and Norhashila Hashim</i>	2463
Cultivation of <i>Rhodotorula Toruloides</i> Using Palm Oil Mill Effluent: Effect on the Growth, Lipid Production, and Waste Removal <i>Ignatia Justine, Grace Joy Wei Lie Chin, Wilson Thau Lym Yong and Mailin Misson</i>	2477
The Relationship Between Bangladesh's Financial Development, Exchange Rates, and Stock Market Capitalization: An Empirical Study Using the NARDL Model and Machine Learning <i>Rehana Parvin</i>	2493
Organ Dose and Radiation Exposure Risk: A Study Comparing Radiation Dose Using Two Software Packages <i>Abdullah Ali M Asiri</i>	2509
Effective Emergency Management: Scrutinizing the Malaysia Lead Responding Agency Planning and Information Management Approach During Disaster Exercise <i>Khairilmizal Samsudin, Fatin Najihah Ghazali, Nur Hannani Abdul Ghani, Mohamad Fahmi Hussin, Ainul Husna Kamarudin and Haikal Kasri</i>	2521
A Topic Modeling and Sentiment Analysis Model for Detection and Visualization of Themes in Literary Texts <i>Kah Em Chu, Pantea Keikhosrokiani and Moussa Pourya Asl</i>	2535
Data Safety Prediction Using Bird's Eye View and Social Distancing Monitoring for Penang Roads <i>Lek Ming Lim, Majid Khan Majahar Ali, Mohd. Tahir Ismail and Ahmad Sufiril Azlan Mohamed</i>	2563
Effect of Process Conditions on Catalytic Hydrothermal Oxidation of <i>p</i> -Xylene to Terephthalic Acid <i>Mohamad Zarqani Yeop, Kamariah Noor Ismail and Ahmad Rafizan Mohamad Daud</i>	2589
Hole Quality Assessment of Drilled Carbon Fiber Reinforced Polymer (CFRP) Panel Using Various Custom Twist Drill Geometries <i>Muhammad Faris Shauqi Kamaruzaman, Muhammad Hafiz Hassan and Muhammad Fauzinizam Razali</i>	2603

ZnO Multilayer Thin Films as the Seed Layer for ZnO Nanorods: Morphology, Structural and Optical properties <i>Rohanieza Abdul Rahman, Muhammad AlHadi Zulkefle, Sukreen Hana Herman and Rosalena Irma Alip</i>	2613
<i>Review Article</i>	
A Review on 3D Augmented Reality Design Technique and Inward Leakage Testing on Protective Face Mask <i>Nur Amirah Kamaluddin, Murizah Kassim and Shahrani Shahbudin</i>	2639
Suppression of Coffee-Ring Effect on Nitrocellulose Membrane: Effect of Polyethylene Glycol <i>Sarah Sorfeena Shahrudin, Norhidayah Ideris and Nur Atikah Kamarulzaman</i>	2659
Multi-Objective Optimal Control of Autocatalytic Esterification Process Using Control Vector Parameterization (CVP) and Hybrid Strategy (HS) <i>Fakhrony Sholahudin Rohman, Dinie Muhammad, Iylia Idris, Muhamad Nazri Murat and Ashraf Azmi</i>	2669
Automated Cryptocurrency Trading Bot Implementing DRL <i>Aisha Peng, Sau Loong Ang and Chia Yean Lim</i>	2683
RNA Isolation from Environmental Samples of a Harmful Algal Bloom for Metatranscriptome Next-Generation Sequencing <i>Diana Lorons, Kenneth Francis Rodrigues, Madihah Jafar Sidik and Grace Joy Wei Lie Chin</i>	2707
A Comparative Study of RNA-Seq Aligners Reveals Novoalign's Default Setting as an Optimal Setting for the Alignment of HeLa RNA-Seq Reads <i>Kristine Sandra Pey Adum and Hasni Arsad</i>	2727
The Kinetic Evaluation and DFT Study of Cis-[Pt(Asc)(NH ₃) ₂] Complex as an Inhibitor to Type 2 Diabetic Human Amylase <i>Khalid Farouk Al-Rawi, Khaldoon Taher Maher, Othman Ibrahim Alajrawy and Firas Taher Maher</i>	2747
Comparative Computational Study of Double Rotating Cylinder Embedded on Selig S1223 Aerofoil and Flat Plate for High Altitude Platform <i>Hidayatullah Mohammad Ali, Azmin Shakrine Mohd Rafie, Mohd Faisal Abdul Hamid and Syaril Azrad Md. Ali</i>	2767
Risk Assessment on Robotic Surgery Using Bayesian Network <i>Teh Raihana Nazirah Roslan and Chee Keong Ch'ng</i>	2789

<i>Review Article</i>	
A Review on the Effect of Zein in Scaffold for Bone Tissue Engineering <i>Istikamah Subuki, Khairun Nor Ashikin Nasir and Nur Azrini Ramlee</i>	2805
Robust Hybrid Classification Methods and Applications <i>Friday Zinzendoff Okwonu, Nor Aishah Ahad, Innocent Ejiro Okoloko, Joshua Sarduana Apanapudor, Saadi Ahmad Kamaruddin and Festus Irimisose Arunaye</i>	2831
Significant Factors in Agile Software Development of Effort Estimation <i>Pantjawati Sudarmaningtyas and Rozlina Mohamed</i>	2851
Transformer Population Failure Rate State Distribution, Maintenance Cost and Preventive Frequency Study Based on Markov Model <i>Nor Shafiqin Shariffuddin, Norhafiz Azis, Jasronita Jasni, Mohd Zainal Abidin Ab Kadir, Muhammad Sharil Yahaya and Mohd Aizam Talib</i>	2879
MOFTI: Mobile Self-Service Framework to Uplift Customer Experience for the Telecommunications Industry in Malaysia <i>Zainab Abu Bakar, Hazrina Sofian and Nazean Jomhari</i>	2903
<i>Review Article</i>	
Dehydrated Food Waste for Composting: An Overview <i>Aziz Khalida, Veknesh Arumugam, Luqman Chuah Abdullah, Latifah Abd Manaf and Muhammad Heikal Ismail</i>	2933
Perceived Benefits, Perceived Barriers and Self-Efficacy Towards Weight Reduction Among Overweight and Obese Children in Kedah, Malaysia <i>Wan Mohd Nurussabah Abdul Karim, Hazizi Abu Saad, Nurzalinda Zalbahar and Nurul Husna Mohd Sukri</i>	2961
Evaluation of Acoustical Performance for Atrium Design with Respect to Skylight Geometry and Material in the Tropics <i>Nazli Che Din, Tio Seng Zhi and Chia Keng Soon</i>	2983

Foreword

Welcome to the fourth issue of 2022 for the *Pertanika Journal of Science and Technology (PJST)*!

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

This issue contains 35 articles; four review articles and the rest are regular articles. The authors of these articles come from different countries namely Bangladesh, Indonesia, Iraq, Japan, Malaysia, Nigeria and Saudi Arabia.

An investigation on optimal material selection for manufacturing prosthetic feet was conducted by Fahad Mohanad Kadhim et al. from Iraq. This investigation simulates chosen materials to find the optimal material selection for manufacturing prosthetic feet by assuming boundary conditions, reaction forces, design considerations, and application. The result shows that the Hybrid composite material has excellent improvement in mechanical properties such as lightweight, stiffness, high mechanical properties, and cost-efficiency. Furthermore, by considering the body weight of the amputee and gait cycle, and analyzing the material properties, the hybrid composite material that is the best suitable should be selected to manufacture foot prostheses. Further details of the investigation can be found on page 2363.

A review on computed tomography and other imaging modalities in pediatric congenital heart disease was conducted by Hongying Chen and colleagues from Universiti Putra Malaysia. Congenital heart defects (CHD) are the most common congenital disabilities. Therefore, an early and accurate diagnosis of coronary heart disease is very important for patients to get timely and effective treatment. At present, the clinical application of echocardiogram (echo), cardiovascular magnetic resonance (CMR) and computed tomography angiography (CTA) in coronary heart disease anatomy and function has increased significantly, which plays an important role in preoperative diagnosis, intraoperative monitoring, and postoperative recovery evaluation. Details of this study are available on page 2439.

A regular article titled “Effective Emergency Management: Scrutinizing the Malaysia Lead Responding Agency Planning and Information Management Approach During Disaster Exercise” was presented by Khairilmizal Samsudin and co-researchers from Malaysia. This study proposes several areas of improvement for planning and information management elements in effective emergency management theory by identifying its challenges through the Fire and Rescue Department Malaysia (FRDM). Four disaster

exercises at the state and federal levels were observed using a structured checklist. The result indicates challenges were found in terms of proactivity in gaining information, disseminating critical information during emergencies, developing a specific incident action plan, and media management. Detailed information about this study is available on page 2521.

In the last 12 months, of all the manuscripts peer-reviewed, 46% were accepted. This seems to be the trend in PJST.

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, Editor-in-Chief 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.

Chief Executive Editor

executive_editor.pertanika@upm.edu.my

Application of Groupwise Principal Sensitivity Components on Unbalanced Panel Data Regression Model for Gross Regional Domestic Product in Kalimantan

Desi Yuniarti^{1,2}, Dedi Rosadi^{1*} and Abdurakhman¹

¹Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Sekip Utara Bulaksumur 21 Yogyakarta 55281, Indonesia

²Department of Mathematics, Faculty of Mathematics and Natural Science, Mulawarman University, Barong Tongkok 04, Kampus Gunung Kelua, Samarinda, East Kalimantan 75123, Indonesia

ABSTRACT

Most robust estimation methods for panel data regression models do not consider the panel data structure consisting of several cross-sections and time-series units. This robust method, which does not consider the panel data structure, can completely remove all observations from a cross-section unit in trimming outlier observations. However, it can cause biased estimation results for the cross-section unit. This study determines the robust estimate for the unbalanced panel data regression model using Groupwise Principal Sensitivity Components (GPSC) by considering grouped structure data. The results were compared with Within-Group (WG) estimation and other robust estimation methods, namely Within-Group estimation with median centering (Median WG), Within-Group Least Trimmed Squares (WG-LTS), and Within Generalized M (WGM) estimators. Comparisons were made based on the Mean Squares Error (MSE) value. In this study, we applied the proposed method to the unemployed and the Gross Regional Domestic Product (GRDP) data at

constant prices in Kalimantan, Indonesia. The analysis showed that GPSC was the best method with the smallest MSE value. Therefore, we can consider implementing and developing the GPSC method to detect and determine the robust estimates for the unbalanced panel data regression model because it fits the panel data structure.

Keywords: Outliers detection, robust estimation, unbalanced panel data regression

ARTICLE INFO

Article history:

Received: 08 January 2022

Accepted: 18 March 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.01>

E-mail addresses:

desiyuniarti@mail.ugm.ac.id (Desi Yuniarti)

dedirosadi@gadjahmada.edu (Dedi Rosadi)

rahmanstat@ugm.ac.id (Abdurakhman)

* Corresponding author

INTRODUCTION

Panel data regression analysis is a regression analysis that uses panel data. Panel data results from observations of several cross-section units, namely companies, households, or individuals, over several periods. So panel data has dimensions of space and time because it consists of several cross-sections and time-series units. Balanced panel data has the same time-series units in each cross-section unit. In contrast, unbalanced panel data has a different number of time-series units in the cross-section units (Gujarati, 2004).

Bramati and Croux (2007) stated that the outliers in panel data might lead to a biased regression estimator. Therefore, it requires a robust panel data regression estimator against outliers. Research on robust estimators against outliers in panel data has begun to develop. Bramati and Croux's (2007) research is currently being developed and has become the basis for many studies on the robust estimators of outliers for panel data regression models. Bramati and Croux (2007) discussed robust estimators for panel data regression models with a fixed-effects approach, namely the Within Groups Generalized M (WGM) estimator by Wagenvoort and Waldmann (2002) and the Within Groups MS (WMS) estimator from Maroona and Yohai (2000) applied to panel data. Several research developments offer different data-centering methods and other estimation methods, such as Aquaro and Čížek (2013), Víšek (2015), Bakar and Midi (2015), and Midi and Muhammad (2018). Another robust estimation method is by Beyaztas and Bandyopadhyay (2020), studying the impact of outlier observations on the Ordinary Least Squares (OLS) estimation method in a linear panel data model and suggested a robust alternative estimation procedure based on weighted likelihood.

The studies described above applied and developed robust methods for panel data regression models, and each method has its advantages. However, we are interested in a robust estimation method considering a panel data structure consisting of several cross-sections and time-series units. The panel data structure consisting of several cross-section units allows for any differences in the average for each cross-section unit. For instance, the cross-section unit shows an area with poverty level data. Therefore, regions with the highest or lowest poverty rates can be considered outliers. The trimming process for outliers using the robust general method can remove observations from a cross-section unit. However, this trimming process can lead to biased estimation results for the appropriate cross-section unit. For this reason, it is necessary to make a robust estimate for the regression model that considers the structure of the panel data, especially the unbalanced panel data.

This study aims to apply the Groupwise Principal Sensitivity Components (GPSC) method by Perez et al. (2013) to detect and determine robust estimates for panel data regression models. The GPSC method is an outlier detection method to obtain robust estimates for grouped data and follows a linear regression model approach with fixed group effects. This GPSC method uses a sensitivity matrix formed by the sensitivity vector of

each cross-section unit. This method develops the Principal Sensitivity Components (PSC) method by Pena and Yohai (1999), adapted for grouped data. The GPSC method, which pays attention to the structure of the grouped data, is suitable for the regression model for panel data consisting of several cross-section units. This GPSC approach can identify the outliers within groups and ensure that outlier trimming does not remove more than 50% of data points from the same group. The outlyingness test of observations is based on a robust estimator from the previous analysis. The development of this GPSC method will be perfect for research on outlier detection and robust estimation for subsequent panel data regression models because this method considers the grouped data structures corresponding to panel data with several cross-section units. In the future, it is expected that research on robust estimators for panel data can be more concerned with the structure of the panel data itself.

Perez et al. (2013) studied outlier detection and robust estimation with data distributed into groups following a linear regression model with fixed group effects. They used several methods, including GPSC, the RDL1 method by Hubert and Rousseeuw (1997), and the MS method from Maroona and Yohai (2000). The results showed that GPSC could detect a high percentage of a true and small number of false outliers. This method was also capable of detecting any hidden high leverage points. In addition, this method could maintain good efficiency properties while maintaining good robustness properties. In their paper, Perez et al. (2013) also explained the deficiency of applying the robust method of M estimation, Generalized M (GM) estimation, least median of squares (LMS) method, Least Trimmed Squares (LTS) method, and Weighted Likelihood Estimator (WLE) method by Agostinelli and Markatou (1998) and Markatou, et al. (1998) on grouped data. Perez et al. (2013) stated that these methods are unsuitable for grouped data and proposed the GPSC method.

The GPSC method was developed based on the PSC method by Pena and Yohai (1999). PSC is a fast iterative procedure to estimate parameters based on a minimal robust scale. The procedure for minimizing this robust scale is obtained by eliminating possible outliers. In the study of Pena and Yohai (1999), each observation is represented by a sensitivity vector, a vector of changes in the least-squares estimate of the observations when each data point is removed. The set of possible outliers obtained as extreme points in the principal components of this vector is called the Principal Sensitivity Components (PSC), or as the set of points with large residuals. The good performance of the proposed procedure by Pena and Yohai (1999) allows the identification of outliers. Pena and Yohai (1999) explained two ways to see the outlyingness of the i th observation, i.e., by using an influence vector and a sensitivity vector. Pena and Yohai (1995) described an analysis based on the influence vector, and Pena and Ruiz-Castillo-Castillo (1998) used it to detect outlier groups in the regression model.

Perez et al. (2013) applied the GPSC method to income data as outcome variables, hectare, food crops, beef and lamb data as covariates variables, and state variables as grouping variables, which states the seven states where agriculture was. This application

of the GPSC method does not clearly state the type of data in each group. However, based on our study, the fixed group effects regression model used by Perez et al. (2013) is more general than panel data because, in each group, there can be data from several objects in the group. If the data in each group is in the form of time series data, it means that the data is panel data. For this reason, we emphasize the application of the GPSC method to panel data, especially for unbalanced panel data. The method can be used for panel data, especially for unbalanced panel data, because the analysis method considers the unbalanced panel data structure consisting of several cross-section units with different time-series units.

This study applies the GPSC method to the GRDP data of Kalimantan, Indonesia, by looking at the effect of unemployment on the GRDP of Kalimantan. The COVID-19 pandemic has dealt a severe blow to the world economy, including Indonesia, thus providing a different pattern for economic growth data. This condition then allows the occurrence of outliers that require a robust estimation method against outliers. We also compare the GPSC estimation results with Within-Group (WG) estimation and several other robust estimation methods, namely, WG estimation with median centering (Median WG), WG Least Trimmed Squares (WG-LTS), and Within Generalized M (WGM) estimator. Finally, we will use the Mean Squares Error (MSE) value to determine the best robust estimation method.

METHODOLOGY

Unbalanced Panel Data Regression Model

This study used an unbalanced panel data regression model in Equation 1 (Baltagi, 2005):

$$Y_{it} = \alpha + \mathbf{X}'_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T_i \quad [1]$$

where Y_{it} is the value of the Y variable for the t th time-series unit in the i th cross-section unit, α is a constant, $\mathbf{X}'_{it} = [X_{1it}, X_{2it}, \dots, X_{Kit}]$ is a vector of independent variables of size $K \times 1$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]'$ is a vector of a parameter of size $K \times 1$, N state the number of cross-section units, T_i state the number of time-series observations in the i th cross-section unit, and the total number of all observations is $n = \sum_{i=1}^N T_i$. Equation 1 is an unbalanced panel data regression model because it has a different number of time-series units for cross-section units.

The panel data regression model in Equation 1 is a one-way error component model if (Equation 2) (Baltagi, 2005):

$$u_{it} = \mu_i + v_{it} \quad [2]$$

where $v_{it} \sim IIN(0, \sigma_v^2)$ and \mathbf{X}_{it} is assumed to be independent of v_{it} . If μ_i is fixed, the model in Equation 1 with the error component in Equation 2 is a one-way panel data regression model with a fixed-effects approach.

Equations 1 and 2 give the following model of Equation 3:

$$Y_{it} = \alpha_i^* + \mathbf{X}'_{it}\boldsymbol{\beta} + v_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T_i \quad [3]$$

where $\alpha_i^* = \alpha + \mu_i$. Equation 3 is a panel data regression model specified for individual effects that are constant over time. For $i = 1, 2, \dots, N$ we can express Equation 3 in the following vector form of Equation 4 (Hsiao, 2003):

$$\mathbf{y}_i = \alpha_i^* \cdot \mathbf{1}_{T_i} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{v}_i \quad [4]$$

where $\mathbf{y}_i = [y_{i1} \ y_{i2} \ \dots \ y_{iT_i}]'$, $\mathbf{1}_{T_i}$ is a vector of elements one of size $T_i \times 1$, vector $\mathbf{X}_i = [\mathbf{X}_{i1} \ \mathbf{X}_{i2} \ \dots \ \mathbf{X}_{iT_i}]'$ of size $K \times T_i$, $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_K]'$ of size $K \times 1$, $\mathbf{v}_i = [v_{i1} \ v_{i2} \ \dots \ v_{iT_i}]'$, and assuming $E[\mathbf{v}_i] = \mathbf{0}$, $E[\mathbf{v}_i \mathbf{v}'_i] = \sigma_v^2 \mathbf{I}_{T_i}$, $E[\mathbf{v}_i \mathbf{v}'_j] = \mathbf{0}$ for $i \neq j$, where \mathbf{I}_{T_i} is the identity matrix of size $T_i \times T_i$.

We could then obtain the estimation of Equation 4 by the Ordinary Least Squares (OLS) method. The OLS estimator of α_i^* and $\boldsymbol{\beta}$ was obtained by minimizing $S = \sum_{i=1}^N \mathbf{v}'_i \mathbf{v}_i$. The within-group covariance matrix of variables X and the within-group covariance vector between the variables X and Y , respectively, are as in Equation 5 (Perez et al., 2013):

$$\begin{aligned} \mathbf{S}_{XX,i} &= \frac{1}{T_i} \sum_{t=1}^{T_i} (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)' \\ \mathbf{S}_{XY,i} &= \frac{1}{T_i} \sum_{t=1}^{T_i} (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) (y_{it} - \bar{y}_i) \end{aligned} \quad [5]$$

for $i = 1, 2, \dots, N$, where $\bar{\mathbf{X}}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \mathbf{X}'_{it}$ is the mean of the variable $X_k, k = 1, 2, \dots, K$ in the i th cross-section unit, and $\bar{y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it}$ is the average of the Y variables in the i th cross-section unit. The combined covariance matrix \mathbf{S}_{XX} and the combined covariance vector \mathbf{S}_{XY} are as in Equation 6 (Perez et al., 2013):

$$\mathbf{S}_{XX} = \sum_{i=1}^N \frac{T_i}{n} \mathbf{S}_{XX,i} \quad ; \quad \mathbf{S}_{XY} = \sum_{i=1}^N \frac{T_i}{n} \mathbf{S}_{XY,i} \quad (6)$$

The least squares (LS) estimator of $\boldsymbol{\beta}$ and α_i^* respectively are stated as Equations 7 and 8 (Perez et al., 2013):

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \mathbf{S}_{XX}^{-1} \mathbf{S}_{XY} \\ \hat{\boldsymbol{\beta}} &= \left[\sum_{i=1}^N \sum_{t=1}^{T_i} (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)' \right]^{-1} \left[\sum_{i=1}^N \sum_{t=1}^{T_i} (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) (y_{it} - \bar{y}_i) \right] \end{aligned} \quad (7)$$

and

$$\hat{\alpha}_i^* = \bar{y}_i - \bar{\mathbf{X}}_i' \hat{\boldsymbol{\beta}} \quad ; \quad i = 1, 2, \dots, N \quad (8)$$

Within-Groups Estimator of Unbalanced Panel Data Regression Model

Within-Groups (WG) estimator for the unbalanced panel data regression model uses the within transformation by forming a $\mathbf{Q} = \text{diag}(\mathbf{E}_{T_i})$ matrix where $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$, \mathbf{I}_{T_i} is an identity matrix of size $T_i \times T_i$, $\bar{\mathbf{J}}_{T_i} = \frac{1}{T_i} \mathbf{J}_{T_i}$, \mathbf{J}_{T_i} is a one-element matrix of size $T_i \times T_i$. The matrix elements of \mathbf{Q} are the deviations from the individual mean. This transformation is applied to the vector form of Equations 1 and 3, giving the following Equations 9 and 10:

$$\mathbf{y} = \alpha \boldsymbol{\iota}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad [9]$$

and

$$\mathbf{u} = \mathbf{Z}_\mu \boldsymbol{\mu} + \mathbf{v} \quad [10]$$

where \mathbf{y} is a $n \times 1$ vector, α is a scalar, $\boldsymbol{\iota}_n$ is a vector of ones of size $n \times 1$, \mathbf{X} is a matrix of independent variable of size $n \times K$, $\boldsymbol{\beta}$ is a parameter vector of size $K \times 1$, \mathbf{u} is an error vector of size $n \times 1$, and $\mathbf{Z}_\mu = \text{diag}(\boldsymbol{\iota}_{T_i})$ is a square diagonal matrix with the vector element $\boldsymbol{\iota}_{T_i}$ on the main diagonal.

Thus, we could get within transformation of Equation 9 to Equation 11 (Baltagi, 2005):

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{v}} \quad [11]$$

where $\tilde{\mathbf{y}} = \mathbf{Q}\mathbf{y}$, $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$, and $\tilde{\mathbf{v}} = \mathbf{Q}\mathbf{v}$. Applying the least-squares method to Equation 11, we could get the WG estimator as Equation 12 (Baltagi, 2005):

$$\tilde{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} \quad [12]$$

provided $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ exists.

Groupwise Principal Sensitivity Components for Unbalanced Panel Data Regression Model

This section explains the GPSC method corresponding to the unbalanced panel data regression model in Equation 1. In applying GPSC to the model of Equation 3, the cross-section unit in the panel data corresponds to the group in the model of GPSC method by Perez et al. (2013). Perez et al. (2013) developed the GPSC algorithm we applied in this study. The GPSC method goes through two stages: the first stage is to determine the clean set from outliers and then determine the initial robust estimate of the clean set. In stage 1, we formed the sensitivity matrix \mathbf{R}_i for i th cross-section unit. Given that $\hat{y}_{it(ij)}$ is the

estimated value of y_{it} if we remove the ij th observation (y_{ij}, \mathbf{X}'_{ij}) as Equation 13 (Perez et al., 2013):

$$\hat{y}_{it(ij)} = \hat{\alpha}_{i(ij)} + \mathbf{X}'_{it(ij)} \hat{\boldsymbol{\beta}}_{(ij)} \tag{13}$$

We determined the estimated value of $\hat{\boldsymbol{\beta}}_{(ij)}$ based on Equation 7 and $\hat{\alpha}_{i(ij)}$ based on Equation 8 using all observations except the ij th observation. Furthermore, for each y_{it} observation in the i th cross-section unit, the sensitivity vector is the change value vector if every point in the i th cross-section unit is deleted as Equation 14 (Perez et al., 2013):

$$\mathbf{r}_{it} = (\hat{y}_{it} - \hat{y}_{it(i1)}, \hat{y}_{it} - \hat{y}_{it(i2)}, \dots, \hat{y}_{it} - \hat{y}_{it(iT_i)})' \tag{14}$$

\mathbf{r}_{it} vector is the sensitivity vector of the i th cross-section unit. Then, we formed the sensitivity matrix \mathbf{R}_i of the i th cross-section unit as in Equation 15:

$$\mathbf{R}_i = [\mathbf{r}_{i1} \ \mathbf{r}_{i2} \ \dots \ \mathbf{r}_{iT_i}]' \tag{15}$$

To avoid any different modeling as much as T_i , we could determine the elements of the matrix \mathbf{R}_i based on the leverage and residuals of the following least square model in Equation 16 (Perez et al., 2013):

$$\hat{y}_{it} - \hat{y}_{it(ij)} = \frac{h_{tj}^i e_{ij}}{1 - h_{jj}^i} \tag{16}$$

where h_{jj}^i is the leverage effect of the j th observation of the i th cross-section unit as in Equation 17:

$$h_{jj}^i = \frac{1}{T_i} + (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)' (n\mathbf{S}_{XX})^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i) \tag{17}$$

Thus, the sensitivity matrix for the i th cross-section is presented in Equation 18 (Perez et al., 2013):

$$\mathbf{R}_i = \mathbf{H}_{ii} \mathbf{W}_i \tag{18}$$

where,

$$\mathbf{H}_{ii} = \begin{bmatrix} h_{11}^i & h_{12}^i & \dots & h_{1T_i}^i \\ h_{21}^i & h_{22}^i & \dots & h_{2T_i}^i \\ \vdots & \vdots & \ddots & \vdots \\ h_{T_i1}^i & h_{T_i2}^i & \dots & h_{T_iT_i}^i \end{bmatrix} \text{ and } \mathbf{W}_i = \begin{bmatrix} \frac{e_{i1}}{1-h_{11}^i} & 0 & \dots & 0 \\ 0 & \frac{e_{i2}}{1-h_{22}^i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{e_{iT_i}}{1-h_{T_iT_i}^i} \end{bmatrix}$$

Next, we form the matrix as in Equation 19 (Perez et al., 2013):

$$\mathbf{M}_i = \mathbf{R}'_i \mathbf{R}_i \quad [19]$$

and determined the non-zero eigenvalues of the \mathbf{M}_i matrix and the set of eigenvectors $\{\mathbf{v}_q^i, q = 1, 2, \dots, K + 1\}$ corresponding to the non-zero eigenvalues of the \mathbf{M}_i matrix. The maximum eigenvalue of \mathbf{M}_i expressed as λ_1^i is a measure of the global influence from the observation of the i th cross-section unit on the predicted values of observations in that cross-section unit. The eigenvector v_1^i corresponding to the eigenvalue λ_1^i is the direction of the maximum sensitivity of the observations at the i th cross-section unit. Eigenvector $\{\mathbf{v}_q^i, q = 1, 2, \dots, p + 1\}$ is the orthogonal direction in which the joint effect of deleting multiple data points from the i th cross-section in the estimated value is maximized. Therefore, the projection of Equation 20 is as follows:

$$\mathbf{z}_q^i = \mathbf{R}_i \mathbf{v}_q^i \quad [20]$$

in the direction of $\mathbf{v}_q^i, q = 1, 2, \dots, K + 1$ detects high leverage points with high mutual influence in the i th cross-section unit. This projection is the principal component of the sensitivity vector. According to Pena and Yohai (1999), the group of points that together have a leverage effect in the i th cross-section unit is expected to have extreme coordinates at least one of $p + 1$ PSC $\{\mathbf{z}_q^i, q = 1, 2, \dots, K + 1\}$. Furthermore, for each principal component of q , a different data set is formed, namely the first set containing all observations from each cross-section unit and the second set, deleting 50% of observations with the largest coordinates in the vector (Equation 21):

$$\mathbf{d}_q^i = |\mathbf{z}_q^i - \text{med}(\mathbf{z}_q^i)|; \quad q = 1, 2, \dots, K + 1 \quad [21]$$

The two sets for each of the i th cross-section units were combined. Furthermore, other small but potentially clean data sets were also formed with the smallest number of eigenvalues provisions. Then, the LS estimates for each of these sets were determined. Based on the results of this LS estimation, we chose the LS estimate that minimized the s -scale estimator. In this study, the robust scale estimate used was the median absolute deviation (MAD) as in Perez et al. (2013). Next, all observations were deleted as in Equation 22:

$$|r_{it}| \geq C_1 \cdot s_i \quad [22]$$

for $C_1 = 2$ and s_i is the Median Absolute Deviation (MAD) robust scale for the i th cross-section unit. Iterations were performed for all remaining observations. For example, $\mathbf{y}^{(r)} = \left((\boldsymbol{\beta}^{(r)})^T, \alpha_1^{(r)}, \alpha_2^{(r)}, \dots, \alpha_N^{(r)} \right)^T$ is the estimator obtained by minimizing

robust scale on the r th iteration. The iteration will end when $\mathbf{y}^{(r+1)} = \mathbf{y}^{(r)}$ and then $\mathbf{y}^* = \mathbf{y}^{(r+1)} = (\boldsymbol{\beta}^{*T}, \alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ is the initial robust estimator.

Based on the results of stage 1 analysis, we obtained a data set that may be clean because we removed observations that could be outliers. Furthermore, in stage 2, we tested these potential outliers using a robust t -test. Finally, we returned the observations not rejected by this robust t -test to the sample data and used them to determine the final estimator. The steps in stage 2 include determining the residuals from the initial robust estimator and eliminating observations by using Equation 23:

$$|r_{it}| > C_2 \cdot s_i \ ; \ i = 1, 2, \dots, N \tag{23}$$

for $C_2 = 3$. Let n^* be the total number of deleted observations. Then, the LS estimator for the remaining $n - n^*$ observations is calculated and expressed as $\tilde{\alpha}_i, i = 1, 2, \dots, N$ and $\tilde{\boldsymbol{\beta}}$. Also, the standard error $\tilde{\sigma}$ using the residuals of these remaining observations and the corresponding leverage \tilde{h}_{it}^i based on Equation 17 was calculated. The outlyingness test of each n^* observation used the following robust t -test statistics in Equation 24:

$$t_{it} = \frac{y_{it} - \tilde{\alpha}_i - \mathbf{X}'_{it}\tilde{\boldsymbol{\beta}}}{\tilde{\sigma}\sqrt{1 + \tilde{h}_{it}^i}} \tag{24}$$

We eliminate every n observation with $|t_{it}| > C_3$, where $C_3 = 3$ (Perez et al., 2013).

Other Robust Estimators for Unbalanced Panel Data Regression Model

In this study, the median WG, WG-LTS, and WGM estimator were other analytical methods used to determine robust estimators. Initially, we determined the centering of variables (dependent and independent) on the median (med) as follows (Equation 25) (Bramati & Croux, 2007):

$$\tilde{Y}_{it} = Y_{it} - \text{med}_t Y_{it} \text{ and } \tilde{X}_{Kit} = X_{Kit} - \text{med}_t X_{Kit} \tag{25}$$

for $1 \leq i \leq N, 1 \leq t \leq T_i, n = \sum_{i=1}^N T_i$ and $1 \leq k \leq K, X_{Kit}$ is the k th independent

variable, $k = 1, 2, \dots, K$ measured at the t th time-series unit in the i th cross-section unit.

We then obtained the median WG estimator by doing the median centering based on Equation 25 first and determined the WG estimator using the variables \tilde{Y}_{it} and \tilde{X}_{Kit} based on Equation 12. Aquaro and Čížek (2013) have also used this robust method before.

WGM estimator is one of Bramati and Croux's (2007) methods, and the WG-LTS estimator is the initial estimator for the WGM estimator. This study applied the WGM

method to determine a robust estimator for the unbalanced panel data regression model in Equation 1. After doing median centering of variables, Bramati and Croux (2007) regressed \tilde{Y}_{it} against \tilde{X}_{Kit} using a robust regression method, namely the LTS method that minimizes the sum of h , the smallest residual squared as in Equation 26:

$$\tilde{\beta}_{LTS} = \arg \min_{\beta} \sum_{j=1}^h [(\tilde{Y}_j - \tilde{\mathbf{X}}'_j \beta)^2]_{j:n} \quad [26]$$

for $n = \sum_{i=1}^N T_i$, $h = 3/4 NT$ is the truncation value. For the estimator β with median centering, we got Equation 27 (Bramati & Croux, 2007):

$$\tilde{\alpha}_i(\beta) = \text{med}_t(Y_{it} - \mathbf{X}'_{it} \beta), \quad i = 1, 2, \dots, N \quad [27]$$

The WGM estimator extends the LTS within-group estimator to improve statistical efficiency while maintaining robustness. To determine the WGM estimator, we formed a diagonal matrix of \mathbf{W}_r of size $n \times n$ to reduce observations' weight with a large residual value from a robust initial LTS fit regression model. The loss function used Tukey's biweight function so that the diagonal element \mathbf{W}_r became Equation 28 (Bramati & Croux, 2007):

$$(W_r)_{it} = \begin{cases} 0, & \text{if } \left| \frac{r_{it}}{\hat{\sigma}_{LTS}} \right| \geq c \\ \left(1 - \left(\frac{r_{it}}{c \hat{\sigma}_{LTS}} \right)^2 \right)^2, & \text{if } \left| \frac{r_{it}}{\hat{\sigma}_{LTS}} \right| < c \end{cases} \quad [28]$$

where $r_{it} = \tilde{Y}_{it} - \tilde{\mathbf{X}}'_{it} \hat{\beta}_{LTS}$ is the residual of the WG-LTS model, $\hat{\sigma}_{LTS}$ is the robust scale estimate of residual, $\hat{\sigma}_{LTS} = c \frac{1}{h} \sum_{j=1}^h [(\tilde{Y}_j - \tilde{\mathbf{X}}'_j \beta)^2]_{j:n}$, and $c = 4.685$ according to Wagenvoort and Waldmann (2002). Furthermore, we formed a \mathbf{W}_x matrix of size $n \times n$. The diagonal elements of the \mathbf{W}_x matrix are presented in Equation 29 (Bramati & Croux, 2007):

$$(W_x)_{it} = \min \left(1, \frac{\sqrt{\chi_{K,0.975}^2}}{RMD_{it}} \right) \quad [29]$$

where $\chi_{K,0.975}^2$ is the upper 97.5% quantile of a Chi-Squared distribution with K degrees of freedom. Robust distance RMD_{it} is a robust of Mahalanobis distance computed for every \tilde{X}_{it} as in Equation 30:

$$RMD_{it} = \sqrt{(\tilde{X}_{it} - \hat{\mu})' \hat{V}^{-1} (\tilde{X}_{it} - \hat{\mu})}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T_i \quad [30]$$

where $\hat{\mu}$ and \hat{V} are the robust location estimates and covariate estimates of the centered independent variables, calculated by applying the S-multivariate location and scale estimator, respectively.

Thus, we could determine WGM estimator for the one-way panel data regression model with a fixed-effects approach as follows (Equation 31) (Bramati & Croux, 2007):

$$\hat{\beta}_{WGM} = (\tilde{X}' \mathbf{W}_x \mathbf{W}_r \tilde{X})^{-1} \tilde{X}' \mathbf{W}_x \mathbf{W}_r \tilde{y} \quad [31]$$

Research Data

The COVID-19 pandemic has hit the economy of the world, including Indonesia. Various problems then occur because of the COVID-19 pandemic, such as business closures and staff reduction, leading to an increase in the unemployment rate and a decrease in people's purchasing power. In the end, it also ultimately affects Indonesia's economic growth, which has decreased. The government is attempting to reduce the impact of COVID-19 on the economy through the National Economic Recovery (NER) program. This program aims to protect, maintain, and improve the economic capacity of business actors in running their businesses during the COVID-19 pandemic.

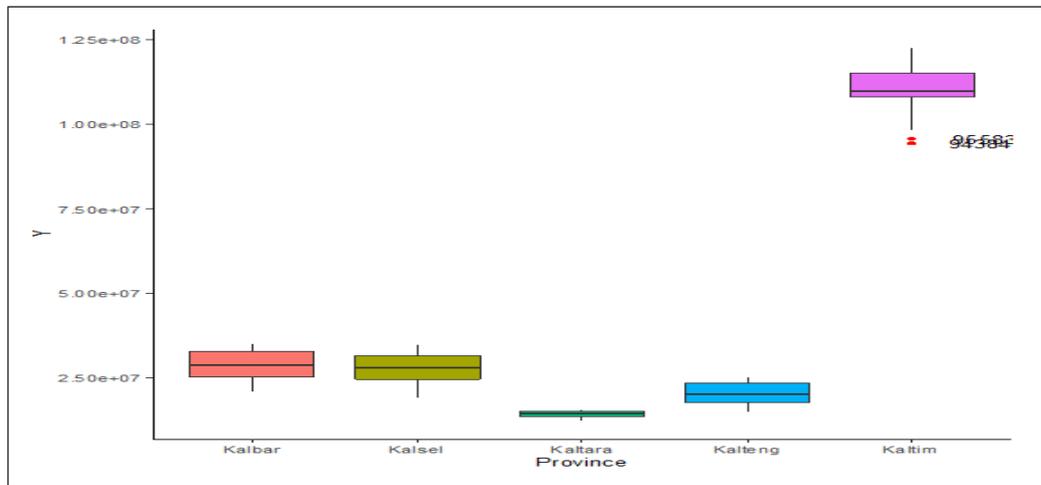
The government's economic restoration efforts are faced with challenges such as the lack of people's purchasing power and a reduction in employees in various businesses. Okun (1962) described the relationship between economic growth (output) and unemployment (input), also known as Okun's Law. Okun's law explains a negative relationship between economic growth and unemployment: when unemployment increase, the economic growth decrease, and vice versa.

This study investigated the effect of unemployment on economic growth using data on the number of unemployed people and the Gross Regional Domestic Product (GRDP) at constant prices (millions of rupiah), showing the economic growth. It used panel data with a cross-section unit covering five provinces on Kalimantan Island, Indonesia, including West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. In addition, the study used quarterly data from 2010 to 2021 as a time-series unit from the Badan Pusat Statistik or Central Bureau of Statistics (BPS) website for each region.

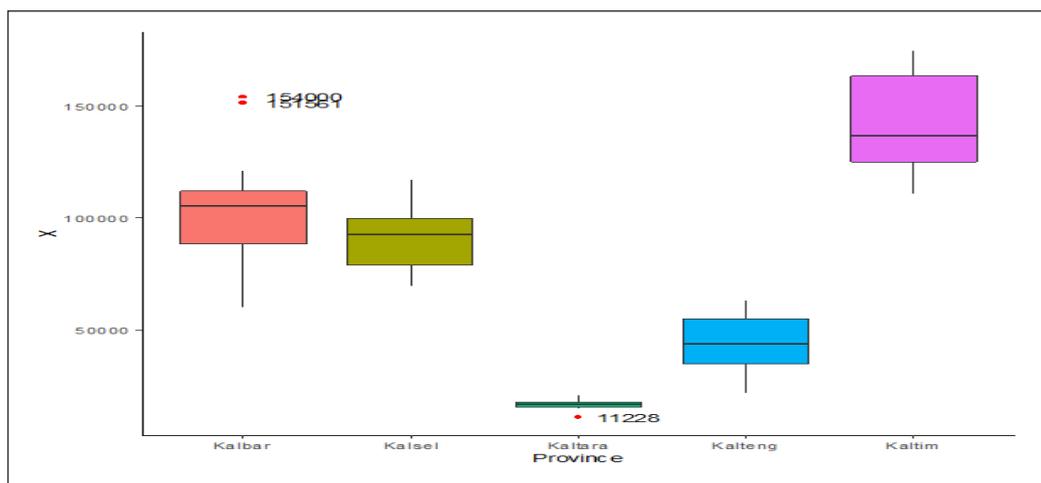
We used some research variables and cross-section units, as shown in Tables 1 and 2. The complete time-series unit referred to quarterly data from 2010 to 2021. Because there were incomplete quarterly data for each variable and unavailable data due to the formation of a new province, causing a different number of time series for each cross-section unit, the data in this study are unbalanced panel data.

Table 3 summarizes data from the variable for each cross-section unit. The average GRDP of East Kalimantan and West Kalimantan provinces were the two highest regions, while the lowest was North Kalimantan, the youngest province on the Kalimantan Island. The highest average unemployment data was in East Kalimantan province, followed by West Kalimantan. However, the average unemployment rate in West and East Kalimantan was not much different, though GRDP was significant. As shown in Table 3, we suspected an outlier in variable X for the province of West Kalimantan and an outlier in variable Y for the province of East Kalimantan. For this reason, we investigated outlier observations through each cross-section unit's boxplots of each research variable.

Figure 1 shows the boxplot of variables Y and X . Based on Figures 1(a) and 1(b), the observations outside the boxplot indicate a presence of outlier observations. Figure 1(a)



(a)



(b)

Figure 1. (a) Boxplot of variable Y; (b) Boxplot of variable X

shows two outliers on variable *Y* in the East Kalimantan Province, 94,384,716 in quarter 1 of 2010 and 95,583,067 in quarter 3 of 2010. Variable *X* in Figure 1(b) has outlier observations in West Kalimantan Province, 151,562 in quarter 3 of 2020 and 154,000 in quarter 1 of 2021. Also, North Kalimantan province has outlier observations of 11,228 in quarter 1 of 2016.

Table 3 and Figure 1 indicate the presence of outlier observations, necessitating an analytical method to produce their robust parameter estimates. The analysis for robust parameter estimation against outliers is discussed in the next section.

RESULTS AND DISCUSSION

This section determined the robust estimation of the panel regression model for GRDP data in Kalimantan, Indonesia, using the GPSC method. This method considers the structure of grouped data to provide better estimation results for panel data consisting of cross-section and time-series units. The results obtained were compared with WG, median WG, WG-LTS, and WGM estimators based on the smallest MSE value. All analyses were conducted using software R. Moreover, the WG estimator was determined using the PLM package, while median WG, WG-LTS, and WGM were performed using R software following the analysis steps described by Bramati and Croux (2007). Finally, the GPSC analysis was conducted using R software based on the syntax given by Perez et al. (2013).

Based on Equation 1 to Equation 3, the one-way unbalanced panel data regression model with a fixed-effects approach for GRDP of Kalimantan data is given as in Equation 32:

$$Y_{it} = \alpha_i^* + \beta X_{it} + v_{it}, \quad i = 1, 2, \dots, 5 \tag{32}$$

where $\alpha_i^* = \alpha + \mu_i$. The *i*-index shows the *i*th cross-section unit according to Table 2. The *t*-index is the index for the *t*th time-series unit, which shows quarterly data from 2010 to 2021. The data in this study possessed many different time-series units for each cross-section unit. Variables *Y* and *X* are following Table 1. The estimation model of Equation 33 is:

$$Y_{it} = \hat{\alpha}_i^* + \hat{\beta} X_{it}, \quad \hat{\alpha}_i^* = \hat{\alpha} + \hat{\mu}_i, \quad i = 1, 2, \dots, 5 \tag{33}$$

Outliers in the GRDP and the unemployment data required a robust estimation model in Equation 33 against outliers.

Table 1
Research variable

Variable	Description	Measure
<i>Y</i>	Gross Regional Domestic Bruto (GRDB) at 2010 Constant Market Prices	Million Rupiahs
<i>X</i>	Number of Unemployment	People

Table 2
Cross-section units

<i>i</i> -index	Province	Abbreviation
1	West Kalimantan	Kalbar
2	Central Kalimantan	Kalteng
3	South Kalimantan	Kalsel
4	East Kalimantan	Kaltim
5	North Kalimantan	Kaltara

Table 3
Data summary of each research variable

Variable	Province	Min	Q_1	Mean	Q_3	Max	NA
Y	Kalbar	2,0760,144	24,714,642	28,357,183	32,700,599	34,995,845	0
	Kalteng	14,942,800	17,896,000	20,575,264	23,970,710	25,414,100	2
	Kalsel	19,181,665	24,593,119	27,717,131	31,463,387	34,989,964	0
	Kaltim	94,384,716	107,887,123	110,379,865	115,089,682	122,535,328	0
	Kaltara	12,360,709	13,504,097	14,271,830	15,237,585	15,584,110	12
X	Kalbar	59,884	88,397	103,097	112,081	154,000	1
	Kalteng	21,838	34,994	44,894	54,995	63,309	3
	Kalsel	69,537	79,227	90,845	99,816	117,209	0
	Kaltim	110,574	125,024	141,546	163,517	174,807	0
	Kaltara	11,228	16,079	16,735	17,290	20,867	10

Estimation Robust of Gross Regional Domestic Product of Kalimantan using Groupwise Principal Sensitivity Components

This section introduces the use of the GPSC method on unbalanced panel data. The method has been proposed to detect and determine robust estimates for linear regression models with fixed group effects corresponding to panel data with several cross-section units.

The first stage of the GPSC method determines the sensitivity matrix for each i th cross-section unit based on Equation 18 and the \mathbf{M}_i matrix based on Equation 19. The \mathbf{M}_i matrix is $T_i \times T_i$ for $i = 1, 2, \dots, 5$, where T_i is the number of time-series units in the i th cross-section unit. Table 4 shows the number of T_i for the i th cross-section unit. The application of GPSC to unbalanced panel data began by eliminating incomplete observations, so the number of observations used in this study was 99.

Tables 5 and 6 show the analysis results from stage 1 using the GPSC method. Table 5 shows the parameter estimation results in stage 1 with an estimated robust s -scale of 3.082×10^6 . Table 6 shows the observations suspected of being outliers at stage 1. Based on Table 6, we could get six observations suspected of being outliers, namely the first and second observations from the first cross-section unit (West Kalimantan), the 43rd observation from the third cross-section unit (South Kalimantan), and the 66th, 67th, and 68th observations from the fourth cross-section unit (East Kalimantan).

Table 4
The number of time-series units of each cross-section unit

<i>i</i> -index	Number of time-series units (T_i)
1	22
2	20
3	23
4	23
5	11
Total (n)	99

Table 5
Parameter estimation of stage 1

Parameter	Parameter Estimation
α_1	3.392×10^7
α_2	2.044×10^7
α_3	2.988×10^7
α_4	1.172×10^8
α_5	1.460×10^7
β	-19.741
MAD	3,081,709

Table 6
Outliers observation of stage 1

<i>i</i> th Observation	Cross-section Unit	Number of Outliers
1, 2	1	2
43	3	1
66, 67, 68	4	3

Table 7
Outliers observation of stage 2

<i>i</i> th Observation	Cross-section Unit	t_{it}	Decision
66	4	-3.749	Outlier
67	4	-3.471	Outlier

Furthermore, in stage 2 of the GPSC method, we tested the outlyingness of the observations as in Table 6. Based on the initial robust estimator, we determined the robust s-scale estimation of each *i*th cross-section unit and deleted the observations according to Equation 23. Based on the analysis results for each cross-section unit, we found that the 66th and 67th observations in the fourth cross-section unit were the potential outliers. Next, the LS estimator for the remaining observations was determined, and an outlyingness test using the robust test statistic based on Equation 24 was performed. Table 7 shows the results of the robust t-test, and we can conclude that the 66th and 67th observations were the outliers. The final step was to determine the final robust estimate in the second stage based on the LS estimation from the remaining observations without the 66th and 67th observations. The final robust estimate for GRDP data using the GPSC method is shown in Table 8.

The Comparison of Robust Estimates for Gross Regional Domestic Product of Kalimantan

We compared the robust results obtained using GPSC with the WG estimation method and several other robust methods for panel data regression, namely Median WG, WG-LTS, and WGM estimators. The comparison was based on the MSE value, as shown in Table 8.

Table 8
Comparison of robust estimates for GRDP in Kalimantan

Parameter	Parameter Estimation				
	WG	Median WG	WG-LTS	WGM	GPSC
α_1	3.156×10^7	3.157×10^7	2.462×10^7	2.856×10^7	2.889×10^7
α_2	2.158×10^7	2.149×10^7	1.834×10^7	2.013×10^7	2.042×10^7
α_3	3.023×10^7	2.989×10^7	2.547×10^7	2.797×10^7	2.789×10^7
α_4	1.143×10^7	1.140×10^8	1.049×10^8	1.096×10^8	1.121×10^8
α_5	1.474×10^7	1.488×10^7	1.386×10^7	1.444×10^7	1.430×10^7
β	-27.701	-24.217	35.710	1.756	-1.899
MSE	2.530×10^{13}	2.416×10^{13}	3.342×10^{13}	2.436×10^{13}	2.038×10^{13}

In Table 8, the GPSC method gave the lowest MSE value, meaning the GPSC method was the best estimate of Kalimantan's GRDP data. These results were expected because the GPSC method considers a data structure that matches the panel data structure. To generate a robust estimate for outlyingness test statistics, eliminating observations suspected of being outliers consists of several stages. First, this method ensures that the deletion does not exceed 50% of the observations in each cross-section unit, thereby preventing the elimination of the observations in a single cross-section unit and providing an appropriate estimation result for each cross-section unit.

Based on Equation 33 and Table 8, the estimation model of the GPSC method is written as Equation 34:

$$GRDP_{it} = \hat{\alpha}_i^* - 1.899UE_{it} \quad [34]$$

$GRDP_{it}$ is the Gross Regional Domestic Product of Kalimantan for the t th time-series unit in the i th cross-section unit, UE_{it} is the number of unemployed for the t th time-series unit in the i th cross-section unit. The value of $\hat{\alpha}_i^*$ in Table 9 represents the estimated intercept for each i th cross-section unit. The cross-section units can be seen in Table 2 and the time-series units for the quarterly period.

The model in Equation 34 means that every additional unemployed person will reduce Kalimantan's GRDP by Rp. 1,890,000.00. This model shows a negative relationship between unemployment and GRDP, appropriate with Okun's Law. Therefore, the government should create more job opportunities to fulfill their lives. Income increases people's purchasing power and the economic growth of Kalimantan in particular and Indonesia in general.

Table 9
Intercept estimation of each cross-section unit of the GPSC model

Province	Intercept Estimation
West Kalimantan	2.889×10^7
Central Kalimantan	$2,042 \times 10^7$
South Kalimantan	2.789×10^7
East Kalimantan	1.121×10^8
North Kalimantan	1.430×10^7

Perez et al. (2013) also concluded that the GPSC method is better if the group means differ significantly. This conclusion is consistent with Kalimantan's GRDP data, where East Kalimantan had a much higher average GRDP and unemployment rate than other regions. Thus, the GPSC method is suitable for determining robust estimates for Kalimantan GRDP data. Perez et al. (2013) compared the RDL1, M-S, and GPSC methods. The GPSC and MS methods by Moronna and Yohai (2000) gave almost similar estimation results to other methods, indicating their credibility. The comparison results in Table 8 showed that the estimated intercept parameters were not much different for the WGM and GPSC methods, but the slope differed in sign. The slope sign for the WGM estimation results did not follow Okun's Law, and the MSE value was still higher.

CONCLUSION

This study applies the GPSC method in detecting outliers and determining robust estimates for unbalanced panel data regression models. We intend to emphasize that GPSC can be applied to panel data regression models, especially unbalanced panel data, because this method considers grouped data structures. So that this method is suitable for an unbalanced panel data structure consisting of several cross-section units with a different number of time-series units. We use unbalanced panel data from data on unemployment and the GRDP at constant prices in Kalimantan, Indonesia. We compare the robust estimation results using GPSC with the WG, Median WG, WG-LTS, and WGM estimation methods. Based on the analysis results, we conclude that the GPSC estimation method provides the best robust estimation results for data of GRDP in Kalimantan.

Based on the results and discussion, we suggest developing a method that considers the panel data structure in detecting outliers and determining robust estimation for the panel data regression model, particularly for the unbalanced data panel. Therefore, we can consider implementing and developing the GPSC method because this method is very suitable for an unbalanced panel data structure consisting of several cross-section units with different time-series units.

ACKNOWLEDGEMENTS

The authors thank the Lembaga Pengelola Dana Pendidikan (LPDP) for its financial support in this doctoral study.

REFERENCES

- Agostinelli, C., & Markatou M. (1998). A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Statistics & Probability Letters*, 37(4), 341-350. [https://doi.org/10.1016/S0167-7152\(97\)00136-3](https://doi.org/10.1016/S0167-7152(97)00136-3)

- Aquaro, M., & Čížek, P. (2013). One-step robust estimation of fixed-effects panel data models. *Computational Statistics and Data Analysis*, 57, 536-548. <https://doi.org/10.1016/j.csda.2012.07.003>
- Bakar, N. M. A., & Midi, H. (2015). Robust centering in the fixed effect panel data model. *Pakistan Journal of Statistics*, 31(1), 33-48.
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd Ed.). John Wiley & Sons Inc.
- Beyaztas, B. H., & Bandyopadhyay, S. (2020). Robust estimation for linear panel data models. *Statistics in Medicine*, 39(29), 4421-4438. <https://doi.org/10.1002/sim.8732>
- Bramati, M. C., & Croux C. (2007). Robust estimators for the fixed effects panel data model. *Econometrics Journal*, 10, 521-540. <https://doi.org/10.1111/j.1368-423X.2007.00220.x>
- Gujarati, D. (2004). *Basic Econometrics* (4th Ed.). McGraw-Hill Companies, Inc.
- Hsiao, C. (2003). *Analysis of Panel Data* (2nd Ed.). Cambridge University Press.
- Hubert, M., & Rousseeuw, P. J. (1997). Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference*, 57(1), 153-163. [https://doi.org/10.1016/S0378-3758\(96\)00041-9](https://doi.org/10.1016/S0378-3758(96)00041-9)
- Markatou, M., Basu, A., & Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, 93(442), 740-750. <https://doi.org/10.2307/2670124>
- Maroona, R. A., & Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference*, 89(1-2), 197-214. [https://doi.org/10.1016/S0378-3758\(99\)00208-6](https://doi.org/10.1016/S0378-3758(99)00208-6)
- Midi, H., & Muhammad, S. (2018). Robust estimation for fixed and random effects panel data models with different centering methods. *Journal of Engineering and Applied Sciences*, 13(17), 7156-7161.
- Okun, A. M. (1962). *Potential GNP, its measurement and significance*. <https://milesorak.files.wordpress.com/2016/01/okun-potential-gnp-its-measurement-and-significance-p0190.pdf>
- Pena, D., & Ruiz-Castillo, J. (1998). The estimation of food expenditures from household budget data in the presence of bulk purchases. *Journal of Business & Economic Statistics*, 16(3), 292-303. <http://dx.doi.org/10.1080/07350015.1998.10524768>
- Pena, D., & Yohai, V. J. (1995). The detection of influence subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society*, 57(1), 145-156.
- Pena, D., & Yohai, V. J. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association*, 94(446), 434-445. <https://doi.org/10.2307/2670164>
- Perez, B., Molina, I., & Pena, D. (2013). Outlier detection and robust estimation in linear regression models with fixed group effects. *Journal of Statistical Computation and Simulation*, 84(12), 2652-2669. <https://doi.org/10.1080/00949655.2013.811669>
- Víšek, J. Á. (2015). Estimating the model with fixed and random effects by a robust method. *Methodology and Computing in Applied Probability*, 17, 999-1014. <https://doi.org/10.1007/s11009-014-9432-5>
- Wagenvoort, R., & Waldmann, R. (2002). On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics*, 106(2), 297-324. [https://doi.org/10.1016/S0304-4076\(01\)00102-6](https://doi.org/10.1016/S0304-4076(01)00102-6)

Evaluation of Subcritical Organic Rankine Cycle by Pure and Zeotropic of Binary and Ternary Refrigerants

Omid Rowshanaie¹, Mohd Zahirasri Mohd Tohir^{1*}, Faizal Mustapha²,
Mohammad Effendy Ya'acob³ and Hooman Rowshanaie⁴

¹Department of Chemical and Environmental Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Department of Aerospace Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

³Department of Process and Food Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

⁴Department of Crop Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

The simulation configuration and process analysis of the Subcritical Organic Rankine Cycle (SORC) system are carried out for the potential comparison between pure, binary, and ternary zeotropic mixtures of R1234ze(E), R1234yf, and R134a as refrigerant working fluids based on applying the flue gas as a heat source with medium temperature. The compression pressure was selected as an optimized variable input parameter of SORC with the lower limit of boundary condition (1.4 MPa); to mitigate air ingress and sub-atmospheric pressure that led to approach optimum net power output generated. Increasing the compression pressure has a positive relationship with the superheated temperature and the mass enthalpy change in the evaporation and, therefore, in the expansion process. In parallel, the enthalpy and entropy changes in the flue gas and cold water positively correlate

with exergy efficiency. So, R1234ze(E)/R1234yf/R134a with 68.35% and R1234yf/R134a with 69.29% as the lowest and highest exergy efficiency in the highest compression pressure; furthermore, the SIC consequences of increasing the cost of each component of the SORC system that has a direct relationship with the PPC and the required exchanger area of evaporation and condensation process and generating a net power output of the turbine. As a result, the maximum to the minimum value of specific

ARTICLE INFO

Article history:

Received: 13 January 2022

Accepted: 05 April 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.02>

E-mail addresses:

omid.rowshanaie@gmail.com (Omid Rowshanaie)

zahirasri@upm.edu.my (Mohd Zahirasri Mohd Tohir)

faizalms@upm.edu.my (Faizal Mustapha)

m_effendy@upm.edu.my (Mohammad Effendy Ya'acob)

hooman.rowshanaie@gmail.com (Hooman Rowshanaie)

* Corresponding author

investment cost (SIC) achieves R134a with 5807402.18-22455670.61 \$.kW⁻¹ and R1234yf with 16.82-17.38% reduction, respectively. To sum up, the lowest payback period (PBP) was R1234yf with 302 days.

Keywords: Exergy efficiency, flue gas, Organic Rankine Cycle (ORC), payback period (PBP), specific investment cost (SIC)

INTRODUCTION

Power and the environment are the two major concerns of this era. The aggravated usage of non-renewable energy sources, especially fossil fuels, has a destructive effect on the environment. It can create climate change and air pollution, cause acid rain, ozone depletion, carbon footprint, greenhouse gases emission effect, and global warming (F. Wang et al., 2021; Z. Wang et al., 2021; Lu et al., 2021; K. Sun et al., 2021; Q. Sun et al., 2021; Ping et al., 2021). However, applying low to medium temperature of waste heat source temperature (<350 °C) reveals that more than 50% of industrial waste heat by driving several waste heat recovery applications can reduce these harmful environmental obstacles and, on the other hand, able to convert these waste heat source into power (Alvi et al., 2021a; Yu et al., 2021; Zhang et al., 2020; Ahmadi et al., 2020; Li et al., 2017; Li et al., 2018; Vélez et al., 2012; Peris et al., 2015; Imran et al., 2018).

Nowadays, the Organic Rankine Cycle (ORC) applications among the various waste heat recovery applications as a most promising and prospective technology are selected to convert low to medium temperature of heat sources, including geothermal energy, biomass energy, solar thermal energy, and industrial waste heat especially flue gas into power (Schilling et al., 2021; Kavathia et al., 2021; Loni et al., 2021; Alvi et al., 2021b; Eyerer et al., 2020; Vaupel et al., 2021; Hamid et al., 2021).

In general, some scholars attempt to investigate and utilize the new, green, and environmentally friendly refrigerant organic working fluids in ORC based on the very low GWP (GWP<1), ODP with zero amount, and more suitable safety grading, low boiling point, and low atmospheric lifetime. Roumpedakis et al. (2020) studied the small-scale ORC system by using a solar as a low-temperature heat source and analysis and fulfillment by applying different typical organic refrigerant working fluids such as R134a, R245fa, R152a, R237ea, R236ea beside a new and environmentally friendly refrigerant working fluid like R1234ze(E) and investigating on the exergy efficiency, thermal efficiency, and thermo-economic parameters. In conclusion, apart from R245fa with the highest exergy and thermal efficiency, the short payback period belonged to a new and environmentally friendly working fluid, R1234ze(E). Molés et al. (2017) considered the performance of an ORC system driven by R1234yf and R1234ze as two working fluids with low GWP and alternative to an old working fluid, R134a, based on a varied range of evaporating

temperatures and condensing temperatures. The results depict that R1234yf would consume higher pump power as a power input compared with R134a by 18.3% to 25.8%.

Furthermore, in various evaporating and condensing temperatures, R1234ze has the highest trend of net cycle efficiency among R1234yf and R134a. Ata et al. (2020) optimized and analyzed an ORC system driven by R1234ze as a new-generation fluid using a heat source with 120°C. The current study performed six performances: thermal efficiency, turbine power, exergy efficiency, total irreversibility, Volume Flow Ratio (VFR), and Environmental Effect Factor (EEF) using the orthogonal design with Taguchi-ANOVA. The main control factors were $\Delta T_{pp,c}$ - $\Delta T_{pp,c}$ - $T_{c,i}$ - T_{sup} - η_t - η_p selected for the statistical analysis. EES's numerical analysis showed that the implemented sensitivity level ranking maximizes thermal efficiency as isentropic efficiency of the turbine, η_t (33.64%). Nevertheless, the scale of sensitivity levels to optimize turbine performance was analyzed as the inlet temperature of cooling water as a heat sink, $T_{c,i}$ (65.21%). Li et al. (2017) focused on the new-kind environmentally friendly, green, and safe organic refrigerant working fluid, R1234ze(E), as a considerable potential to be applied in subcritical and transcritical ORC applications used by the hot water as a heat source with 100–200 °C inlet temperature and no restriction of outlet temperature limit. Therefore, two of the optimized parameters in the expansion process were the inlet pressure and temperature of the turbine. In a nutshell, in this study, the highest system net power output of R1234ze(E) in comparison with R600a and also R245fa at 100–167 °C heat sources with no restriction of outlet temperature was maximum with 31.4% and 25.8% larger than that of R245fa and R600a, respectively.

In addition to paying attention to the new and environmentally friendly refrigerant organic working fluids in ORC, the main working conditions, especially subcritical besides transcritical, are carried out by particular groups of researchers, which are in parallel with the current study. Zhang et al. (2018) selected a subcritical as a working condition of air-cooled ORC using the 150 °C as a low-temperature geothermal brine. This research focused on applying R245fa as the typical working fluid. Also, R1234ze(E) and R1234ze(Z) were two low ODP and GWP working fluids to reach the maximum exergy efficiency main achievement of the system. These scholars exposed that, for 100 kg/s geothermal sources, the highest and maximum exergy efficiency of the system besides the highest total efficiency of the system between these typical and low GWP and zero ODP organic refrigerant working fluids belonged to an environmentally friendly and new working fluid, R1234ze(E) in the chemical process industry (CPI). Another research in the case of subcritical working conditions is Yang et al. (2015). They considered the choice of the most suitable working fluid for an ORC in terms of subcritical working conditions and made a comparison between several typical and new environmentally friendly working fluids, including R600, R600a, R601a, R245fa beside R1234ze and R1234yf and also utilizing a heat source such as the diesel engine with 200–370 °C as an exhaust waste heat recovery application. They achieved

that after analyzing and investigating all of the working fluids that drive this subcritical ORC (SORC), R1234ze(E) was chosen as a significant thermodynamic performance. Chagnon-Lessard et al. (2020) performed different numerical tools of a geothermal ORC, which was driven by 20 other potential typical and new environmental friendly working fluids; and (ORC/S/SC) as the subcritical ORC using the single-pressure heater, (ORC/S/TC) as the transcritical ORC using the single-pressure heater, (ORC/D/SC) as the subcritical ORC using the dual-pressure heater. Last but not least was (ORC/D/TC) as the transcritical ORC using the dual-pressure heater. Furthermore, the specific work output was chosen as an objective function of this SORC system, and some variables, such as work pressures, the brine, the working fluid mass flow ratios, effectiveness of superheaters, and the range of the cooling tower, were chosen in this study. Their main achievement was among the 20 working fluids of this research, R1234yf, R115, R125, R218, R227ea, R134a, R22, R124, R32, RC318, R134a, R12, RE245cb2, and R152a selected as the highest specific work. In addition, Hu et al. (2020) evaluated an ORC that utilized R1234ze(E), R1234ze(Z), R1234yf, R134a, R1243zf, R600a, R245fa, R1234yd(Z), R1233zd(E), and R1336mzz(Z) as typical, famous, and low GWP organic refrigerant working fluids as well. This study conducted a numerical target preference and multi-criteria decision-making (MCDM) method. Overall, they found the R1234ze(E) was optimal and had the most considerable power output of 50.8 kW, which was 14% higher than R245fa. Zhai et al. (2016) analyzed the 30 different working fluids, including a new environmentally friendly it, R1234yf, by applying subcritical ORC for open-type heat sources with temperatures from 150 to 350 °C, water to air, respectively as heat sources. Their achievement illustrated that Cyclohexane had the highest exergy efficiency, and RC318 had the lowest. Manente et al. (2017), in terms of using subcritical working conditions besides the comparison between dual pressure and the single pressure layout in the ORC systems that utilized a heat source, geothermal 100-200 °C performed a simulation by EES (Engineering Equation Solver) on some new, low GWP and ODP environmental friendly and also new working fluids like; R1234yf, R1234ze(E), and R1234ze(Z) in parallel with some old, famous, and typical working fluids such as; R134a, iC4, R245fa, iC5, and cC5. Their achievements revealed that R1234ze(E) and R1234yf had the highest and optimum net power output, total cycle efficiency, and thermal efficiency compared with the other working fluids.

Many scholars try to consider, evaluate, and compare pure and zeotropic mixtures of working fluids that drive ORC and pay attention to the thermodynamic analysis. For instance, Zhai et al. (2018) paid attention to zeotropic mixtures and pure working fluids of various types of old and new environmental working fluids in ORC and focus analyzed the performance parameters. Based on this study, applied in subcritical working conditions, the highest optimum cycle exergy efficiency belonged to 0.3R1234ze(E)/0.7R245fa with 44.45% in a subcritical ORC using a 210 °C heat source. Regarding thermo-economic

parameters, R236ea, R227ea, R245fa, and 0.2R227ea/0.8R245fa had the lowest cost value of the application, and R1234ze(E) had the highest cost value of the present ORC application. Kang et al. (2015) analyzed ten groups of old and low GWP working fluids, including HCs, HFCs, and HFOs. Furthermore, by applying Matlab and Refprop as numerical simulation software, they investigated the performances of an ORC in terms of geothermal heat source (120 °C). In conclusion, the impact of different evaporation temperatures on net power output, exergy efficiency, and thermal efficiency depicted the R1234yf/R601a (0.7/0.3) had the maximum optimum value compared with other zeotropic mixtures of working fluids. Zheng et al. (2018) performed a solar ORC integrated with vapor compression cycle (VCC) driving by R290, R161, R152a, R134a, R600a, R227ea, R1234yf, and R1234ze as pure working fluids and R290/R600a, R152a/R600a, R161/R600a, R227ea/R600a, and R1234yf/R600a as binary zeotropic mixtures of working fluid to improve the overall system performance. In a nutshell, among their working fluids, R161/R600a (0.25/0.75) revealed the highest optimum system efficiency with 0.3089. Moreover, in comparison, with pure working fluids R161 and R600a, the total efficiency of their binary zeotropic mixture showed a 54.7% and 39.6% increase, respectively.

Most present scholars pay attention to modeling and performing the organic Rankine cycle base on complex and particular initial working conditions by applying the pure and finally the fixed binary zeotropic mixtures of working fluids with small scale power generate and low exergy efficiency in parallel to the high investment cost and payback period. In contrast, this study uses the subcritical organic Rankine cycle in a simple structure and driven by pure, fixed binary zeotropic mixtures and a fixed ternary zeotropic mixture of two green and environmental friendly refrigerant working fluids beside an old but famous working fluid, efforts to generate optimum and large-scale power that suitable for CPI (>1MW) by utilizing the flue gas as a heat source with medium temperature, open kind, and no restriction of the outlet temperature to increase the exergy efficiency and reduce the investment cost and payback period in optimum values.

In current scholar, designing, modeling, simulating, analyzing, and investigating the Subcritical Organic Rankine Cycle (SORC) under steady-state conditions and simple structure also driven by pure, fixed binary zeotropic mixtures and a fixed ternary zeotropic mixture of two environmentally friendly and new working fluids like R1234ze(E) and R1234yf and an old and famous organic refrigerant working fluid like R134a, based on applying the flue gas as a heat source with medium temperature, open kind, and no restriction of the outlet temperature of heat source in the evaporation process, also the flue gas is released from industrial boilers, and then try to improve the exergy and total efficiency parameters and parallel with SIC, PPC, and PBP as thermo-economic parameters cause to generating optimum net power output.

MATERIALS AND METHODS

The System of ORC and its Thermodynamic Process

Figure 1 shows that the Subcritical Organic Rankine Cycle (SORC) applies R1234ze(E), R1234yf, and R134a as pure working fluids, R1234ze(E)/R1234yf (0.5/0.5), R1234ze(E)/R134a (0.5/0.5), and R1234yf/R134a (0.5/0.5) as fixed binary zeotropic mixtures of working fluids, the R1234ze(E)/R1234yf/R134a (0.4/0.3/0.3) as a fixed ternary zeotropic mixture of working fluid and the current SORC system is including; a pump, an evaporator, a vapor generator, a turbine, and a condenser. Figures 1 and 2 depict the schematic diagram of the current SORC system and, in parallel, the T-S diagram of the SORC thermodynamic process. Furthermore, in Figures 1 and 2, the thermodynamic state points are shown the thermodynamic process of the present SORC system.

Figures 1 and 2 belong to the SORC system; first, under a subcritical state, the working fluid in the saturated liquid state by using a feed pump is pressurized to subcooled liquid (1–2 process) as a compression process. Then the working fluid in a high-pressure by using an evaporation process, including an evaporator and a vapor generator, absorbs heat by applying a heat

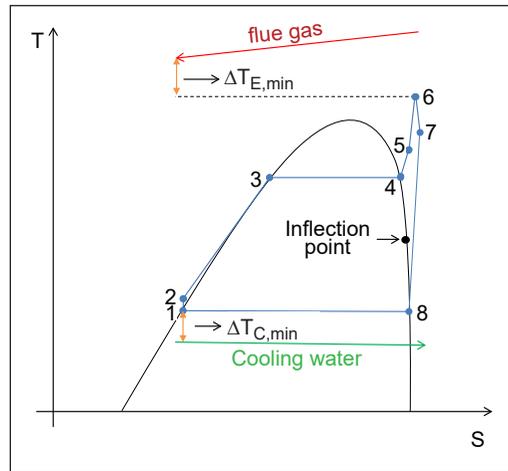


Figure 2. SORC thermodynamic process by applying flue gas as a heat source, cooling water as a heat sink, and using the Pure, Binary Zeotropic, and Ternary Zeotropic Mixtures of R1234ze(E), R1234yf, and R134a as Working Fluids

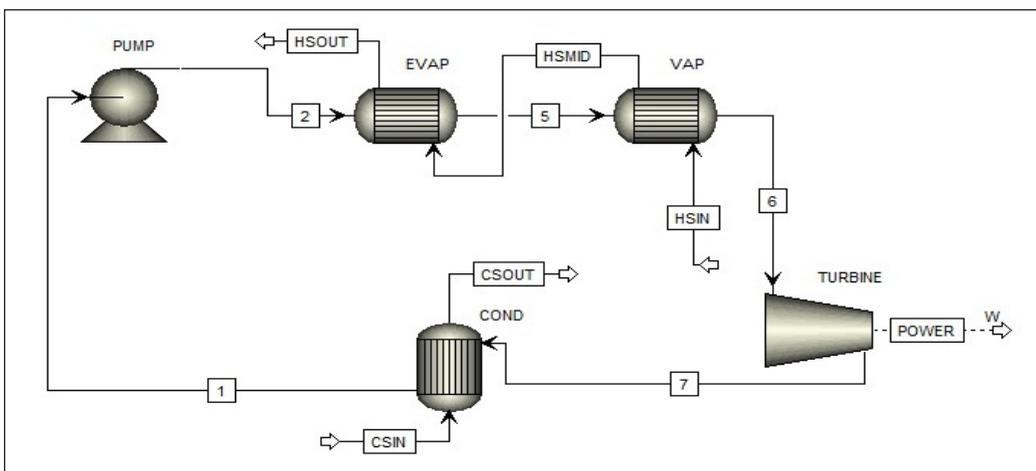


Figure 1. Schematic of the considered SORC system driven by pure, binary zeotropic, and ternary zeotropic mixtures of R1234ze(E), R1234yf, and R134a as Working Fluids

source, flue gas, and converts into saturated and then superheated vapor (2–3 process: preheating, 3–4 process: evaporation, 4–6 process: superheating). The vapor expands in the turbine and generates optimum power (6–7 process) as an expansion process. Finally, the working fluid discharges in a condenser and cools into saturated liquid (7–1 process) as a condensation process by contributing a heat sink and cooling water to complete this thermodynamic cycle.

Thermophysical Properties of Working Fluids

Table 1 presents the thermophysical properties of R1234ze(E), R1234yf, and R134a (Yang et al., 2020; Wang et al., 2020; Tian et al., 2020; Braimakis et al., 2020; Liu et al., 2020; Ata et al., 2020).

Table 1
Thermophysical properties of R1234ze(E), R1234yf, and R134a

Working Fluid	R1234ze(E)	R1234yf	R134a
Chemical Name	trans-1,3,3,3-tetrafluoroprop-1-ene	2,3,3,3-tetrafluoroprop-1-ene	1,1,1,2-tetrafluoroethane
Molecular Formula	CHF=CHCF ₃ (trans)	CF ₃ CF=CH ₂	CF ₃ CH ₂ F
Normal Boiling Point at 101.3 kPa/ °C	-18.97	-29.48	-26.07
Critical Pressure/ MPa	3.635	3.382	4.059
Critical Temperature/°C	109.36	94.70	101.06
Molecular Weight/ g/mol	114.04	114.04	102.03
Appearance	Colorless	Colorless	Colorless
GWP	<1	<1	1300
ODP	0	0	0
Lifetime in the Atmosphere/days	~4900	10.5	16.4
Safety Classification	A2L	A2L	A1

R1234ze(E) and R1234yf are isentropic, but the type of R134a is wet. Besides, in the inflection point (S_{ip}), the entropy reaches the maximum value of the two-phase zone on the saturation vapor curve. Likewise, the saturation vapor curve slope is negative above this inflection point and $(\frac{ds}{dT})_{sat} < 0$, so the working fluid presents a wet property. On the other hand, the saturation vapor curve slope is positive below this inflection point and $(\frac{ds}{dT})_{sat} > 0$; therefore, the working fluid shows a dry-type property.

Model Boundary Condition of SORC system

Tables 2, 3, and 4 show a list of the boundary conditions and the constraints of the present SORC system. The flue gas is a heat source with medium temperature, open kind, and no restriction of the outlet temperature of heat source chosen in the evaporation process; also, this kind of flue gas is released from industrial boilers. The inlet temperature of flue gas as

the medium temperature fixed at 220 °C and 0.1013 MPa; moreover, the heat source outlet temperature can decrease to ambient temperature or $T_2 + \Delta T_{e,min}$ without restriction. On the other hand, the superheated temperature was set to 5 °C increases to avoid the expansion process not passing through the two-phase region and based on the flash tolerance of each

Table 2
SORC system boundary conditions and constraints of pure working fluids

Parameter	Symbol	R1234ze(E)	R1234yf	R134a
Working Fluid mass flow rate/kg.s ⁻¹	\dot{m}_{wf}	277.778	277.778	277.778
flue gas mass flow rate/kg.s ⁻¹	\dot{m}_{HS}	555.556	555.556	555.556
Cooling water mass flow rate/kg.s ⁻¹	\dot{m}_{CS}	555.556	555.556	555.556
Mole Fractions	-	1	1	1
flue gas pressure/kPa	P_{HS}	101.325	101.325	101.325
flue gas inlet temperature/°C	$T_{HS,in}$	220	220	220
Condenser minimal temperature difference/°C	$\Delta T_{cond,min}$	15.098	4.329	5.735
Evaporator minimal temperature difference/°C	$\Delta T_{evap,min}$	73.038	85.23	83.469
Cooling water pressure/kPa	P_{CS}	101.325	101.325	101.325
Cooling water inlet temperature/°C	$T_{CS,in}$	10	10	10
Environment pressure/kPa	P_0	101.325	101.325	101.325
Environment temperature/°C	T_0	20	20	20
Feed pump efficiency/%	η_p	85	85	85
Feed pump pressure head/m	H	105.236	108.335	98.621
Turbine efficiency/%	η_t	72	72	72

Table 3
SORC System boundary conditions and constraints of binary zeotropic working fluids

Parameter	Symbol	R1234ze(E)/	R1234ze(E)/	R1234yf/
		R1234yf	R134a	R134a
Working Fluid mass flow rate/kg.s ⁻¹	\dot{m}_{wf}	277.778	277.778	277.778
flue gas mass flow rate/kg.s ⁻¹	\dot{m}_{HS}	555.556	555.556	555.556
Cooling water mass flow rate/kg.s ⁻¹	\dot{m}_{CS}	555.556	555.556	555.556
Mole Fractions	-	0.5/0.5	0.5/0.5	0.5/0.5
flue gas pressure/kPa	P_{HS}	101.325	101.325	101.325
flue gas inlet temperature/°C	$T_{HS,in}$	220	220	220
Condenser minimal temperature difference/°C	$\Delta T_{cond,min}$	8.538	7.325	4.982
Evaporator minimal temperature difference/°C	$\Delta T_{evap,min}$	79.844	81.433	84.516
Cooling water pressure/kPa	P_{CS}	101.325	101.325	101.325
Cooling water inlet temperature/°C	$T_{CS,in}$	10	10	10
Environment pressure/kPa	P_0	101.325	101.325	101.325
Environment temperature/°C	T_0	20	20	20
Feed pump efficiency/%	η_p	85	85	85
Feed pump pressure head/m	H	106.697	102.126	103.675
Turbine efficiency/%	η_t	72	72	72

Table 4
SORC system boundary conditions and constraints of ternary zeotropic working fluids

Parameter	Symbol	R1234ze(E)/R1234yf/R134a
Working Fluid mass flow rate/kg.s ⁻¹	\dot{m}_{wf}	277.778
flue gas mass flow rate/kg.s ⁻¹	\dot{m}_{HS}	555.556
Cooling water mass flow rate/kg.s ⁻¹	\dot{m}_{CS}	555.556
Mole Fractions	-	0.4/0.3/0.3
flue gas pressure/kPa	P_{HS}	101.325
flue gas inlet temperature/°C	$T_{HS,in}$	220
Condenser minimal temperature difference/°C	$\Delta T_{cond,min}$	7.0725
Evaporator minimal temperature difference/°C	$\Delta T_{evap,min}$	81.805
Cooling water pressure/kPa	P_{CS}	101.325
Cooling water inlet temperature/°C	$T_{CS,in}$	10
Environment pressure/kPa	P_0	101.325
Environment temperature/°C	T_0	20
Feed pump efficiency/%	η_p	85
Feed pump pressure head/m	H	104.33
Turbine efficiency/%	η_t	72

working fluid in the evaporator. Likewise, each working fluid's mass flow rate is set to 277.778 kg.s⁻¹, and the mass flow rate of flue gas and cooling water is set to 555.556 kg.s⁻¹ as the main effective parameters in parallel expansion enthalpy change on the optimum power generating. There are some constraints of this study as well.

The lower and upper limits of P_2 (compression pressure) based on subcritical working conditions investigate because these limit values are accessible, efficient, relatively safe, and economical to achieve. Furthermore, to prevent the influence of the significant thermophysical property variations close to the critical zone of each working fluid. The optimized variable input parameter is compression pressure. Also, it defines as a turbine inlet pressure. In the cycle, the pump's inlet pressure lower limit increase sets into the 0.4 MPa (reach into the 1.4 MPa as a minimum boundary condition of compression pressure) to mitigate air ingress and prevent sub-atmospheric pressure leading to approach optimum net power output generated (>1MW). Moreover, the upper limit of compression pressure increase is set to 1.2 MPa (reach into the 2.2 MPa as a maximum boundary condition of compression pressure) to prevent the flash calculation failed, and temperature cross detected in the T-Q diagram of the evaporator.

Assumptions

The following general assumptions are made to simplify the current SORC system analysis:

- Simplifying the complexity of the model.
- Steady-state operating of the SORC system.
- Neglecting heat and pressure loss in heat exchangers and pipes.

- Friction pressure drop is neglected in pipelines and heat exchangers.
- Neglecting the influence of gravitational potential energy and fluid kinetic.
- The heat exchangers arrangement in a countercurrent flow type.
- Supplying the constant temperatures for the evaporation system's heat source and the condenser's heat sink.

The SORC System Theoretical Equations

The Thermodynamic Equations. The net power output generated by the SORC system is as in Equation 1:

$$W_{net} = W_T - W_P \quad [1]$$

Where W_t is the turbine power generated, W_p is the pump power consumed.

The heat absorption capacity of the SORC system is as in Equation 2:

$$Q_{SORC} = \dot{m}_{HS} (h_{HS,in} - h_{HS,out}) \quad [2]$$

Where \dot{m}_{HS} is the flue gas mass flow rate, also $h_{HS,in}$ and $h_{HS,out}$ are the flue gas inlet and outlet enthalpies, respectively.

The SORC total efficiency is as in Equation 3:

$$\eta_{SORC} = \frac{W_{net}}{Q_{SORC}} \quad [3]$$

The $T_0 = 20^\circ\text{C}$ and $P_0 = 101.325\text{ kPa}$ are chosen as the environment state reference. The exergy released in the evaporator and vapor generator system by the heat source is as in Equation 4:

$$\Delta E_{HS} = \dot{m}_{HS} (h_{HS,in} - h_{HS,out} - T_0 (S_{HS,in} - S_{HS,out})) \quad [4]$$

Where $S_{HS,in}$ and $S_{HS,out}$ are the inlet and outlet of the heat source entropies, respectively. The exergy absorbed in the evaporator and vapor generator system by the working fluid is as in Equation 5:

$$\Delta E_{wf,abs} = \dot{m}_{wf} (h_6 - h_2 - T_0 (S_6 - S_2)) \quad [5]$$

Where \dot{m}_{wf} is the working fluid mass flow rate, likewise h_6 and h_2 are the vapor generator outlet enthalpy and the evaporator inlet enthalpy, respectively. Moreover, S_6 and S_2 are the vapor generator outlet entropy and the evaporator inlet entropy, respectively.

Hence, the evaporator and vapor generator exergy destruction rates (the exergy loss or the unused exergy) of application are as in Equation 6:

$$I_{E,V} = \Delta E_{HS} - \Delta E_{wf,abs} = T_0 (\dot{m}_{wf} (S_6 - S_2) - \dot{m}_{HS} (S_{HS,in} - S_{HS,out})) \quad [6]$$

The turbine exergy destruction rate is as in Equation 7:

$$I_T = T_0 \dot{m}_{wf} (S_7 - S_{7s}) \quad [7]$$

Where S_7 and S_{7s} are the turbine outlet and isentropic state entropies, respectively. The pump exergy destruction rate is as in Equation 8:

$$I_P = T_0 \dot{m}_{wf} (S_2 - S_{2s}) \quad [8]$$

Where S_2 and S_{2s} are the pump outlet and isentropic state entropies, respectively. Similar to the evaporator and vapor generator system heat transfer process, the condenser exergy release is as in Equation 9:

$$\Delta E_{wf,rel} = \dot{m}_{wf} (h_7 - h_1 - T_0 (S_7 - S_1)) \quad [9]$$

Where h_1 and h_7 are the condenser outlet and inlet enthalpies, respectively. Moreover, S_1 and S_7 are the condenser outlet and inlet entropies, respectively. Also, the exergy absorbed from the condenser by the heat sink is as in Equation 10:

$$\Delta E_{CS} = \dot{m}_{CS} (h_{CS,out} - h_{CS,in} - T_0 (S_{CS,out} - S_{CS,in})) \quad [10]$$

Where \dot{m}_{CS} is the cold stream as a heat sink mass flow, moreover $h_{CS,out}$ also $h_{CS,in}$ are the cold stream outlet and inlet enthalpies, respectively. Furthermore, $S_{CS,out}$ and $S_{CS,in}$ are the cold stream outlet and inlet entropies, respectively.

So, the condenser exergy destruction rate is as in Equation 11:

$$I_C = \Delta E_{wf,rel} - \Delta E_{CS} = T_0 (\dot{m}_{CS} (S_{CS,out} - S_{CS,in}) - \dot{m}_{wf} (S_7 - S_1)) \quad [11]$$

The exergy balance of the SORC system is declared by Equation 12:

$$\Delta E_{HS} + \Delta E_{CS} = W_t - W_p + \sum I_i \quad [12]$$

In each process of the present SORC system, the exergy destruction rate coefficient is as in Equation 13:

$$\xi_i = \frac{I_i}{\Delta E_{HS} - \Delta E_{CS}} \quad [13]$$

To sum up, the exergy destruction rate coefficient of the SORC system rate is given by Equation 14:

$$\xi_{SORC} = \sum \xi_i = \frac{I_{E,V} + I_T + I_P + I_C}{\Delta E_{HS} - \Delta E_{CS}} = 1 - \frac{W_t - W_p}{\Delta E_{HS} - \Delta E_{CS}} = 1 - \frac{W_{net}}{\Delta E_{HS} - \Delta E_{CS}} \quad [14]$$

All in all, the total exergy efficiency is as in Equation 15:

$$\eta_{ex} = \frac{W_{net}}{\Delta E_{HS} - \Delta E_{CS}} = 1 - \sum \xi_i \quad [15]$$

The Thermo-Economic Equations of the SORC System

Evaluation of the thermo-economic parameters as a substitute for the efficiency of the SORC system consists of specific investment cost (SIC), power production cost (PPC), and payback period (PBP).

The cost of all components of the current SORC system is calculated using Equation 16:

$$Cost_{2001} = C_{bmE} + C_{bmV} + C_{bmC} + C_{bmT} + C_{bmP} \quad [16]$$

Where C_{bm} is the cost of SORC components and $Cost_{2001}$ is the cost of all components in 2001 presented by the chemical engineering plant cost index (CEPCI) (Turton et al., 2012).

Furthermore, converting the cost of all components from 2001 to 2021 can be achieved by using Equation 17:

$$Cost_{2021} = \frac{Cost_{2001} CEPCI_{2021}}{CEPCI_{2001}} \quad [17]$$

Which $CEPCI_{2001} = 397$, $CEPCI_{2021} = 624.0269$ (Turton et al., 2012; Mignard, 2014).

Hence, the cost of SORC components is given by Equations 18-29:

For evaporator:

$$\log C_{pE} = K_1 + K_2 \log (A_E) + K_3 [\log (A_E)]^2 \quad [18]$$

$$C_{bmE} = C_{pE} F_{bmE} \quad [19]$$

For vapor generator:

$$\log C_{pV} = K_1 + K_2 \log (A_V) + K_3 [\log (A_V)]^2 \quad [20]$$

$$C_{bmV} = C_{pV} F_{bmV} \quad [21]$$

For condenser:

$$\log C_{pC} = K_1 + K_2 \log (A_C) + K_3 [\log (A_C)]^2 \quad [22]$$

$$C_{bmC} = C_{pC} F_{bmC} \quad [23]$$

For pump:

$$\log C_{pP} = K_1 + K_2 \log (W_P) + K_3 [\log (W_P)]^2 \quad [24]$$

$$\log F_{pP} = C_1 + C_2 \log (W_P) + C_3 [\log (W_P)]^2 \quad [25]$$

$$F_{bmP} = B_1 + B_2 F_M F_{pP} \quad [26]$$

$$C_{bmP} = C_{pP} F_{bmP} \quad [27]$$

For turbine:

$$\log C_{pT} = K_1 + K_2 \log (W_T) + K_3 [\log (W_T)]^2 \quad [28]$$

$$C_{bmT} = C_{pT} F_{bmT} \quad [29]$$

Where the $K_1, K_2, K_3, C_1, C_2, C_3, B_1, B_2, F_{bm}, F_M$ are the cost price correction factors of evaporator “E”, vapor generator “V”, condenser “C”, pump “P”, and turbine “T”, as displayed in Table 5.

Table 5
Correlation coefficient of SORC components cost (Turton et al., 2012)

	K_1	K_2	K_3	C_1	C_2	C_3	B_1	B_2	F_M	F_{bm}
Evaporator, Vapor Generator, and Condenser	4.6420	0.3698	0.0025	/	/	/	/	/	/	2.9
Pump	3.3892	0.0536	0.1538	0	0	0	1.89	1.35	1.5	/
Turbine	2.7051	1.4398	0.1776	/	/	/	/	/	/	3.5

Besides, the cost of labor is calculated using Equation 30:

$$\text{Cost}_{\text{Labor}} = 0.3 \text{Cost}_{2021} \quad [30]$$

In a nutshell, one of the foremost performance parameters which equivalent to the profitability of the SORC system is the specific investment cost (SIC) that, based on neglecting the maintenance and insurance annual cost, is calculated using Equation 31:

$$\text{SIC} = \frac{\text{Cost}_{2021} + \text{Cost}_{\text{Labor}}}{W_{\text{net}}} \quad [31]$$

Likewise, the heat source of the SORC system, flue gas, is assumed to be cost-free.

The capital recovery factor (CRF) is a prerequisite factor of power production cost (PPC) and as a converter parameter to convert a current value into an annual cost base on a specified time and discount rate is calculated using Equation 32:

$$\text{CRF} = \frac{i(1+i)^{LT}}{[(1+i)^{LT} - 1]} \quad [32]$$

Where i is the annual interest rate and assumed to be 5%, LT is the system lifetime and assumed to be 15 years.

The cost of the SORC system to generate the net power output is power production cost (PPC) and achieved by using Equation 33:

$$PPC = \frac{[CRFCost_{2021} + iCost_{2021}]}{(W_{net}t_{op})} \quad [33]$$

Where t_{op} is the system operation time and assumed to be 7500 h annually (Nafey & Sharaf, 2010).

The payback period (PBP) is the period that is needed to recover the total investment cost of the SORC system and is calculated using Equation 34:

$$PBP = \frac{\ln[(W_{net}C_e)/(iCost_{2021} - W_{net}C_e)]}{\ln(1+i)} \quad [34]$$

Where C_e is the power price in 2021 in the U.S. and assumed to be 0.0696 \$/kWh (<https://www.eia.gov/outlooks/steo/report/electricity.php>).

RESULT AND DISCUSSION

Simulation Software Validation

The validity confirmation of SORC system components, R1234ze(E), R1234yf, and R134a as working fluids, heat source (flue gas), heat sink (cold water) in the AspenPlus (v10) simulation software has been valid extensively. Most current results in this study have been validated and tested using established data and compared with several results in a similar condition from published works of literature as having been done, accepted, and published in the past. The current results from the AspenPlus (v10) are consistent, with no significant deviation arising for all trials.

The AspenPlus (v10) software's library source comprehensively uses several equations of state as a calculation method to evaluate and estimate the number of thermodynamic parameters in the SORC system. Hence, based on the current SORC system condition for more accuracy, the REFPROP selects as an equation of state in this simulation. In the present study, the significant number of results before this is done and published by other scholars with similar conditions compared and verified by the current results of this study. The proposed SORC system is a novel configuration and has not been studied and considered in previous literature. So, to verify and validate the current study conducted in the same condition in Li et al. (2017). As shown in Table 6, the total efficiency and net power output of R1234ze(E) in the present SORC system investigate by comparison with Li et al. (2017), with the maximum derivation from the date in reference is 10.00% for the

total efficiency of the SORC system. Since each section’s derivation of this SORC system is reasonably negligible, the simulation model and the results were accurate enough for further investigations.

Table 6
Verification and validation of the simulation results of the present work with Li et al. (2017)

P_T (MPa)	T_T (°C)	$\eta_{Present}$ (%)	$\eta_{Ref.}$ (%)	%	$W_{net,Present}$ (kW)	$W_{net,Ref.}$ (kW)	%
1.58	109.144	0.064	0.071	9.86	44.12	44.80	1.52
1.38	103.285	0.057	0.062	8.06	38.47	39.00	1.36
1.18	96.748	0.048	0.051	5.88	31.77	33.00	3.73
0.98	89.318	0.036	0.040	10.00	23.64	24.80	4.68

Performance Analysis of Subcritical Organic Rankine Cycle (SORC)

Figure 3 illustrates the effect of different compression pressure (1.4–2.2 MPa) on the net power output of R1234ze(E), R1234yf, and R134a as pure working fluids, R1234ze(E)/R1234yf (0.5/0.5), R1234ze(E)/R134a (0.5/0.5), and R1234yf/R134a (0.5/0.5) as fixed binary zeotropic mixtures of working fluids, and the R1234ze(E)/R1234yf/R134a (0.4/0.3/0.3) as a fixed ternary zeotropic mixture of working fluid. As a result, the highest

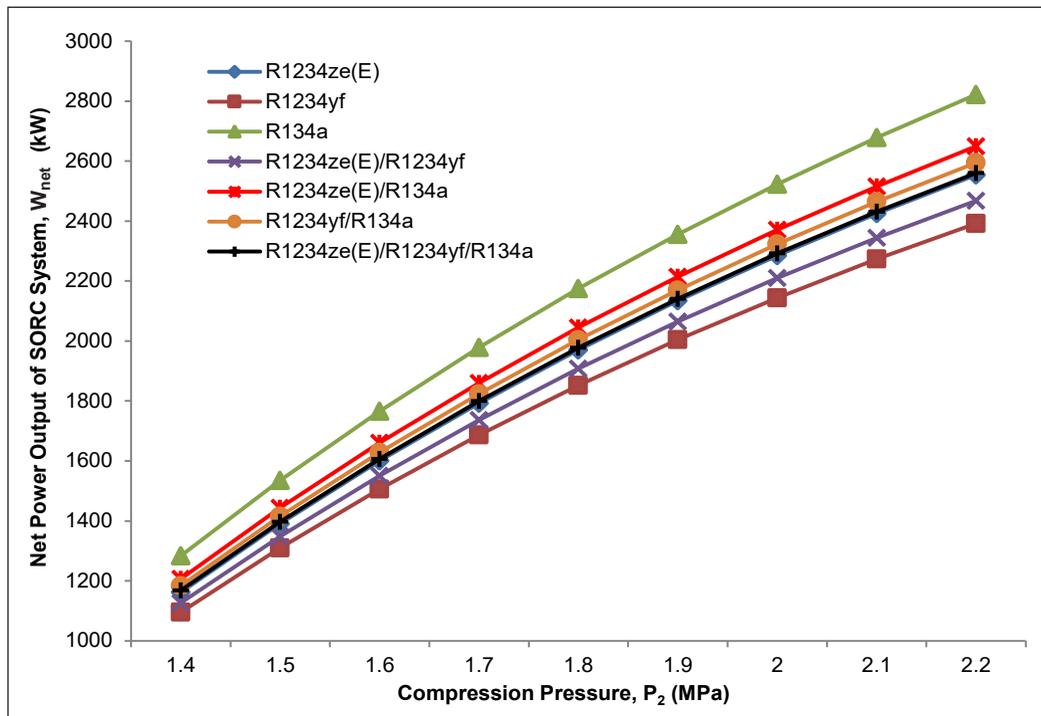


Figure 3. The net power output generated by SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

SORC net power output range is 1283.22 KW to 2822.04 KW and belongs to the R134a. Also, the lowest range of net power output corresponding to R1234yf with 14.62–15.24% reduction shows 1095.58 KW to 2391.81 KW. Therefore, between the highest to the lowest range of SORC system net power output, the R1234ze(E)/R134a with 1207.02–2650.02 KW, R1234yf/R134a with 1183.86–2595.13 KW, R1234ze(E)/R1234yf/R134a with 1167.92–2560.27 KW, R1234ze(E) with 1162.40–2554.04 KW, and R1234ze(E)/R1234yf with 1127.27–2467.84 KW are adapted, respectively. As depicted in Figure 3, increasing the compression pressure causes an increase in the evaporation temperature as a consequence of enthalpy change of flue gas and the evaporation process growth, and totally the conclusion increases the turbine output power and, as a result, increases the net power output in each working fluid. The main reason for these fluctuations between different kinds of working fluids is that the increasing expansion enthalpy change of each working fluid leads to an increase in the net power output because of the turbine enthalpy conversion into power. In parallel, the flue gas enthalpy changes in the evaporation system and, therefore, the heat absorption capacity reveals to a large extent the positive relationship with the SORC system net power output. These results agree with Invernizzi et al. (2016) and Bianchi et al. (2020).

As depicted in Figure 4, the impact of different compression pressure (1.4–2.2 MPa) considered on the heat absorption capacity of R1234ze(E), R1234yf, and R134a as pure working fluids, R1234ze(E)/R1234yf (0.5/0.5), R1234ze(E)/R134a (0.5/0.5), and R1234yf/

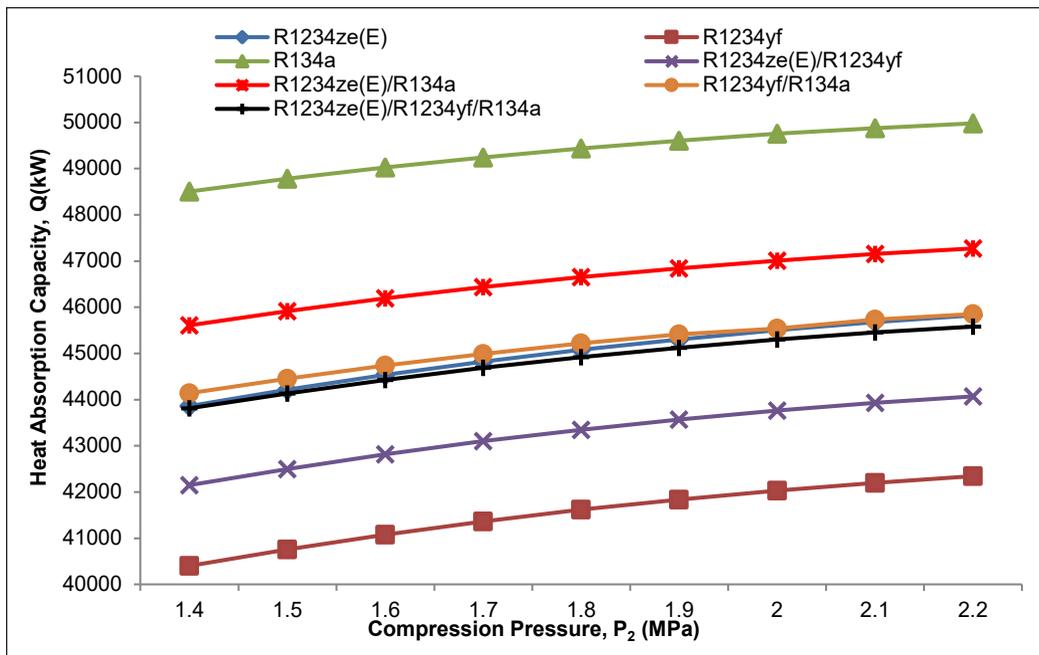


Figure 4. The heat absorption capacity of SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

R134a (0.5/0.5) as fixed binary zeotropic mixtures of working fluids, and the R1234ze(E)/R1234yf/R134a (0.4/0.3/0.3) as a fixed ternary zeotropic mixture of working fluid. The quantitative amount of heat absorption capacity of SORC belongs to the R134a as the highest amount with the value range of 48505.59–49983.37 kW, and the lowest amount belongs to the R1234yf, with a 16.71–15.28% reduction for the minimum to maximum compression pressure, as well. Likewise, the R1234ze(E)/R134a, R1234yf/R134a, R1234ze(E), R1234ze(E)/R1234yf/R134a, and R1234ze(E)/R1234yf are placed between highest amount to lowest amount of heat absorption capacity, respectively. By analyzing and investigating the heat absorption capacity as the main parameter of energy recovery and its equation, the flue gas enthalpy change significantly affects the heat absorption capacity of the SORC system. This parameter is directly related to the mass enthalpy change in the evaporation and, therefore, in the expansion process of each working fluid as a most effective parameter to generate optimum turbine power output. To sum up, these results similar to the results of Ji et al. (2021), Rowshanaie et al. (2020), and Rowshanaie et al. (2015).

As illustrated in Figure 5, the simulation model of SORC overall efficiency using Matlab Simulink software in Equations 1, 2, and 3 are considered and calculated to enhance the analysis accuracy in the current study.

From Figure 6, the total efficiency of SORC system for different compression pressures (1.4–2.2 MPa) of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids are disputed. Figure 6 indicates that the total efficiency of SORC has a positive relationship with compression pressure and, as a result, with the inlet pressure of the turbine. In parallel, the significant power generated by the turbine, minimal impact of power consumption by the pump, and the present SORC heat absorption capacity is the foremost reason for different insignificant values of total efficiency between each pure, binary zeotropic. Ternary zeotropic working

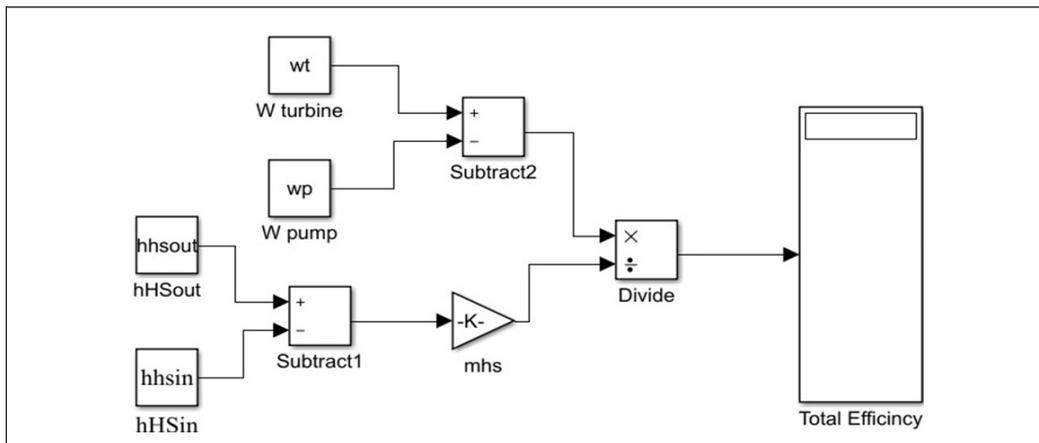


Figure 5. The total efficiency simulation analysis model of the SORC system

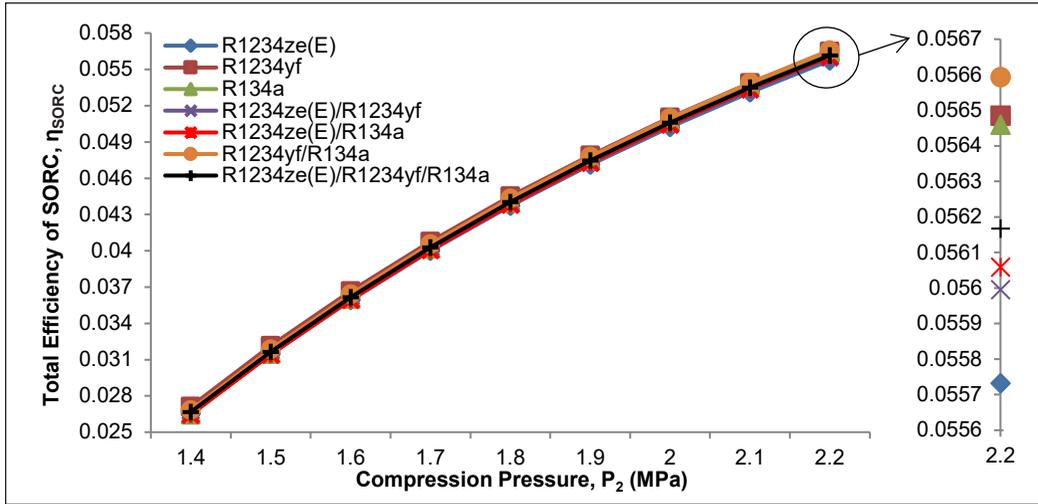


Figure 6. The total efficiency of SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids.

fluids in each point of compression pressure or inlet pressure of turbine are minimum temperature differences at evaporation process, which has a direct relationship with a SORC overall efficiency. In brief, these results are similar to the results of Li et al. (2017), Vera et al. (2020), Cambi et al. (2016), and Rowshanaie et al. (2020).

Figure 7 shows the simulation model of exergy efficiency that calculates in the Matlab Simulink software by contributing Equations 1, 4, 10, and 15.

Figure 8 tries to indicate the influence of different compression pressures on the exergy efficiency of SORC for pure working fluids like; R1234ze(E), R1234yf, R134a, and binary zeotropic working fluids such as R1234ze(E)/R1234yf (0.5/0.5), R1234ze(E)/R134a (0.5/0.5), R1234yf/R134a (0.5/0.5), and ternary zeotropic working fluid including; R1234ze(E)/R1234yf/R134a. As illustrated in Figure 7, at the minimum boundary condition of compression pressure (1.4 MPa), the increasing trend of exergy efficiency starts from R1234ze(E)/R134a with 0.4866, R1234ze(E)/R1234yf with 0.4871, R134a with 0.4898, R1234yf/R134a with 0.4908, R1234ze(E) with 0.4946, R1234ze(E)/R1234yf/R134a with 0.4962 and reach to R1234yf with 0.5019 as the highest exergy efficiency in the lowest compression pressure of current SORC system. Hence, with increasing the compression pressure (1.4-2.2 MPa), the exergy efficiency of each working fluid grows dramatically. In this condition, at the maximum boundary condition of compression pressure (2.2 MPa), the increasing trend of exergy efficiency is started from R1234ze(E)/R1234yf/R134a with 0.6835, R1234ze(E)/R134a with 0.6855, R1234ze(E) with 0.6866, R1234yf with 0.6872, R1234ze(E)/R1234yf with 0.6881, R134a with 0.6905, and achieve to R1234yf/R134a with 0.6929 as the highest exergy efficiency in the highest compression pressure of present SORC system. The main impacts of these minimal differences between these pure, binary,

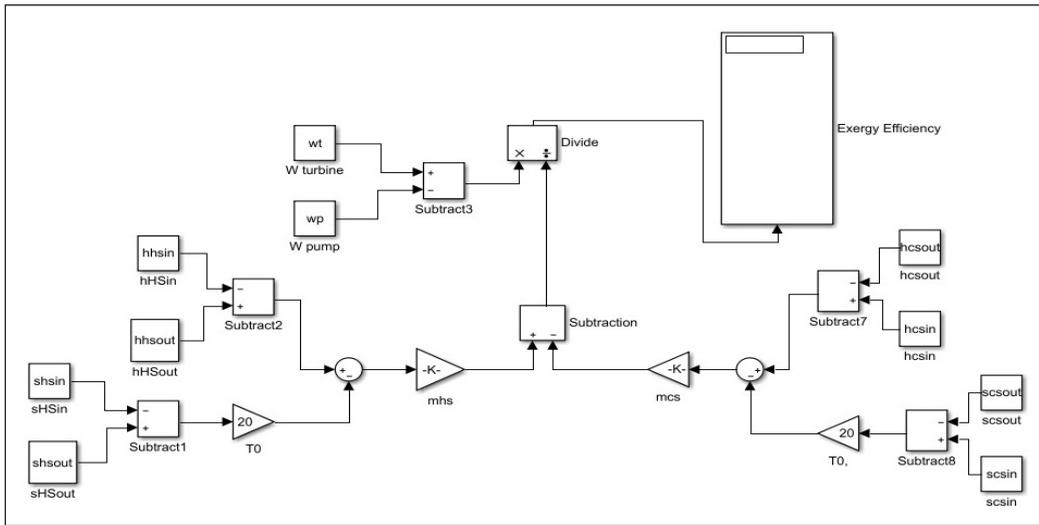


Figure 7. The exergy efficiency simulation analysis model of the SORC system

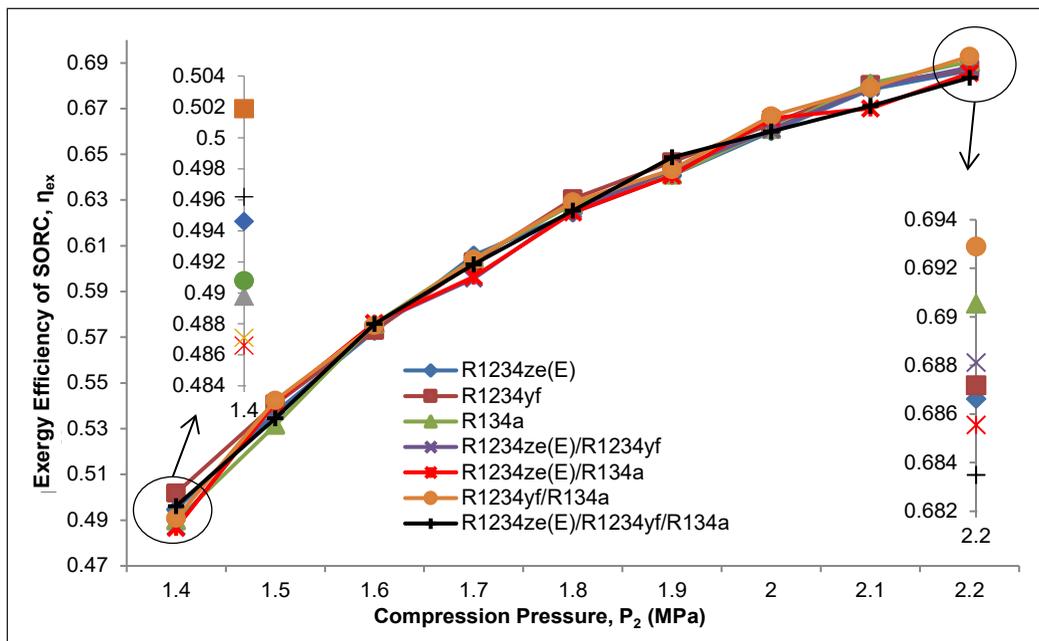


Figure 8. The SORC exergy efficiency at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

and ternary zeotropic working fluids are differences in power generated by the turbine and power consumed by the pump as a direct relationship with exergy efficiency. Also, in parallel, the flue gas and the cold water enthalpy different and entropy different have a direct relationship with exergy efficiency. Overall, these results are supported by Guo et al. (2021), Shengjun et al. (2011), and Rowshanaie et al. (2020).

The simulation calculation model of specific investment cost (SIC) as an essential economic efficiency parameter by Simulink of Matlab simulation software depicts in Figure 9, based on Equations 1 and 16-31.

Figure 10 compares the influence of different compression pressure (1.4–2.2 MPa) on specific investment cost (SIC) as economic efficiency of the current SORC system. As illustrated in Figure 10, increasing the compression pressure and enhancing the turbine inlet pressure will drastically glide the SIC. The maximum value to minimum value of SIC in terms of different compression pressure belongs to R134a with 5807402.18–22455670.61 \$.kW⁻¹, and compared with this, the R1234ze(E)/R134a with 7.12–7.32% reduction, R1234yf/R134a with 9.11–9.42% reduction, R1234ze(E)/R1234yf/R134a with 10.52–10.88% reduction, R1234ze(E) with 10.99–10.95% reduction, R1234ze(E)/R1234yf with 14.12–14.46% reduction, and reach to the R1234yf with 16.82–17.38% reduction and achieve to the amount of 4830207.61–18551143.42 \$.kW⁻¹ as a minimum value of SIC. The foremost reasons are to increase the required exchanger area and, in parallel, to increase the net power generated by the turbine, leading to increasing cost of each component, including the pump, evaporator, vapor generator, turbine, and condenser; as a result cause to increasing SIC of SORC system, dramatically. It notes that these achievements are similar to Quoilin et al. (2011).

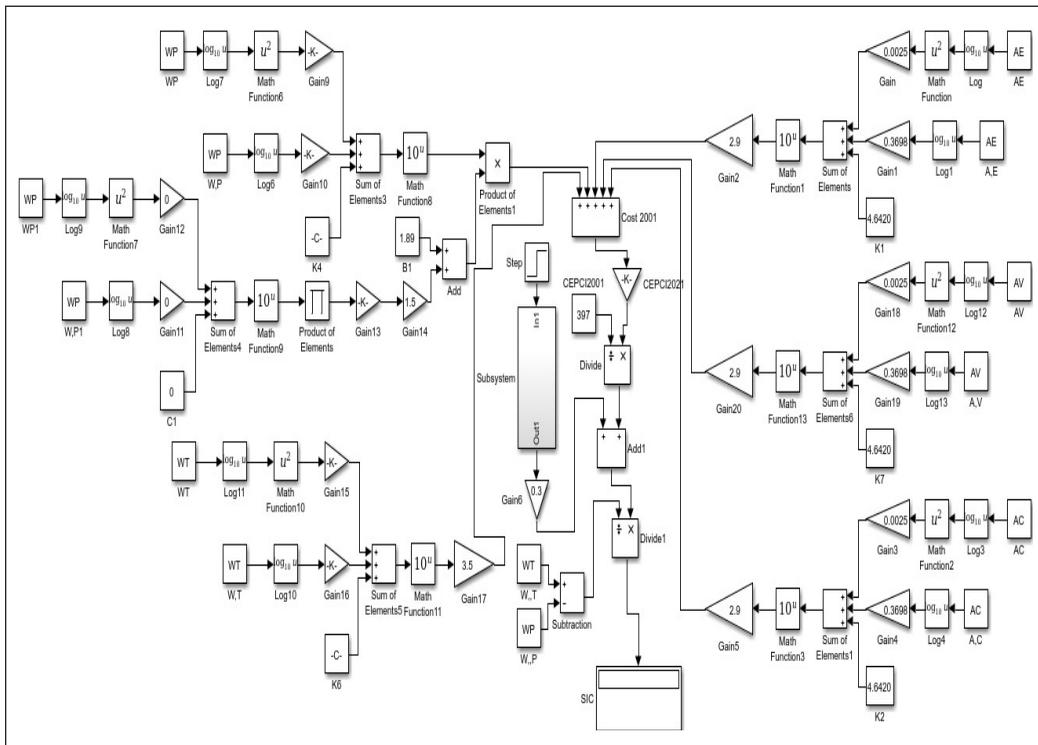


Figure 9. The SIC simulation analysis model of the SORC system

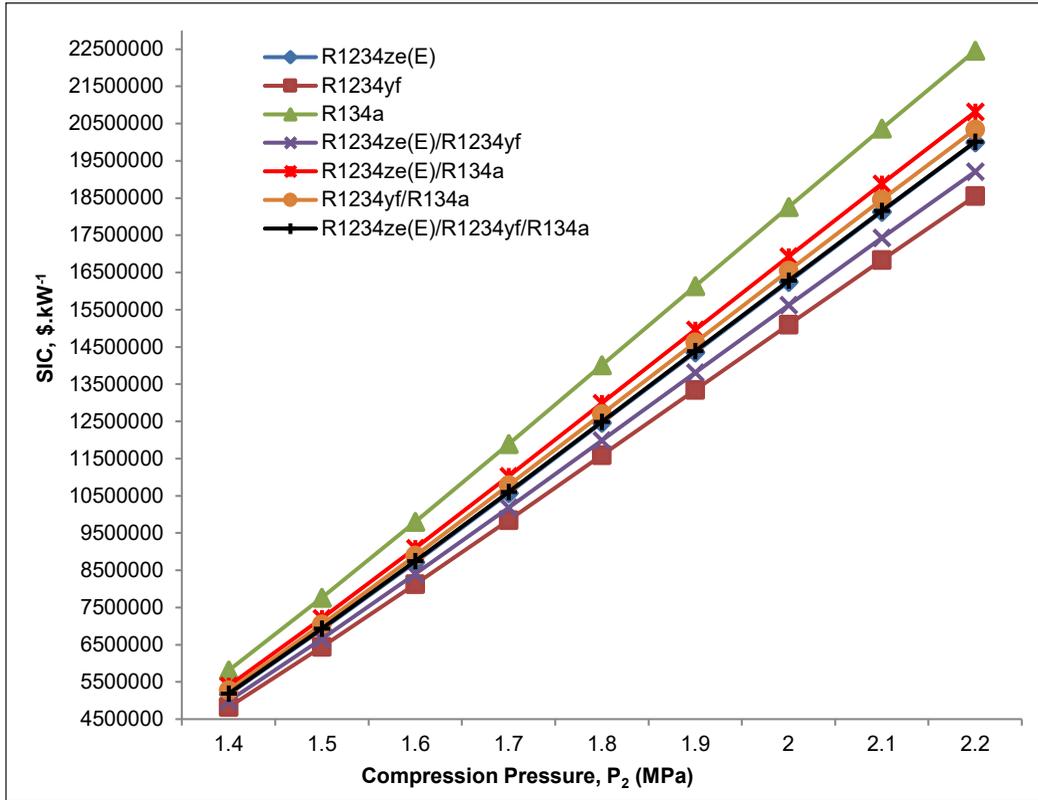


Figure 10. The specific investment cost (SIC) of SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

Figure 11 illustrates the calculation model of power production cost (PPC) in the Simulink of Matlab simulation software in Equations 1, 16–29, 32, and 33.

As considered in Figure 12, the relationship between the compression pressures and, in the same meaning, the turbine inlet pressure (1.4–2.2 MPa) and the power production cost (PPC) of the SORC system are discussed. As illustrated in this graph, the trend of PPC with increasing compression pressure depicts a positive correlation. In this condition, increasing the total cost of the present SORC application and the SORC system net power output at the same time is affected by flue gas enthalpy change, leading to increased PPC, remarkably. In Figure 12, the maximum to minimum range amount of PPC to comparison between different pure, binary, and ternary zeotropic working fluids belongs the R134a, R1234ze(E)/R134a, R1234yf/R134a, R1234ze(E)/R1234yf/R134a, R1234ze(E), R1234ze(E)/R1234yf, and R1234yf with values range of 81.74–316.07, 75.91–292.94, 74.29–286.30, 73.08–281.67, 72.76–281.44, 70.20–270.37, and 67.99–261.11 $\$/kW \cdot year^{-1}$, respectively. The result of this economic parameter agrees with Li et al. (2019) and Zhang et al. (2017).

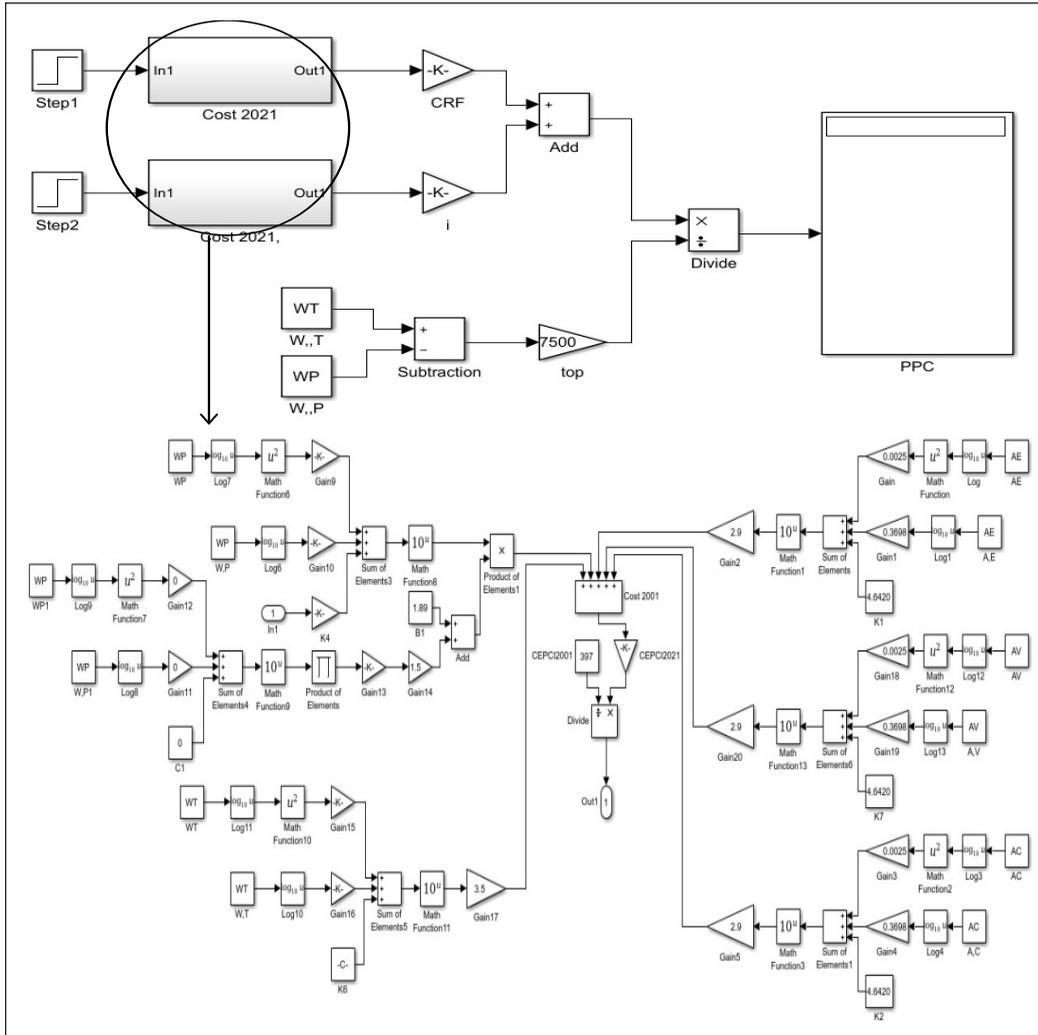


Figure 11. The PPC simulation analysis model of the SORC system

Figure 13 illustrates the simulation analysis method by applying the Simulink of Matlab simulation software for the payback period (PBP) of the SORC system by utilizing Equations 1 and 16–29.

The payback period (PBP) of the SORC system investigates different compression pressure; likewise, turbine inlet pressure (1.4–2.2 MPa) for each pure, binary, and ternary zeotropic working fluid is well present in Figure 14. As shown in this graph, the PBP of the current SORC system has a positive relationship with compression pressure and in parallel with turbine inlet pressure. Moreover, some essential thermodynamic parameters have a pivotal influence on enhancing the PBP in general. The main reasons for these fluctuation differences between each type of working fluids are the required exchanger's area, net power output, cost of each component, and heat absorption capacity. The lowest

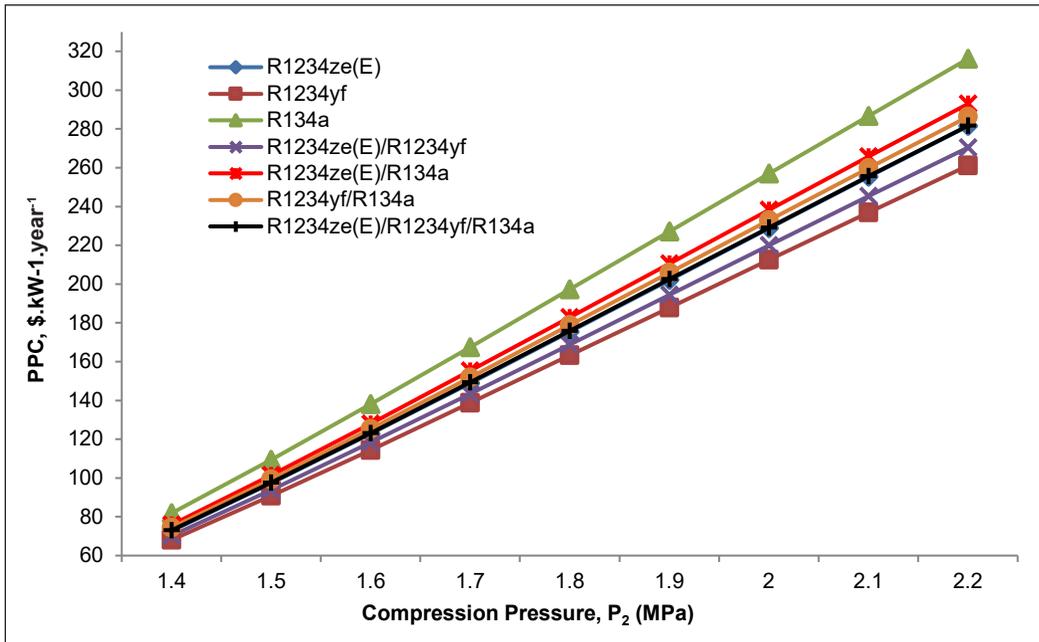


Figure 12. The power production cost (PPC) of SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

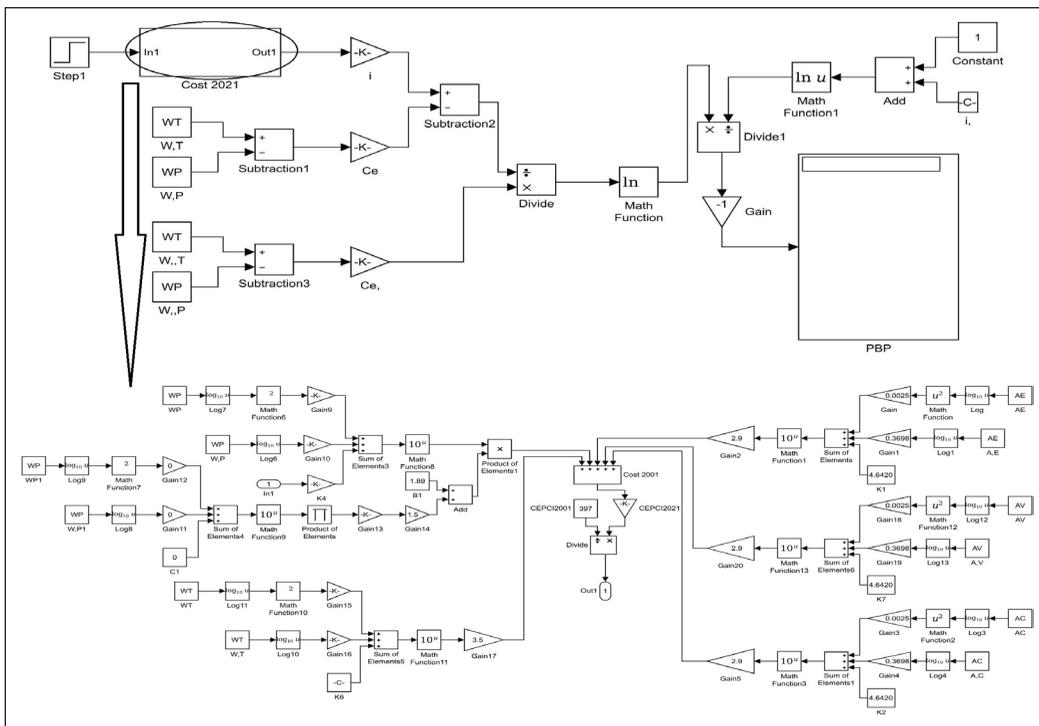


Figure 13. The PBP simulation analysis model of the SORC system

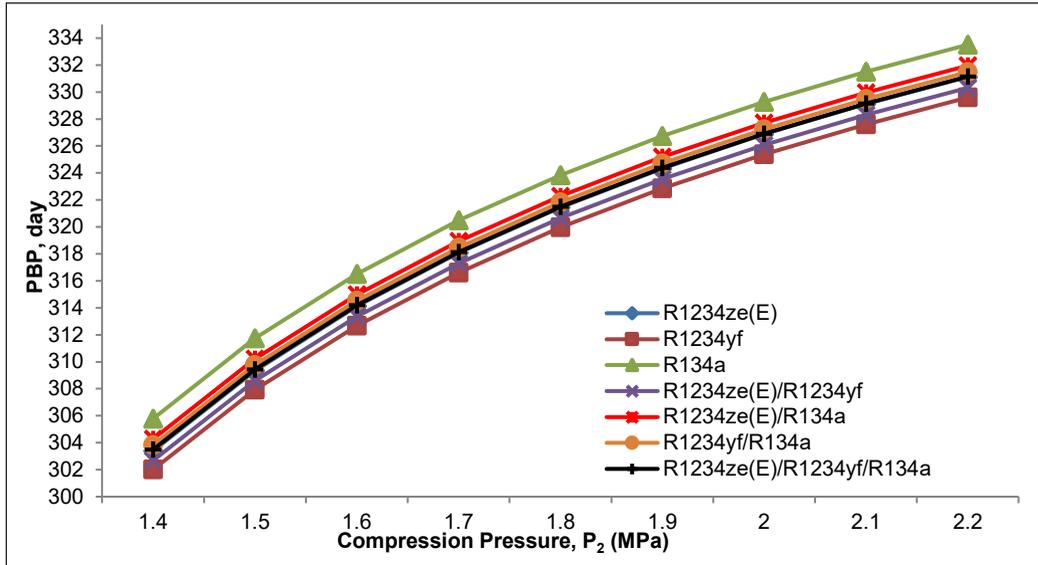


Figure 14. The payback period (PBP) of SORC system at different compression pressures of R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf, R1234ze(E)/R134a, R1234yf/R134a, and R1234ze(E)/R1234yf/R134a as working fluids

and highest PBP of the SORC system belongs to the R1234yf and R134a with 302-330 and 306-333 days, respectively. Moreover, followed by them the R1234ze(E)/R1234yf, R1234ze(E)/R1234yf/R134a, R1234ze(E), R1234yf/R134a, and R1234ze(E)/R134a locate between these lowest to highest PBP of the SORC system. It is essential to say that these achievements are similar to the results of Li et al. (2019), Yang (2018), and Zhang et al. (2017).

CONCLUSION

In a nutshell, this study focused on designing, modeling, simulating, conducting, analyzing, and evaluating the Subcritical Organic Rankine Cycle (SORC), which is driven by pure, fixed binary and ternary zeotropic working fluids such as; R1234ze(E), R1234yf, R134a, R1234ze(E)/R1234yf (0.5/0.5), R1234ze(E)/R134a (0.5/0.5), R1234yf/R134a (0.5/0.5), and R1234ze(E)/R1234yf/R134a (0.4/0.3/0.3), respectively. Also, try to apply the flue gas (220°C), released from industrial boilers, as a heat source with medium temperature, open kind, and no restriction of the outlet temperature of the heat source in the evaporation process. Furthermore, a heat sink (cold water) and try to model, design, and simulation by utilizing the AspenPlus (v10), Simulink of Matlab (vR2017a), and REFPROP (v10) to analyze and achieve the optimum value of the thermo-physical and thermo-economic parameters based on compression pressure (1.4-2.2 MPa) as the main boundary condition. The main conclusions are made as follows.

Increasing the compression pressure has a positive relationship with the superheated temperature and the mass enthalpy change in the evaporation. Therefore, in the expansion process, the flue gas enthalpy change significantly affects the heat absorption capacity of the SORC system. This parameter is directly related to the mass enthalpy change in the evaporation and, therefore, in the expansion process of each working fluid as an effective parameter to generate optimum turbine power output ($>1\text{MW}$) of each working fluid. The highest and optimum range of the SORC net power output is 1283.22–2822.04 KW when R134a is adapted. Also, the lowest and optimum range of net power output belongs to R1234yf with only a 14.62–15.24% reduction. Likewise, R1234ze(E)/R134a, R1234yf/R134a, R1234ze(E)/R1234yf/R134a, R1234ze(E), and R1234ze(E)/R1234yf find out between the highest to lowest range of net power output, respectively.

The result of optimum exergy efficiency establishes that the R1234yf/R134a with 69.29% as the highest and the lowest exergy efficiency in the highest compression pressure of the present SORC system belongs to R1234ze(E)/R1234yf/R134a with only 0.94% reduction, compare with R1234yf/R134a is selected. Furthermore, it notes that the enthalpy and entropy changes in the flue gas and cold water have a positive relationship with exergy efficiency.

Considering the thermo-economic efficiency parameters, the SIC increases the cost of each component of the SORC system because of a similar trend to the PPC and in parallel with the required exchanger area and the net power generated by the turbine. As a result, the power consumes by the pump significantly. Moreover, SIC reveals the same results as PBP of the SORC system. For example, the maximum value to minimum value of SIC based on increasing the compression pressure achieves R134a with 5807402.18–22455670.61 $\$.kW^{-1}$ and R1234yf with 16.82–17.38% reduction, respectively. On the other hand, the highest and lowest amount of PPC belongs to the R134a and R1234yf with 81.74–316.07 and 67.99–261.11 $\$.kW^{-1}.year^{-1}$, respectively. To sum up, the PBP shows the same result as SIC and PPC with 306–333 and 302–330 days, which achieve R134a and R1234yf investigated at different inlet turbine pressure.

Further research should focus on developing and performing the present ORC by designing an EES (Engineering Equation Solver) method for greater computational numeric focus to understand the current ORC errors better to conduct in nonfiction.

ACKNOWLEDGEMENT

The authors want to acknowledge the financial support from Universiti Putra Malaysia via Geran Putra Matching (Matching Geran Putra/2019/9300467).

REFERENCES

- Ahmadi, A., El Haj Assad, M., Jamali, D. H., Kumar, R., Li, Z. X., Salameh, T., Al-Shabi, M., & Ehyaei, M. A. (2020). Applications of geothermal Organic Rankine Cycle for electricity production. *Journal of Cleaner Production*, 274, Article 122950. <https://doi.org/10.1016/j.jclepro.2020.122950>
- Alvi, J. Z., Feng, Y., Wang, Q., Imran, M., & Pei, G. (2021a). Effect of phase change materials on the performance of direct vapor generation solar Organic Rankine Cycle system. *Energy Reports*, 223, Article 120006. <https://doi.org/10.1016/j.energy.2021.120006>
- Alvi, J. Z., Feng, Y., Wang, Q., Imran, M., & Pei, G. (2021b). Effect of working fluids on the performance of phase change material storage based direct vapor generation solar Organic Rankine Cycle system. *Energy Reports*, 7, 348-361. <https://doi.org/10.1016/j.egy.2020.12.040>
- Ata, S., Kahraman, A., & Sahin, R. (2020). Prediction and sensitivity analysis under different performance indices of R1234ze ORC with Taguchi's multi-objective optimization. *Case Studies in Thermal Engineering*, 22, Article 100785. <https://doi.org/10.1016/j.csite.2020.100785>
- Bianchi, M., Branchini, L., De Pascale, A., Melino, F., Ottaviano, S., Peretto, A., & Torricelli, N. (2020). Replacement of R134a with low-GWP fluids in a kW-size reciprocating piston expander: Performance prediction and design optimization. *Energy*, 206, Article 118174. <https://doi.org/10.1016/j.energy.2020.118174>
- Braimakis, K., Mikelis, A., Charalampidis, A., & Karellas, S. (2020). Exergetic performance of CO₂ and ultra-low GWP refrigerant mixtures as working fluids in ORC for waste heat recovery. *Energy*, 203, Article 117801. <https://doi.org/10.1016/j.energy.2020.117801>
- Cambi, M., Tascioni, R., Cioccolanti, L., & Bocci, E. (2016). Converting a commercial scroll compressor into an expander: Experimental and analytical performance evaluation. *Energy Procedia*, 129, 363-370. <https://doi.org/10.1016/j.egypro.2017.09.234>
- Chagnon-Lessard, N., Mathieu-Potvin, F., & Gosselin, L. (2020). Optimal design of geothermal power plants: A comparison of single-pressure and dual-pressure Organic Rankine Cycles. *Geothermics*, 86, Article 101787. <https://doi.org/10.1016/j.geothermics.2019.101787>
- Eyerer, S., Dawo, F., Wieland, C., & Spliethoff, H. (2020). Advanced ORC architecture for geothermal combined heat and power generation. *Energy*, 205, Article 117967. <https://doi.org/10.1016/j.energy.2020.117967>
- Guo, H., Gong, M., & Sun, H. (2021). Performance analysis of a novel energy storage system based on the combination of positive and reverse Organic Rankine Cycles. *Energy*, 231, Article 120905. <https://doi.org/10.1016/j.energy.2021.120905>
- Hamid, M. R. A., Yaw, T. C. S., Tohir, M. Z. M., Ghani, W. A. W. A. K., Sutrisna, P. D., & Jeong, H. K. (2021). Zeolitic imidazolate framework membranes for gas separations: Current state-of-the-art, challenges, and opportunities. *Journal of Industrial and Engineering Chemistry*, 98, 17-41.
- Hu, S., Li, J., Yang, F., Yang, Z., & Duan, Y. (2020). Multi-objective optimization of Organic Rankine Cycle using hydrofluorolefins (HFOs) based on different target preferences. *Energy*, 203, Article 117848. <https://doi.org/10.1016/j.energy.2020.117848>

- Imran, M., Haglind, F., Asim, M., & Zeb Alvi, J. (2018). Recent research trends in Organic Rankine Cycle technology: A bibliometric approach. *Renewable and Sustainable Energy Reviews*, *81*, 552-562. <https://doi.org/10.1016/j.rser.2017.08.028>
- Invernizzi, C. M., Iora, P., Preßinger, M., & Manzolini, G. (2016). HFOs as substitute for R-134a as working fluids in ORC power plants: A thermodynamic assessment and thermal stability analysis. *Applied Thermal Engineering*, *103*, 790-797. <https://doi.org/10.1016/j.applthermaleng.2016.04.101>
- Ji, W. T., Xiong, S. M., Chen, L., Zhao, C. Y., & Tao, W. Q. (2021). Effect of subsurface tunnel on the nucleate pool boiling heat transfer of R1234ze(E), R1233zd(E) and R134a. *International Journal of Refrigeration*, *122*, 122-133. <https://doi.org/10.1016/j.ijrefrig.2020.11.002>
- Kang, Z., Zhu, J., Lu, X., Li, T., & Wu, X. (2015). Parametric optimization and performance analysis of zeotropic mixtures for an Organic Rankine Cycle driven by low-medium temperature geothermal fluids. *Applied Thermal Engineering*, *89*, 323-331. <https://doi.org/10.1016/j.applthermaleng.2015.06.024>
- Kavathia, K., & Prajapati, P. (2021). A review on biomass-fired CHP system using fruit and vegetable waste with regenerative Organic Rankine Cycle (RORC). *Materialstoday: Proceedings*, *43*, 572-578. <https://doi.org/10.1016/j.matpr.2020.12.052>
- Li, J., Liu, Q., Ge, Z., Duan, Y., & Yang, Z. (2017). Thermodynamic performance analyses and optimization of subcritical and transcritical Organic Rankine Cycles using R1234ze(E) for 100-200 °C heat sources. *Energy Conversion and Management*, *149*, 140-154. <https://doi.org/10.1016/j.enconman.2017.06.060>
- Li, L., Ge, Y. T., Luo, X., & Tassou, S. A. (2018). Experimental analysis and comparison between CO₂ transcritical power cycles and R245fa Organic Rankine Cycles for low-grade heat power generations. *Applied Thermal Engineering*, *136*, 708-717. <https://doi.org/10.1016/j.applthermaleng.2018.03.058>
- Li, T., Meng, N., Liu, J., Zhu, J., & Kong, X. (2019). Thermodynamic and economic evaluation of the Organic Rankine Cycle (ORC) and two-stage series organic Rankine cycle (TSORC) for flue gas heat recovery. *Energy Conversion and Management*, *183*, 816-829. <https://doi.org/10.1016/j.enconman.2018.12.094>
- Liu, P., Shu, G., Tian, H., Feng, W., Shi, L., & Wang, X. (2020). Experimental study on transcritical Rankine cycle (TRC) using CO₂/R134a mixtures with various composition ratios for waste heat recovery from diesel engines. *Energy Conversion and Management*, *208*, Article 112574. <https://doi.org/10.1016/j.enconman.2020.112574>
- Loni, R., Najafi, G., Bellos, E., Rajaei, F., Said, Z., & Mazlan, M. (2021). A review of industrial waste heat recovery system for power generation with Organic Rankine Cycle: Recent challenges and future outlook. *Journal of Cleaner Production*, *287*, Article 125070. <https://doi.org/10.1016/j.jclepro.2020.125070>
- Lu, P., Luo, X., Wang, J., Chen, J., Liang, Y., Yang, Z., Wang, C., & Chen, Y. (2021). Thermo-economic design, optimization, and evaluation of a novel zeotropic ORC with mixture composition adjustment during operation. *Energy Conversion and Management*, *230*, Article 113771. <https://doi.org/10.1016/j.enconman.2020.113771>
- Manente, G., Lazzaretto, A., & Bonamico, E. (2017). Design guidelines for the choice between single and dual pressure layouts in Organic Rankine Cycle (ORC) systems. *Energy*, *123*, 413-431. <https://doi.org/10.1016/j.energy.2017.01.151>

- Mignard, D. (2014). Correlating the chemical engineering plant cost index with macroeconomic indicators. *Chemical Engineering Research and Design*, 92(2), 285-294. <https://doi.org/10.1016/j.cherd.2013.07.022>
- Molés, F., Navarro-Esbri, J., Peris, B., Mota-Babiloni, A., & Mateu-Royo, C. (2017). R1234yf and R1234ze as alternatives to R134a in Organic Rankine Cycles for low temperature heat sources. *Energy Procedia*, 142, 1192-1198. <https://doi.org/10.1016/j.egypro.2017.12.380>
- Nafey, A. S., & Sharaf, M. A. (2010). Combined solar organic Rankine cycle with reverse osmosis desalination process: energy, exergy, and cost evaluations. *Renewable Energy*, 35(11), 2571-2580. <https://doi.org/10.1016/j.renene.2010.03.034>
- Peris, B., Navarro-Esbri, J., Moles, F., Martí, J. P., & Mota-Babiloni, A. (2015). Experimental characterization of an Organic Rankine Cycle (ORC) for micro-scale CHP applications. *Applied Thermal Engineering*, 79, 1-8. <https://doi.org/10.1016/j.applthermaleng.2015.01.020>
- Ping, X., Yang, F., Zhang, H., Zhang, J., Zhang, W., & Song, G. (2021). Introducing machine learning and hybrid algorithm for prediction and optimization of multistage centrifugal pump in an ORC system. *Energy*, 222, Article 120007. <https://doi.org/10.1016/j.energy.2021.120007>
- Quoilin, S., Declaye, S., Tchanche, B. F., & Lemort, V. (2011). Thermo-economic optimization of waste heat recovery Organic Rankine Cycles. *Applied Thermal Engineering*, 31(14-15), 2885-2893. <https://doi.org/10.1016/j.applthermaleng.2011.05.014>
- Roumpedakis, T. C., Loumpardis, G., Monokrousou, E., Braimakis, K., Charalampidis, A., & Karellas, S. (2020). Exergetic and economic analysis of a solar driven small scale ORC. *Renewable Energy*, 157, 1008-1024. <https://doi.org/10.1016/j.renene.2020.05.016>
- Rowshanaie, O., Mustapha, S., Ahmad, K. A., & Rowshanaie, H. (2015). Simulation of Organic Rankine Cycle through flue gas to large scale electricity generation purpose. *Jurnal Teknologi*, 77(27), 9-18. <https://doi.org/10.11113/jt.v77.6878>
- Rowshanaie, O., Tohir, M. Z. M., Mustapha, F., Ya'acob, M. E., & Rowshanaie, H. (2020). Optimization and analyzing of subcritical Organic Rankine Cycle using R1234ze(E) for low and medium temperature heat source. *IOP Conf. Series: Materials Science and Engineering*, 778, Article 012074. <https://doi.org/10.1088/1757-899X/778/1/012074>
- Schilling, J., Entrup, M., Hopp, M., Gross, J., & Bardow, A. (2021). Towards optimal mixtures of working fluids: Integrated design of processes and mixtures for Organic Rankine Cycles. *Renewable and Sustainable Energy Reviews*, 135, Article 110179. <https://doi.org/10.1016/j.rser.2020.110179>
- Shengjun, Z., Huaixin, W., & Tao, G. (2011). Performance comparison and parametric optimization of subcritical Organic Rankine Cycle (ORC) and transcritical power cycle system for low-temperature geothermal power generation. *Applied Energy*, 88(8), 2740-2754. <https://doi.org/10.1016/j.apenergy.2011.02.034>
- Sun, K., Zhao, T., Wu, S., & Yang, S. (2021). Comprehensive evaluation of concentrated solar collector and Organic Rankine cycle hybrid energy process with considering the effects of different heat transfer fluids. *Energy Reports*, 7, 362-384. <https://doi.org/10.1016/j.egypr.2021.01.004>
- Sun, Q., Lin, D., Khayatnezhad, M., & Taghavi, M. (2021). Investigation of phosphoric acid fuel cell, linear Fresnel solar reflector and Organic Rankine Cycle polygeneration energy system in different climatic

- conditions. *Process Safety and Environmental Protection*, 147, 993-1008. <https://doi.org/10.1016/j.psep.2021.01.035>
- Tian, R., Xu, Y., Shi, L., Song, P., & Wei, M. (2020). Mixed convection heat transfer of supercritical pressure R1234yf in horizontal flow: Comparison study as alternative to R134a in organic Rankine cycles. *Energy*, 205, Article 118061. <https://doi.org/10.1016/j.energy.2020.118061>
- Turton, R., Bailie, R. C., Whiting, W. B., Shaeiwitz, J. A., & Bhattacharyya, D. (2012). *Analysis, synthesis and design of chemical processes* (4th Ed.). Pearson Education.
- Vaupel, Y., Huster, W. R., Mhamdi, A., & Mitsos, A. (2021). Optimal operating policies for Organic Rankine Cycles for waste heat recovery under transient conditions. *Energy*, 224, Article 120126. <https://doi.org/10.1016/j.energy.2021.120126>
- Vélez, F., Segovia, J. J., Martín, M. C., Antolín, G., Chejne, F., & Quijano, A. (2012). A technical, economical and market review of Organic Rankine Cycles for the conversion of low-grade heat for power generation. *Renewable and Sustainable Energy Reviews*, 16(6), 4175-4189. <https://doi.org/10.1016/j.rser.2012.03.022>
- Vera, D., Baccioli, A., Jurado, F., & Desideri, U. (2020). Modeling and optimization of an ocean thermal energy conversion system for remote islands electrification. *Renewable Energy*, 162, 1399-1414. <https://doi.org/10.1016/j.renene.2020.07.074>
- Wang, F., Wang, L., Zhang, H., Xia, L., Miao, H., & Yuan, J. (2021). Design and optimization of hydrogen production by solid oxide electrolyzer with marine engine waste heat recovery and ORC cycle. *Energy Conversion and Management*, 229, Article 113775. <https://doi.org/10.1016/j.enconman.2020.113775>
- Wang, J., Wang, J., Li, J., & Liu, N. (2020). Pressure drop of R134a and R1234ze(E) flow boiling in microchannel arrays with single- and double-side heating. *International Journal of Heat and Mass Transfer*, 161, Article 120241. <https://doi.org/10.1016/j.ijheatmasstransfer.2020.120241>
- Wang, Z., Xia, X., Pan, H., Zuo, Q., Zhou, N., & Xie, B. (2021). Fluid selection and advanced exergy analysis of dual-loop ORC using zeotropic mixture. *Applied Thermal Engineering*, 185, Article 116423. <https://doi.org/10.1016/j.applthermaleng.2020.116423>
- Yang, F., Zhang, H., Song, S., Bei, C., Wang, H., & Wang, E. (2015). Thermoeconomic multi-objective optimization of an Organic Rankine Cycle for exhaust waste heat recovery of a diesel engine. *Energy*, 93, 2208-2228. <https://doi.org/10.1016/j.energy.2015.10.117>
- Yang, M. H. (2018). Payback period investigation of the Organic Rankine Cycle with mixed working fluids to recover waste heat from the exhaust gas of a large marine diesel engine. *Energy Conversion and Management*, 162, 189-202. <https://doi.org/10.1016/j.enconman.2018.02.032>
- Yang, Z., Valtz, A., Coquelet, C., Wu, J., & Lu, J. (2020). Experimental measurement and modelling of vapor-liquid equilibrium for 3,3,3-Trifluoropropene (R1243zf) and trans-1,3,3,3-Tetrafluoropropene (R1234ze(E)) binary system. *International Journal of Refrigeration*, 120, 137-149. <https://doi.org/10.1016/j.ijrefrig.2020.08.016>
- Yu, X., Huang, Y., Li, Z., Huang, R., Ghang, J., & Wang, L. (2021). Characterization analysis of dynamic behavior of basic ORC under fluctuating heat source. *Applied Thermal Engineering*, 189, Article 116695. <https://doi.org/10.1016/j.applthermaleng.2021.116695>

- Zhai, H., An, Q., & Shi, L. (2016). Analysis of the quantitative correlation between the heat source temperature and the critical temperature of the optimal pure working fluid for subcritical Organic Rankine Cycles. *Applied Thermal Engineering*, 99, 383-391. <https://doi.org/10.1016/j.applthermaleng.2016.01.058>
- Zhai, H., An, Q., & Shi, L. (2018). Zeotropic mixture active design method for Organic Rankine Cycle. *Applied Thermal Engineering*, 129, 1171-1180. <https://doi.org/10.1016/j.applthermaleng.2017.10.027>
- Zhang, C., Fu, J., Kang, J., & Fu, W. (2018). Performance optimization of low-temperature geothermal organic Rankine cycles using axial turbine isentropic efficiency correlation. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 40, Article 61. <https://doi.org/10.1007/s40430-018-0996-9>
- Zhang, C., Liu, C., Wang, S., Xu, X., & Li, Q. (2017). Thermo-economic comparison of subcritical organic Rankine cycle based on different heat exchanger configurations. *Energy*, 123, 728-741. <https://doi.org/10.1016/j.energy.2017.01.132>
- Zhang, Y., Lei, B., Masaud, Z., Imran, M., Wu, Y., Liu, J., Qin, X., & Muhammad, H. A. (2020). Waste heat recovery from diesel engine exhaust using a single-screw expander Organic Rankine Cycle system: Experimental investigation of exergy destruction. *Energies*, 13(22), Article 5914. <https://doi.org/10.3390/en13225914>
- Zheng, N., Wei, J., & Zhao, L. (2018). Analysis of a solar Rankine cycle powered refrigerator with zeotropic mixtures. *Solar Energy*, 162, 57-66. <https://doi.org/10.1016/j.solener.2018.01.011>

Optimal Material Selection for Manufacturing Prosthetic Foot

Fahad Mohanad Kadhim^{1*}, Samah Falah Hasan² and Sadiq Emad Sadiq³

¹University of Al-Nahrain, Prosthetic and orthotics Engineering Department, Baghdad, Iraq

²Al Farahidi University, Aeronautical Technical Engineering Department, Baghdad, Iraq

³Al-Furat Al-Awsat Technical University, Department of Aeronautical Technical Engineering, Technical Engineering College of Najaf, Baghdad, Iraq

ABSTRACT

The foot is an essential part of the components of the prosthesis. Therefore, the selected materials' mechanical properties, cost, and weight must be considered when manufacturing the prosthetic foot. This study studied the mechanical properties of selected materials used for prosthetic feet. The material chosen is Carbon Fiber, Glass fiber, and hybrid composite material. This study aims to simulate chosen materials to find the optimal material selection for manufacturing prosthetic feet by assuming boundary conditions, reaction forces, design consideration, and application. The simulation was done by the finite element analysis ANSYS-14.5 program. The result of the force plate test shows the ground reaction force equal to 750N at heel strike, 700N at mid-stance, and 650N at the toe-off stage. The finite element result shows the maximum Von-Misses stress equal to 119MPa at the toe-off stage, and the hybrid composite material has the maximum safety factor. Furthermore, the results showed that the mechanical properties of the hybrid composite materials are the best, as the yield stress is 560MPa, the ultimate stress is 678MPa, and the modulus of elasticity is 6.2GPa. The result shows that the Hybrid composite material has excellent improvement in

mechanical properties such as lightweight, stiffness, high mechanical properties, and cost-efficiency. Hence by considering the body weight of the amputee, gait cycle, and analyzing the material properties, the hybrid composite material is the best suitable should be selected to manufacture foot prostheses.

ARTICLE INFO

Article history:

Received: 14 January 2022

Accepted: 03 March 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.03>

E-mail addresses:

fahad.mohanad@nahrainuniv.edu.iq (Fahad Mohanad Kadhim)

Sama.Falah@alfarahidiuc.edu.iq (Samah Falah Hasan)

sadaq.emad@atu.edu.iq (Sadiq Emad Sadiq)

* Corresponding author

Keywords: Carbon fiber, composite material, foot, hybrid material, prosthetic

INTRODUCTION

A prosthesis is an artificial device used by people who suffer from amputations due to disease or accidents (Kadhim et al., 2020a; Herbert et al., 2005). The lower limb prosthetic parts are the pylon, the socket, and the foot, as shown in Figure 1 (Kadhim et al., 2020b). A prosthetic foot is a manufactured device that replaces a lost limb and restores some function to amputees (Awad & Kadhim, 2022).

In recent years, different prosthetic parts have been manufactured and developed, such as knee parts (Arteaga et al., 2020; Zhang et al., 2021; Kadhim et al., 2020c; Wang et al., 2020) and prosthetic sockets (Estillore et al., 2021; Vitali et al., 2017; Sakuri et al., 2020; Marable et al., 2020, Kadhim et al., 2019; Monette et al., 2020), and hip joint application (Annur et al., 2020; Abdullah et al., 2019; Delikanli & Kayacan, 2019; Gavali et al., 2016). Many researchers were interested in the design and manufacturing of prosthetic feet. Hadi and Oleiwi (2015) investigated the tensile strength of polymer blends as prosthetic foot material reinforcement by carbon fiber. Mohammed and Salman (2020) studied designing and modeling a prosthetic foot made of suitable composite materials (high-density polyethylene (HDPE) filled with 60% Date Palm Wood) (DPW). Tao et al. (2017) used 3D printing of polylactic acid to design and optimize the prosthetic foot. They printed a prosthetic foot using filament made of Polylactic Acid (PLA) and a hobby-level printer. Tryggvason et al. (2020) investigated the Dynamic FEA used to modify the design and analyze the energy of a variable stiffness prosthetic foot. Bence and Dávid (2017) created a 3D-printed Energy Storage and Return (ESAR) foot prosthesis with an intriguing novel shape suitable for people with trans-tibial amputations and employing ABS material as a filament.

Oleiwi and Hadi (2021) Most prosthetic feet currently available are made from polyethylene and polyurethane materials. These feet always have mechanical failure due to fatigue problems in the metatarsal region due to the dorsiflexion and plantarflexion movements, as shown in Figure 2.

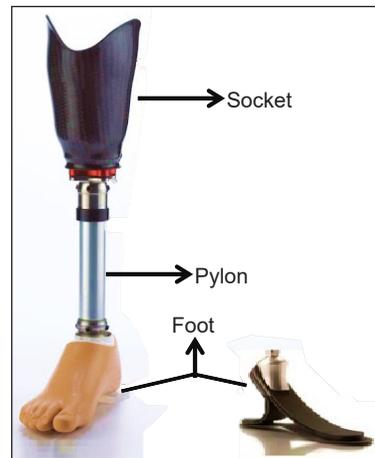


Figure 1. The parts of the lower limb prosthetic (Awad et al., 2022)



Figure 2. Mechanical failure of the traditional prosthetic foot (Yousif et al., 2018)

This study proposed three common materials to manufacture the prosthetic foot. The materials are carbon fiber, glass fiber, and the hybrid composite material that consists of a mixture of carbon fiber and glass fiber in addition to resin. The purpose of suggesting these materials is because all of these materials have high mechanical properties and are suitable for this application. Also, to compare these materials to choose the best material for manufacturing the prosthetic foot. The proposed materials for use in the manufacture of the prosthetic foot were tested for mechanical properties. In addition, the foot prosthetic model has been tested by the Finite element method as a case study to understand better the stress distribution and safety factors and optimize the selection of materials.

MATERIALS AND METHOD

Materials

In this study, three groups of materials will be tested. Each material will be a case study for manufacturing the prosthetic foot. Each material group consists of several layers according to the study cases listed below:

1. Case (1) (twenty layers of carbon fiber with lamina)
2. Case (2) (twenty layers of fiberglass with lamina)
3. Case (3) (ten layers of carbon fiber + ten layers of fiberglass, the arrangement is periodically recurring a layer of carbon fiber, followed by a layer of fiberglass, and then a layer of carbon fiber).

The lamination process produces samples of tensile and fatigue tests to examine the materials to know their mechanical properties. The material needed in lamination is the following (two Polyvinylalcohol PVA bags, layers of each material group, epoxy with hardener with a ratio of 80:30, a vacuum device, and Jepson mold). Figure 3 shows the positioning of the Jepson mold at the stand of the vacuum pressure system.

Put the first piece of the PVA on the gypsum molds and suction the air between the PVA and gypsum with the vacuum system, then put the materials' layers on the plaster mold, and then put the second piece of the PVA and suction the air between two bags with the vacuum system, then mix the overlying resin 80:20 polyurethane with the hardener and put this mixture on these layers. Thus, using a vacuum system, the air and the mixture of the

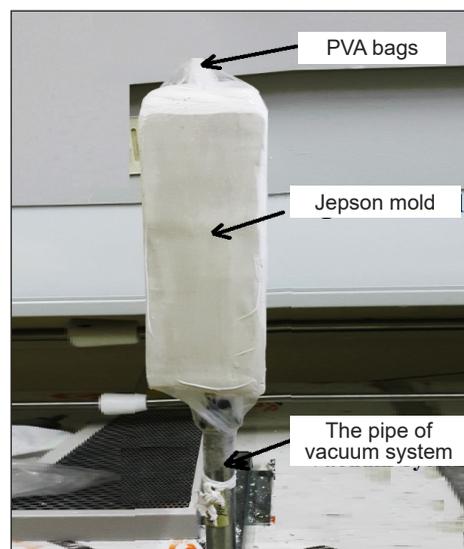


Figure 3. The Jepson mold at the stand of the vacuum pressure to lay up the layers of fibers

lamina are suctioned from the space in which the casting process is made. Finally, maintain the vacuum at constant pressure at room temperature until the laminations are cool, then cut them according to the size of the samples for the mechanical test.

Tensile Test

A CNC machine cuts each casting sample into three samples for tensile testing. The dimensions of the specimens are cut according to ASTM D638 (2014). The thickness of samples varies depending on the layers of each material case, as shown in Figure 4. The samples were tested at a 3 mm/min speed by a Testometric device type. The specimen was placed in the grip of the tensile testing machine, and the test was performed by applying tension until it underwent fracture, as shown in Figure 5. Figure 6 shows the carbon fibers and glass fibers group specimen. The figure of the hybrid sample is similar to the shape of the carbon fiber specimen because its outer surface consists of a layer of carbon fiber.

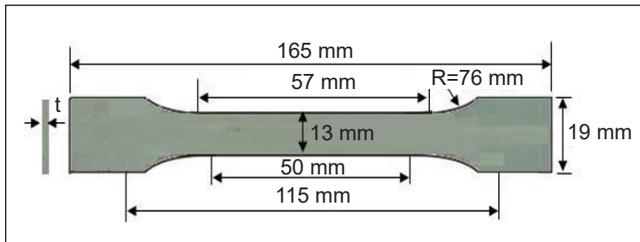
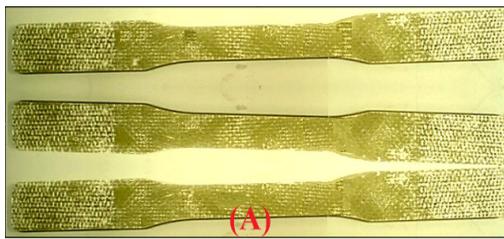


Figure 4. The ASTM D638 dimensions of tensile specimen



Figure 5. The clamped samples with the jig and tensile device



(a)



(b)

Figure 6. (a) The tensile specimen of fiberglass; (b) The tensile specimen of carbon fiber

Fatigue Test

The CNC machine cut eight samples of each casting case to test the fatigue machine (Roberts & Hart, 2001). The dimension of the spearman's cutting is according to the spearman's standard of fatigue device test. The length of the fatigue specimen is 100 mm

and the width 10 mm, while thickness varies with each group of a composite material layout, as shown in Figure 7. Figure 8 shows the fatigue specimen of carbon fibers and glass fibers group without the figure of the hybrid sample due to its similarity to the shape of the carbon fiber specimen.

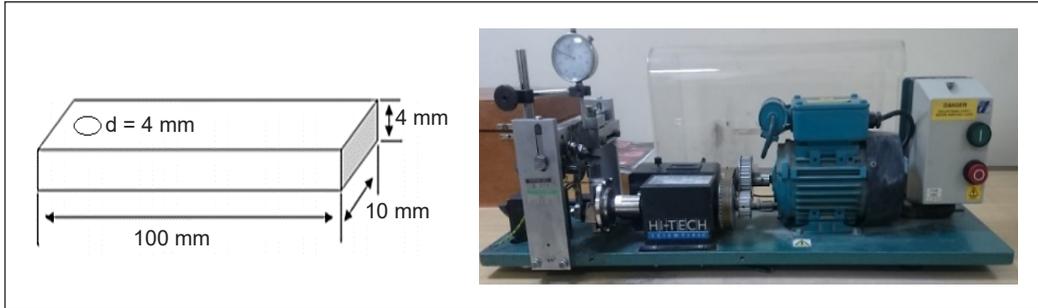


Figure 7. The fatigue test specimen's dimensions with a fatigue machine (HI-TEICH)



Figure 8. The fatigue specimens of carbon fibers and fiberglass

Test of Ground Reaction Force

The ground exerts the ground reaction force on a body in contact with it (Müller & Schiffer, 2020). When the normal or amputee person walks on the ground, the reaction force affecting on the body through the foot, the reaction force consists of three components of forces, two in a horizontal direction and one in a vertical direction of the foot, as shown in Figure 9. In this test, the patient walks normally on the device, and the device records the values of the reaction forces on the body for the steps taken by the patient while passing on the device. This test aims to measure the ground reaction values on the body that will apply as a boundary condition during the simulation for the material's suggested feet. The patient has a transtibial amputee who walked on the force plate device to measure the reaction force that must be applied on foot during finite element analysis, as shown in Figure 10.

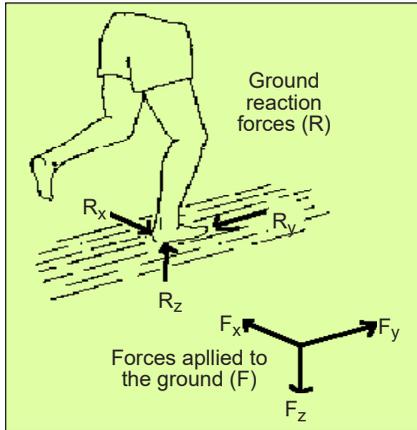


Figure 9. The components of ground reaction force

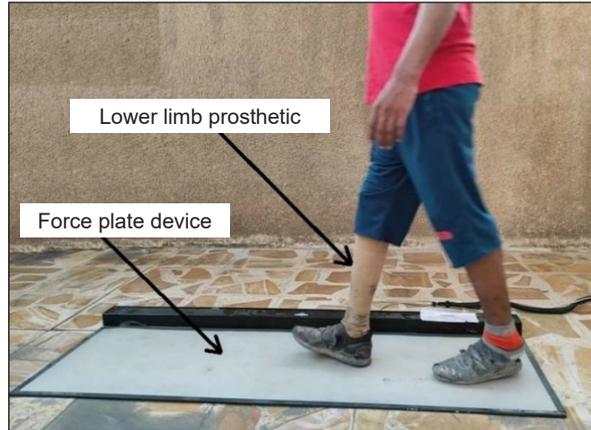


Figure 10. The patient with a transtibial amputee walked on the force plate device

Finite Element Analysis

It was a useful tool for understanding load transfer through prosthetic applications. The finite element technique is a full-field analysis for calculating the state of stresses, safety factors, and other mechanical parameters in a specific field:

1. The Solid-works 2020 software designed the three-dimensional model of the prosthetic foot.
2. The designed model was transferred to ANSYS 14.5 software to analyze a static load test and find the Von-Misses stress and the safety factor when applying the boundary conditions.
3. Meshing was done with tetrahedrons, as shown in Figure 11.

Mesh convergence tests were done previously to determine the best mesh size. A mesh size of 2 mm was chosen for this model based on the convergence analysis.

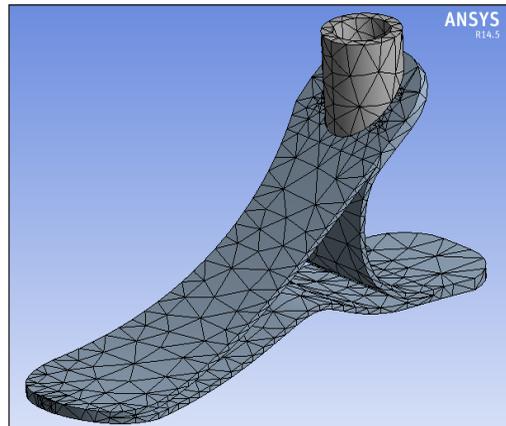


Figure 11. The mesh of the foot prosthetic model

RESULTS AND DISCUSSION

The tensile tests showed the mechanical properties of the three study cases of the materials used to manufacture the prosthetic foot. Figure 12 shows the stress-strain curve of the three tested materials, whose mechanical properties can be summarized in Table 1. The results showed that the hybrid composite material has better mechanical properties than

carbon fiber and fiberglass due to the anisotropic properties in all the hybrids. Failure is generally noncatastrophic when hybrid composites are stressed under tension (i.e., It does not happen suddenly). The carbon fibers fail first when the load is applied to the hybrid composites, and the load is then transmitted to the glass fibers. The matrix phase must maintain the applied load when the glass fibers fail. Therefore, the failure of the composite event coincides with the failure of the matrix phase (Callister, 2007).

The modulus of elasticity for hybrid material equals 6.2 GPa, ultimate stress is 678 MPa, and yield stress is 560 MPa. The fatigue test results showed the stress endurance equal to (425, 300,90) MPa for the hybrid composite material, carbon fiber, and glass fiber, respectively. Glass fiber has low stress compared to others due to the higher strength of carbon fiber. Therefore, it behaves like elastic material with good bearing capacity during tensile loading (Naito & Oguma, 2017; Dong, 2016). Anisotropic properties may be seen in all the hybrids. Failure is generally noncatastrophic when hybrid composites are stressed under tension (i.e., It does not happen suddenly). The carbon fibers fail first when the load is applied to the hybrid composites, and the load is then transmitted to the glass fibers. The matrix phase must maintain the applied load when the glass fibers fail. The composite event's failure coincides with the matrix phase's failure.

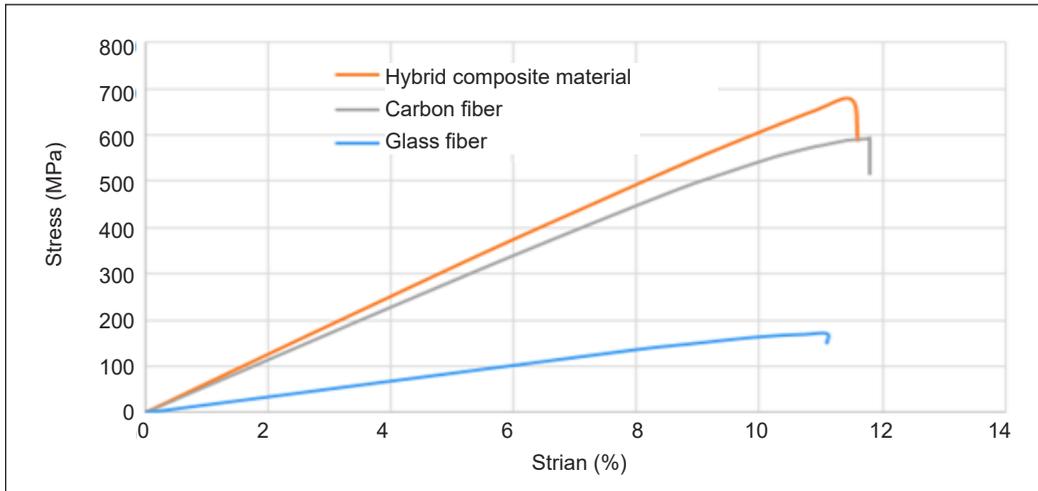


Figure 12. The stress-strain curve of the three tested materials

Table 1
Overview of the tested materials' mechanical properties

Number of cases	Number of layers	E (GPa)	σ_{ult} (MPa)	σ_y (MPa)	Thickness (mm)
Case (1)	20	5.7	580	448	4.1
Case (2)	20	1.7	170	160	4.48
Case (3)	20	6.2	678	560	4.22

Figure 13 shows the relationship between repeated stresses and the number of cycles of fatigue testing for the three tested materials.

Figure 14 shows the values of ground reaction force resulting from the gait cycle test. The ground reaction force curve is divided into three regions (heel strike, mid stance, and toe-off). The reaction force value equals 750 N at heel strike, 700 at midstance, and 650 at toe-off). The values of ground reaction force are used as a boundary condition in the FEM analysis of the foot.

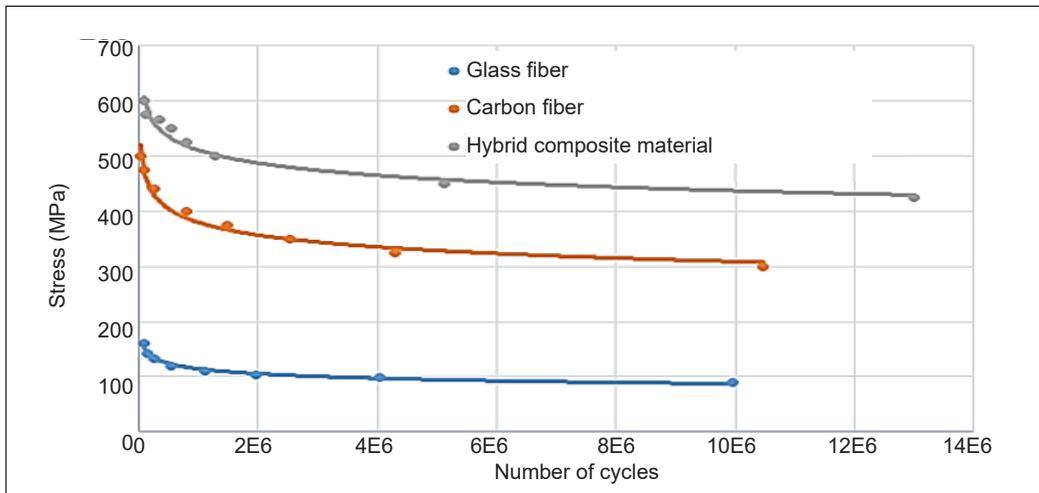


Figure 13. The S-N curve of the three tested materials

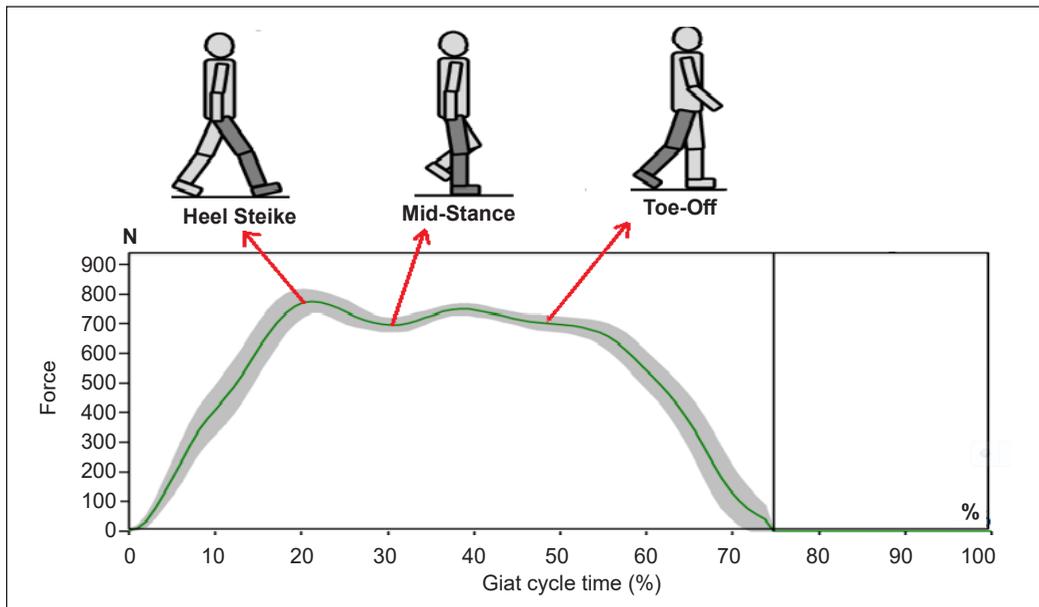


Figure 14. The curve of the ground reaction force of an amputee person

The Boundary Condition for Analysis of the Prosthetic Foot. For the analysis of the prosthetic foot, the boundary conditions must be applied to evaluate the generated stresses and the values of the safety factor for each study case of the materials group for the manufacture of the prosthetic foot. The boundary condition includes applying the value of ground reaction force at shank with fixed the foot at the heel at the initial contact stage of the gait cycle, as shown in Figure 15(a), the region (A) is fixed support and applied load at region (B). Next, for the stage of midstance, the reaction force is applied at the shank and fixed at the forefoot and rearfoot as shown in Figure 15(b), the regions (B) are fixed support and applied load at region (A). Finally, the boundary condition at the toe-off stage, the reaction force applied at the shank and fixed at the forefoot as shown in Figure 15(c), the region (B) is fixed support and applied load at region (A). The applied boundary conditions for simulating a prosthetic foot derived from the applied forces resulting from bodyweight on foot during gait cycle analyses (Baker, 2013; Levine et al., 2012).

The result of numerical analysis of prosthetic foot shows the stress generated due to applying the boundary condition. For example, Figure 16(a) shows the maximum stress value generated equal to 103.64 MPa when the patient walks at the heel strike. At the same time, Figure 16(b) shows the maximum stress equal to 65 MPa when applying the

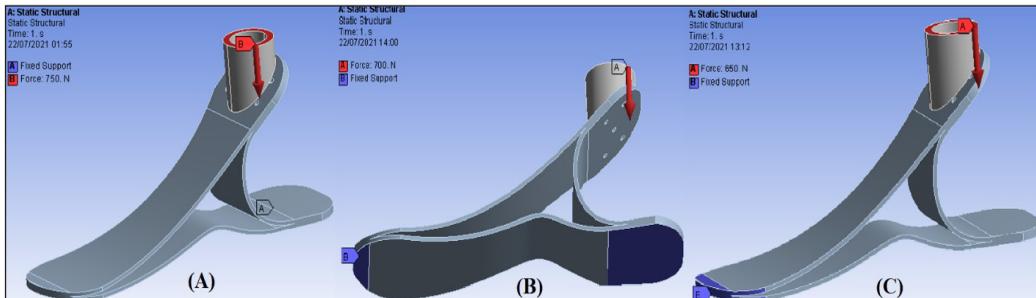


Figure 15. (a) The prosthetic foot model's boundary condition at heel strike position; (b) The prosthetic foot model's boundary condition at midstance position; (c) The prosthetic foot model's boundary condition at the toe-off position

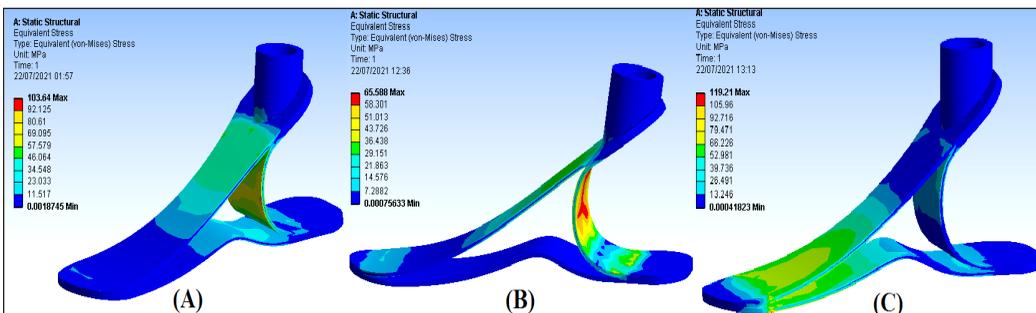


Figure 16. (a) The foot model's Von Mises stress at the heel strike stage; (b) The foot model's Von Mises stress at the midstance stage; (c) The foot model's Von Mises stress at the toe-off stage

boundary condition at the midstance phase. Therefore, maximum stress equals 119.2 MPa for toe-off boundary conditions, as shown in Figure 16(c). It was noted from the stress analysis results that there is a large difference between the maximum stress generated and the yield stress for each material, which explains the possibility of the success of the three groups of materials in manufacturing the prosthetic foot.

Also, the analysis shows that the safety factor for the prosthetic foot's material groups has been passed in design. The safety factor's value varies by region, based on the distribution of stresses generated and the endurance stress for each material group. Figures 17 to 19 show the foot model's safety factor when using carbon fiber, fiberglass, and hybrid composite material in three gait cycle phases (heel strike, midstance, toe-off). The results showed that the safety factor values were equal to more than 1.25 for all groups of materials and at all stages of the gait cycle. Therefore, the design will be safe if the fatigue safety factor is greater than or equal to 1.25 (Miller, 2002). The values of safety factors of three materials at three stages of the gait cycle can be summarized in Table 2.

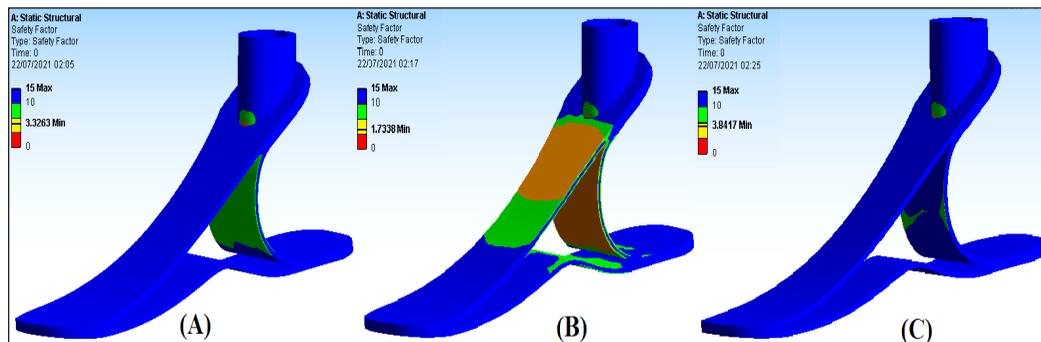


Figure 17. (a) The safety factor value for carbon fiber reinforced foot model at the heel strike stage; (b) The safety factor value for fiberglass reinforced foot model at the heel strike stage; (c) The safety factor value for hybrid material reinforced foot model at the heel strike stage

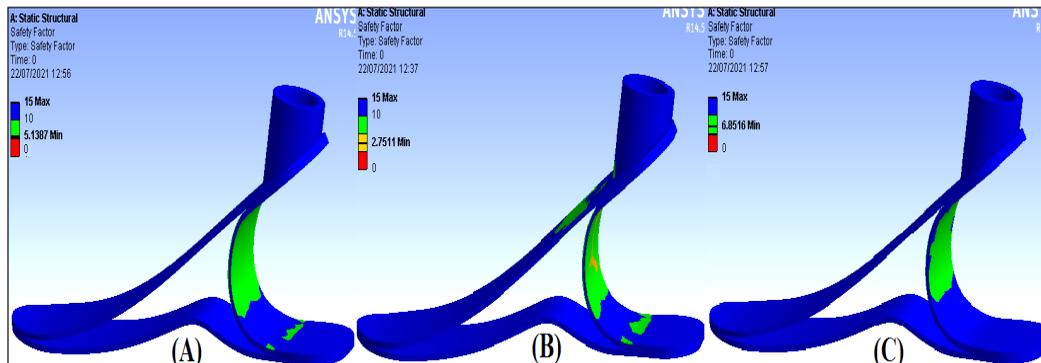


Figure 18. (a) The safety factor value for carbon fiber reinforced foot model at the heel strike stage; (b) The safety factor value for fiberglass reinforced foot model at the heel strike stage; (c) The safety factor value for hybrid material reinforced foot model at the midstance stage

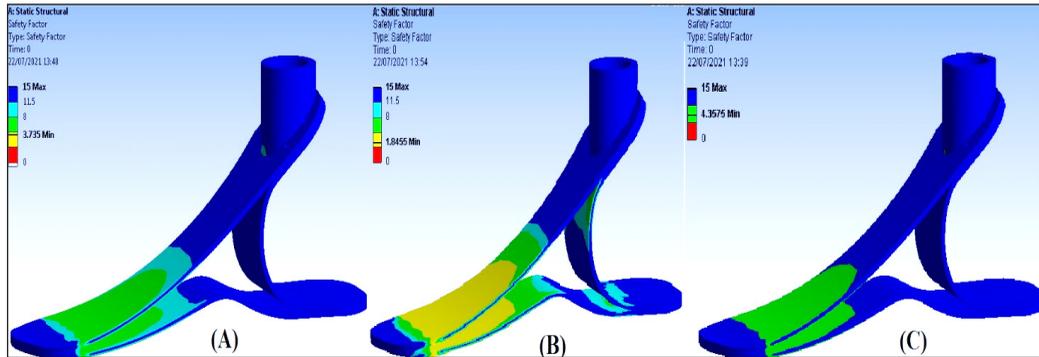


Figure 19. (a) The safety factor value for carbon fiber reinforced foot model at the heel strike stage; (b) The safety factor value for fiberglass reinforced foot model at the heel strike stage; (c) The safety factor value for hybrid material reinforced foot model at the toe-off stage

Table 2

The safety factors summarized the values of three materials at different stages of the gait cycle

Material	Heel strick	Mid-stance	Toe-off
Carbon fiber	5.7	580	448
Glass fiber	1.7	170	160
Hybrid composite material	6.2	678	560

CONCLUSION

The previous results concluded that all the materials groups passed in design, but the best material to manufacture the prosthetic foot is the material in case (3), which is hybrids composite material (carbon and glass fibers). For manufacturing the prosthetic foot, it was decided to use a hybrid composite material because it has a better overall combination of properties than composites with only one fiber type. Various fiber combinations and matrix materials are used, but combining carbon and glass fibers in polymeric resin is the most common method. Carbon fibers are strong and stiff in comparison to other materials. They provide low-density reinforcement but are expensive. Glass fibers are less expensive than carbon fibers but lack their stiffness. The glass fiber–carbon fiber hybrid is stronger and harder, has greater impact resistance, and can be manufactured cheaper than similar all-carbon or all-glass reinforced polymers. Therefore, it was concluded that the hybrid material is the most suitable choice for manufacturing the prosthetic foot because of its high mechanical properties and lightweight compared to the prosthetic foot was made of carbon fiber or fiberglass.

ACKNOWLEDGEMENT

The authors appreciate the cooperation of the Al-Nahrain University, Iraq, in completing the experimental work.

REFERENCES

- Abdullah, T. Ş., Murat, Ç., Irfan, K., & Fehim, F. (2019). Optimal material selection for total hip implant: A finite element case study. *Arabian Journal for Science and Engineering*, 44, 10293-10301.
- Annur, D., Utomo, M. S., Asmaria, T., Malau, D. P., Supriadi, S., Suharno, B., Rahyussalim, A. J., Prabowo, Y., & Amal, M. I. (2020). Material selection based on finite element method in customized iliac implant. *Materials Science Forum*, 1000, 82-89. <https://doi.org/10.4028/www.scientific.net/msf.1000.82>
- Arteaga, O., Terán, H. C., Morales, H., Argüello, E., Erazo, M. I., Ortiz, M., & Morales, J. J. (2020). Design of human knee smart prosthesis with active torque control. *International Journal of Mechanical Engineering and Robotics Research*, 9(3), 347-352. <https://doi.org/10.18178/ijmerr.9.3.347-352>
- ASTM D638. (2014). *Standard test method for tensile properties of plastics*. American Society of Testing and Materials. West Conshohocken. <https://www.astm.org/d0638-14.html>
- Awad, F. S., & Kadhim, F. M. (2022). Compare EMG signals by using myo-ware muscle sensor and myo-trace device for measuring the electrical activity of the muscles. In *AIP Conference Proceedings* (Vol. 2386, No. 1, p. 040009). AIP Publishing LLC. <https://doi.org/10.1063/5.0067239>
- Awad, F. S., Kadhim, F. M., & Aboud, S. W. (2022). Strain and deformation measurement for prosthetic parts using the Arduino microcontroller and strain gauges instruments. *International Journal of Mechanical Engineering*, 7(1), 1049-1055.
- Baker, R. W. (2013). *Measuring Walking: A Handbook of Clinical Gait Analysis*. Wiley.
- Bence, R., & Dávid, P. (2017). Design and analysis of 3D printable foot prosthesis. *Periodical Polytechnic Mechanical Engineering*, 61(4), 282-287. <https://doi.org/10.3311/PPme.11085>
- Callister, W. D. (2007). *Materials Science and Engineering - An Introduction*. John Wiley & Sons. Inc.
- Delikanli, Y. E., & Kayacan, M. C. (2019). Design, manufacture, and fatigue analysis of lightweight hip implants. *Applied Biomaterials & Functional Materials*, 17(2), <https://doi.org/10.1177/2280800019836830>
- Dong, C. (2016). Uncertainties in flexural strength of carbon/glass fibre reinforced hybrid epoxy composites. *Composites Part B: Engineering*, 98, 176-181. <https://doi.org/10.1016/j.compositesb.2016.05.035>
- Estillore, J. V., Dungo, C. A., Guzman, K. N., Maniaul, J. M., & Magdaluyo Jr, E. (2021). Optimal material selection study of prosthetic socket and pylon tube in transtibial prosthesis fabrication. *Engineering Research Express*, 3(2), Article 025030. <https://doi.org/10.1088/2631-8695/ac0094>
- Gavali, S., L., Gawande, S. H., Patil, S. R., & Yerrawar, R. N. (2016). Experimental stress analysis of hip joint implant for fracture analysis. *IOSR Journal of Mechanical & Civil Engineering*, 1, 20-31.
- Hadi, A. N., & Olewi, J. K. (2015). Improving tensile strength of polymer blends as prosthetic foot material reinforcement by carbon fiber. *Journal of Material Science & Engineering*, 4(2), 2169-0022.
- Herbert, N., Simpson, D., Spence, W. D., & Ion, W. (2005). A preliminary investigation into the development of 3-D printing of prosthetic sockets. *Journal of Rehabilitation Research & Development*, 42(2), 141-146. <https://doi.org/10.1682/jrrd.2004.08.0134>
- Kadhim, F. M., Takhakh, A. M., & Chiad, J. S. (2020a). Modeling and evaluation of smart economic transfemoral prosthetic. *Defect and Diffusion Forum*, 398, 48-53. <https://doi.org/10.4028/www.scientific.net/ddf.398.48>

- Kadhim, F. M., Chiad, J. S., & Enad, M. A. S. (2020b). Evaluation and analysis of different types of prosthetic knee joint used by above knee amputee. *Defect and Diffusion Forum*, 398, 34-40. <https://doi.org/10.4028/www.scientific.net/ddf.39>
- Kadhim, F. M., Chiad, J. S., & Enad, M. A. S. (2020c). Evaluation and analysis of different types of prosthetic knee joint used by above knee amputee. *Defect and Diffusion Forum Journal*, 398, 34-40. <https://doi.org/10.4028/www.scientific.net/DDF.398.34>
- Kadhim, F. M., Takhakh, A. M., & Abdullah, M. A. (2019d). Mechanical properties of polymer with different reinforcement material composite that used for fabricates prosthetic socket. *Journal of Mechanical Engineering Research & Developments*, 42(4), 118-123. <https://doi.org/10.26480/jmerd.04.2019.118.123>
- Levine, D., Richards, J., & Whittle, M. W. (2012). *Whittle's gait analysis*. Elsevier health sciences.
- Marable, W. R., Smith, C., Sigurjónsson, B. P., Atlason, I. F., & Johannesson, G. A. (2020). Transfemoral socket fabrication method using direct casting: outcomes regarding patient satisfaction with device and services. *Canadian Prosthetics & Orthotics Journal*, 3(2), Article 6. <https://doi.org/10.33137/cpoj.v3i2.34672>
- Miller, B. A. (2002). Failure analysis and prevention, fatigue failures. *ASM International Handbook*, 11, Article 1470.
- Mohammed, H. S., & Salman, J. M. (2020). Design and modeling the prosthetic foot from suitable composite materials. *American Journal of Engineering and Applied Sciences*, 13(3), 516-522. <https://doi.org/10.3844/ajeassp.2020.516.522>
- Monette, D., Dumond, P., Chikhaoui, I., Nichols, P., & Lemaire, E. D. (2020). Preliminary material evaluation of flax fibers for prosthetic socket fabrication. *Journal of Biomechanical Engineering*, 143(2), Article 021006. <https://doi.org/10.1115/1.4048079>
- Müller, P., & Schiffer, Á. (2020). Human gait cycle analysis using kinect V2 sensor. *Pollack Periodica*, 15(3), 3-14. <https://doi.org/10.1556/606.2020.15.3.1>
- Naito, K., & Oguma, H. (2017). Tensile properties of novel carbon/glass hybrid thermoplastic composite rods. *Composite Structures*, 161, 23-31. <https://doi.org/10.1016/j.compstruct.2016.11.042>
- Olewi, J. K., & Hadi, A. N. (2021). Properties of materials and models of prosthetic feet: A review. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1094, No. 1, p. 012151). IOP Publishing. <https://doi.org/10.1088/1757-899X/1094/1/012151>
- Roberts, N. P., & Hart, N. R. (2001). Alternating bending fatigue machine (HSM20), instruction manual. *Hi-Tech Ltd. UK*, 150, 200-250.
- Sakuri, S., Surojo, E., Ariawan, D., & Prabowo, A. R. (2020). Investigation of Agave cantala-based composite fibers as prosthetic socket materials accounting for a variety of alkali and microcrystalline cellulose treatments. *Theoretical and Applied Mechanics Letters*, 10(6), 405-411. <https://doi.org/10.1016/j.taml.2020.01.052>
- Tao, Z., Ahn, H. J., Lian, C., Lee, K. H., & Lee, C. H. (2017). Design and optimization of prosthetic foot by using polylactic acid 3D printing. *Journal of Mechanical Science and Technology*, 31(5), 2393-2398. <https://doi.org/10.1007/s12206-017-0436-2>

- Tryggvason, H., Starker, F., Lecomte, C., & Jonsdottir, F. (2020). Use of dynamic FEA for design modification and energy analysis of a variable stiffness prosthetic foot. *Applied Sciences*, *10*(2), Article 650. <https://doi.org/10.3390/app10020650>
- Vitali, A., Regazzoni, D., Rizzi, C., & Colombo, G. (2017). Design and additive manufacturing of lower limb prosthetic socket. In *ASME International Mechanical Engineering Congress and Exposition* (Vol. 58462, p. V011T15A021). American Society of Mechanical Engineers. <https://doi.org/10.1115/imece2017-71494>
- Wang, X., Meng, Q., Zhang, Z., Sun, J., Yang, J., & Yu, H. (2020). Design and evaluation of a hybrid passive-active knee prosthesis on energy consumption. *Mechanical Sciences*, *11*(2), 425-436. <https://doi.org/10.5194/ms-11-425-2020>
- Yousif, L. E., Resan, K. K., & Fenjan, R. M. (2018). Temperature effect on mechanical characteristics of a new design prosthetic foot. *International Journal of Mechanical Engineering and Technology*, *9*(13), 1431-1447.
- Zhang, Z., Yu, H., Cao, W., Wang, X., Meng, Q., & Chen, C. (2021). Design of a semi-active prosthetic knee for transfemoral amputees: Gait symmetry research by simulation. *Applied Sciences*, *11*(12), Article 5328. <https://doi.org/10.3390/app11125328>

Development of Regression Models for the Prediction of NiTi Archwire Forces in an Orthodontic Bracket System Using Response Surface Methodology

Sareh Aiman Hilmi Abu Seman, Ching Wei Ng, Muhammad Fauzinizam Razali* and Abdus Samad Mahmud

School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 USM, Nibong Tebal, Pulau Pinang, Malaysia

ABSTRACT

Superelastic NiTi archwire is extensively employed in the early stage of orthodontic therapy due to its capacity to transmit constant and light force to the tooth. The archwire force prediction for orthodontic treatment planning becomes a challenging process as the generated friction at the wire-bracket interface modifies the force exerted by the NiTi archwire. If plotted, the typical force plateau behavior of the superelastic NiTi archwire now gives way to a slope. This study established regression models for estimating the magnitude of forces released by NiTi archwire when bent at various settings in an orthodontic bracket system. Four bending settings parameters were considered: archwire geometry, inter bracket distance, the magnitude of archwire deflection, and testing temperature. The relationships between the settings and the wire forces were investigated using a response surface methodology approach based on data obtained from bending simulation. The magnitude and slope of the unloading force of superelastic NiTi archwire decrease gradually as the amount of wire deflection and inter-bracket distance increase, respectively. NiTi archwires with a diameter of 0.016 inches are best used in the early stages of orthodontic therapy because of their lower unloading forces of 0.57 N to 1.71 N and lower force slope values of 0.13 N/mm to 0.72 N/mm. The developed regression models have strong R-squared values

for the loading force, unloading force, and force slope of 0.996, 0.9830, and 0.9748, respectively, and may thus be used to aid the orthodontist in forecasting the amount of force being exerted on a tooth in various malocclusion conditions.

Keywords: Bending, NiTi archwire, orthodontic bracket, response surface methodology, unloading force

ARTICLE INFO

Article history:

Received: 21 January 2022

Accepted: 29 March 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.04>

E-mail addresses:

sarehaiman@usm.my (Sareh Aiman Hilmi Abu Seman)

chingwei90@hotmail.com (Ching Wei Ng)

mefauzinizam@usm.my (Muhammad Fauzinizam Razali)

abdus@usm.my (Abdus Samad Mahmud)

* Corresponding author

INTRODUCTION

Fixed appliance therapy is one of the orthodontic options available to treat malocclusions, as it encourages correct tooth alignment and overall facial esthetics (Papageorgiou et al., 2017). The fixed appliance is installed by attaching an orthodontic bracket on every tooth with adhesive, followed by carefully inserting an archwire into the bracket slot. This insertion procedure usually induces localized bending along the wire length due to the bracket's irregular position as a result of the tooth's malocclusion. The archwire is then secured inside the bracket slot using elastomeric ligatures, fine wires, or a metal door, depending on the bracket's ligation type considered for the treatment. As the orthodontic archwire attempts to restore its original straight form during the therapy, the malposed tooth is gradually pushed upwards or downwards, depending on the direction of bending recovery of the archwire.

In order to accurately anticipate the tooth movement, the magnitude of the force exerted by the bent wire on the tooth must be within the suggested range. Many variables that influence tooth movement, such as the remodeling of tissues and bone concerning the applied orthodontic pressure, are uncontrollable. Contrarily, the force being applied to the tooth is a controlled variable, and sufficient knowledge of the physics behind the force delivery mechanism may aid in minimizing undesired tooth movement. The recommended orthodontic forces often suggested, ranging from 0.10 N to 1.20 N (Wu et al., 2018; Theodorou et al., 2019), are regarded as the optimum force levels for a speedier and more comfortable treatment experience.

Nickel-titanium (NiTi) orthodontic archwire is commonly selected in orthodontic treatment as it can deliver light and continuous force during bending, which promotes tooth movement. However, it is reported that in orthodontic bracket configurations, friction modifies the constant force behavior of superelastic NiTi archwire to a slope (Nucera et al., 2014; Razali et al., 2018). The friction intensity encountered by NiTi archwire as it slides within the bracket slot could also be affected by the distance set between brackets, a vertical discrepancy of bracket placement, oral temperature, wire sizes, and ligating type used to secure the archwire (Kusy & Whitley, 2000; Higa et al., 2017). While the friction component is unavoidable, the bending settings are adjustable and could be controlled to minimize the deviation of the wire force from the optimal force levels. The work aimed to establish regression models for anticipating the forces produced by a NiTi archwire when bent in various orthodontic bracket systems. The magnitude of the archwire force was determined at different bending settings using a three-dimensional finite-element model of wire bending in a three-bracket configuration. The bending settings considered were the archwire geometry, the inter bracket distance, the magnitude of archwire deflection, and the testing temperature. The wire force regression models were built using the response surface methodology approach. In the future, this regression model may enable the orthodontist to

anticipate the NiTi archwire force early in the treatment process, regardless of the patient's malocclusion state.

MATERIALS AND METHODS

The force-deflection data of the NiTi archwires in an orthodontic bracket configuration were determined using the three-dimensional finite element model developed in our earlier work (Razali et al., 2018). The model considered the bending of a single superelastic NiTi wire in a three-bracket configuration and has been validated against experimental work. The model was developed using Abaqus 6.12.2 software. A user material subroutine (UMAT/Nitinol) developed by Auricchio and Taylor (1997) was employed to assign the superelastic behavior of NiTi alloy on the archwire model.

The half-cut view of the developed three-dimensional finite element model is shown in Figure 1. A bilinear rigid quadrilateral element (R3D4) was used to create three orthodontic brackets with a slot dimension of 0.46 mm (height), 2.8 mm (length), and 0.63 mm (width). The bracket slot's dimensions were determined directly from scanning electron microscopy images. A total of 6,496 elements were employed to model each bracket instance. These brackets were symmetrically positioned by spacing their midpoints apart by 7.5 mm. The middle bracket was displaced downward by 0.06 mm compared to the adjacent bracket. Therefore, it is critical to remove the space between the archwire and the middle bracket's top surface prior to initiating the bending. On the other hand, two NiTi archwire of 0.016 x 0.022-inch and 0.016-inch size with a length of 30 mm was considered in this study. Mesh was applied to the wire model using eight-node linear brick elements with reduced integration (C3D8R). For the round and rectangular wire instances, 72,000 and 72,144 elements were utilized, respectively.

The bending simulation of the superelastic NiTi wire in the bracket system is divided into two stages: loading and unloading. These steps characterize the archwire's

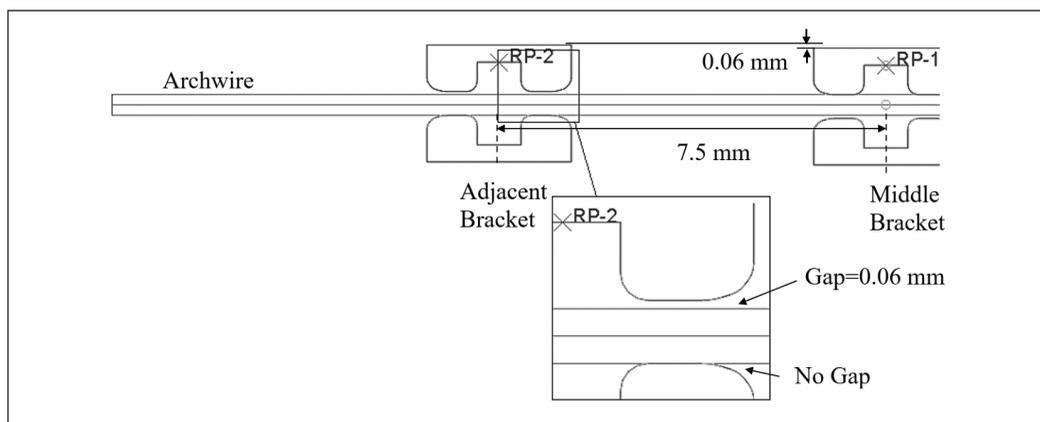


Figure 1. Positioning of archwire and brackets during the assembly stage

deformation and recovery in terms of the application and removal of the bending load, which is accomplished by downwards, and upwards displacement of the middle bracket at a displacement rate of 0.016 mm/s. The initial time increment was set to 0.01 seconds for both steps, with the minimum and maximum permissible time increments set to 0.00001 and 0.1 seconds, respectively. Only the y-axis of the middle bracket was permitted to move, while the other degree of freedom was set to zero. In addition, all degrees of freedom of the adjacent brackets were set to zero to prevent movement throughout the bending course. No boundary conditions were assigned to the wire instance, allowing the wire to move while bending freely. The oral temperature of 36°C was used as the constant for the bending environment.

This study defined the contact behavior between the wire and bracket surfaces using surface-to-surface discretization. The bracket’s surface was designated the master surface, while the wire’s surface was designated the slave surface. In this master-slave method, every slave node search for the nearest point on the master surface, and the contact direction is always normal to the master surface. The surface-to-surface discretization between the archwire and bracket surfaces was described using two interaction properties: normal and tangential behavior. The coefficient of friction was set at 0.27 between the NiTi archwire and the stainless-steel bracket surfaces (Thalman, 2008).

The archwire’s force-deflection behavior during bending was determined using the vertical response force (RF2) and displacement (U2) acquired from the middle bracket reference point (RP-1). As shown in Figure 2, three force criteria were obtained from the force-deflection curve to create the force regression model. First, the loading force

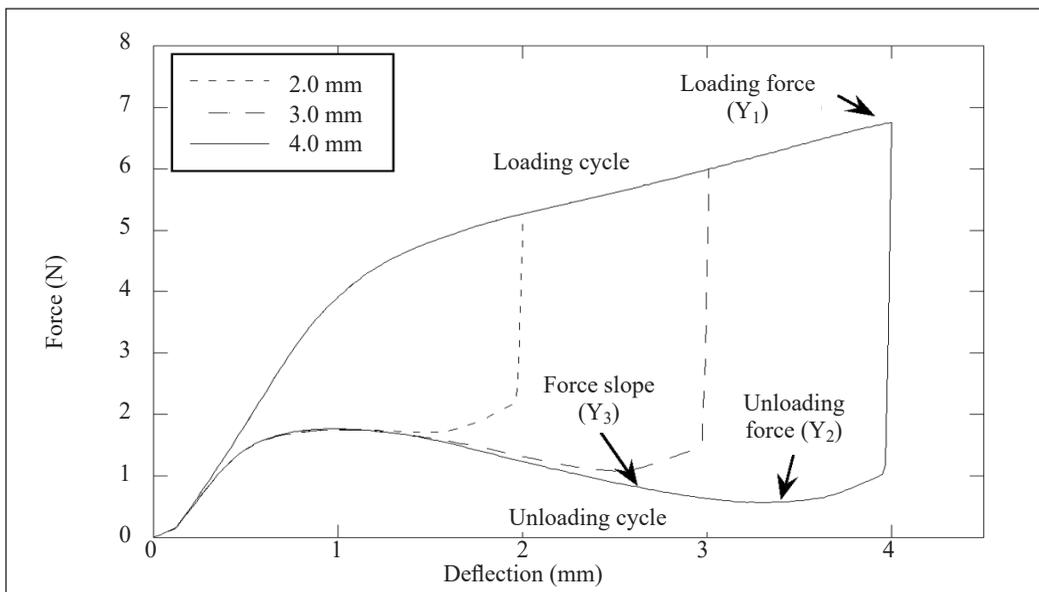


Figure 2. Force-deflection behavior of the NiTi archwire bent at different deflection magnitudes

is determined at the maximum deflection applied on the wire. This response denotes the force the patient experienced upon the NiTi wire installation on the bracket. First, the unloading force is determined when the bracket was deactivated by 0.8 mm from the loading deflection, and this response shows the force transmitted from the archwire to the malposed tooth with the least magnitude. Meanwhile, the slope was calculated using the unloading curve's best linear section. This parameter signifies the rate of changes recorded in the force magnitude as the wire recovered during the unloading cycle.

A statistical design of experiments (DOE) was created using the central composite design (CCD), a tool for designing response surface models. In summary, the design utilizes three types of design points: two-level factorial (2^k), axial points (2^k), and center points. The central composite design method is used to fit and estimate the coefficients of a second-order equation. Equation 1 illustrates the standard form of this second-order equation model with two factors.

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} X_i X_j + \varepsilon \quad [1]$$

where Y is the response of interest, β_0 is a constant coefficient, β_i , β_{ii} , and β_{ij} are the interaction coefficients of the linear, the quadratic, and the second-order terms, respectively. X_i and X_j are the factors, k is the number of studied factors, and e is the random error.

As detailed in Table 1, each of the statistical model's bending settings (factors) has three levels, denoted mathematically by minus one (-1), zero (0), and plus one (+1). The inter-bracket distance, wire deflection, and testing temperature varied between 7.0 mm and 8.0 mm, 2.0 mm and 4.0 mm, and 26°C to 46°C, respectively. These factors were taken into account in this study to account for the variance in wire-bracket placement and oral condition found between patients, and the ranges for each factor were drawn from the literature (Nucera et al., 2014; Franchi et al., 2009; Elayyan et al., 2010; Badawi et al., 2009; Arreghini et al., 2016; Alavi & Hosseini, 2012; Wilkinson et al., 2002). The amount

Table 1
Actual and coded values for each central composite design factor

Numeric Factor (symbol)	Coded Value		
	-1	0	1
	Actual Value		
Inter-bracket distance, mm (<i>A</i>)	7	7.5	8
Wire deflection, mm (<i>B</i>)	2	3	4
Testing temperature, °C (<i>C</i>)	26	36	46
Categoric Factor (symbol)	Coded Value		
	Round	Rectangular	
Geometry (<i>D</i>)	{1}	{-1}	

of the wire deflection of 1.0 mm was omitted from this investigation due to its linear force-deflection characteristic, which is typical in common alloys (Thalman, 2008).

The design included two categorical factors: round and rectangular wire geometry. A total of forty simulation runs were carried out with reference to the following equation $CCD = (2^k + 2k + R) * L$, where k is the number of factors, R is the number of replications at the design center, and L is the level of the categorical factors. Each simulation took fifteen hours to finish on a 2.67GHz CPU with a 24-GB memory.

RESULTS

The magnitudes of the forces obtained from the simulation of archwire bending at different configurations are summarized in Table 2, which lists the possible combination of bending conditions designed by the central composite design (CCD). As can be observed, the loading and unloading forces ranged between 4.08 N and 18.20 N and 0 N and 4.72 N, respectively. On the other hand, the force slope varied between 0.00 N/mm and 2.72 N/mm. Notably, run 30 is the most critical bending condition (combination of 7.0 mm inter-bracket distance, 4.0 mm wire deflection, and 46°C testing temperature) since it has the highest loading force magnitude of 18.20 N. Additionally, it is noteworthy that the rectangular wire's unloading force was zero for running 4, where the wire was deflected to 4.0 mm in 26°C environments with the brackets distanced at 7.0 mm in between. The zero-newton force demonstrated that the wire's unloading force had weakened to zero as it slid along the bracket slot in a high friction condition. However, this zero-newton force was not recorded when the wire's temperature reached 46°C. It is due to the superelastic NiTi wire strengthening at greater temperatures, where the enhanced unloading force aids in overcoming the sliding friction.

The regression models for the archwire force prediction were generated using Design Expert (Version 7.0, Trial Version) software. The best-fitting models, with probability values (Prob>F) of less than 0.0001, were obtained by selecting a quadratic model for the loading force (Y_1), the unloading force (Y_2), and the force slope (Y_3). The final regression models, in terms of the coded factors, are as in Equations 2-4:

$$Y_1 = 9.59 - 1.10A + 1.21B + 1.17C - 3.57D - 0.11AB - 0.056AC + 0.42AD + 0.067BC - 0.48BD - 0.42CD \quad [2]$$

$$Y_2 = 1.63 + 0.050A - 1.03B + 0.37C - 0.56D + 0.28AB - 0.22BC + 0.41BD - 0.13CD \quad [3]$$

$$Y_3 = 1.00 - 0.31A - 0.50B + 0.19C - 0.36D - 0.083AB - 0.063AC + 0.12AD - 0.18BD - 0.069CD - 0.21B^2 \quad [4]$$

where A, B, C, and D are the factors of Table 1.

Table 2
 Summary of the force data obtained from the simulated force-deflection curves

Run	Factor (Coded)				Response (Y)					
	Interbracket Distance (A)	Wire Deflection (B)	Temperature (C)	Wire Geometry (D)	Simulation		Prediction			
					Loading Force, (N)	Unloading Force, (N)	Force Slope, (N/ mm)	Loading Force, (N)	Unloading Force, (N)	Force Slope, (N/ mm)
1	1	1	1	{1}	6.82	0.73	0.69	6.78	0.80	0.61
2	1	-1	1	{1}	5.33	1.91	0.00	5.34	1.84	0.02
3	1	0	0	{-1}	11.73	2.17	1.07	11.75	2.29	0.89
4	-1	1	-1	{-1}	14.77	0.00	2.19	14.79	-0.11	1.84
5	0	-1	0	{-1}	11.30	3.61	0.44	11.40	3.65	0.37
6	1	-1	-1	{-1}	8.77	2.55	0.04	8.68	2.74	0.05
7	-1	1	-1	{1}	6.67	0.13	0.86	6.65	0.07	0.80
8	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
9	0	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.36
10	0	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.36
11	1	-1	-1	{1}	4.08	1.20	0.00	4.08	0.98	0.00
12	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
13	-1	0	0	{-1}	14.93	2.36	1.66	14.91	2.04	1.94
14	0.	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.36
15	0	1	0	{1}	6.75	0.57	0.68	6.73	0.45	0.75
16	1	1	-1	{-1}	11.49	0.87	0.99	11.58	0.79	1.09
17	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
18	-1	-1	1	{1}	6.71	2.23	0.47	6.67	2.44	0.40
19	0	1	0	{-1}	14.88	1.11	1.48	14.83	0.69	1.77
20	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
21	1	-1	1	{-1}	11.55	4.11	0.15	11.53	4.20	0.19
22	0	-1	0	{1}	5.28	1.71	0.13	5.26	1.69	0.09
23	0	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.36
24	1	0	0	{1}	5.35	1.06	0.47	5.37	1.09	0.38
25	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
26	0	0	1	{-1}	14.72	2.70	1.71	14.69	2.72	1.68
27	-1	-1	-1	{1}	5.14	1.41	0.19	5.16	1.53	0.18
28	1	1	-1	{1}	5.28	0.47	0.44	5.27	0.73	0.47
29	-1	0	0	{1}	6.80	1.13	0.72	6.74	1.06	0.93
30	-1	1	1	{-1}	18.20	0.20	2.72	18.05	0.62	2.73
31	0	0	0	{1}	6.00	1.08	0.62	5.99	1.07	0.62
32	0	0	1	{1}	6.72	1.31	0.69	6.74	1.31	0.73
33	0	0	-1	{-1}	11.55	1.68	1.08	11.44	1.62	1.07
34	-1	1	1	{1}	8.26	0.31	1.27	8.37	0.19	1.21
35	0	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.37
36	0	0	-1	{1}	5.22	0.84	0.46	5.23	0.83	0.49
37	-1	-1	1	{-1}	14.44	4.72	1.14	14.46	4.59	1.07
38	-1	-1	-1	{-1}	11.22	3.02	0.42	11.23	3.07	0.51
39	0	0	0	{-1}	13.16	2.22	1.40	13.18	2.17	1.37
40	1	1	1	{-1}	14.83	1.44	1.56	14.81	1.46	1.55

The results from the analysis of variance (ANOVA) of each regression model are summarized in Table 3. The coefficient of determination (R-squared) is the variable of interest in this statistical analysis, which determines how well the regression line fits the simulated force data from the bending simulations. Higher R-squared values indicate that

Table 3

Summary of ANOVA for the loading force (Y_1), the unloading force (Y_2), and the force slope (Y_3)

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
ANOVA for Response Surface Quadratic model					
Model (Y_1)	603.35	10	60.33	7771.79	< 0.0001
<i>A</i> -Inter-bracket Distance	24.00	1	24.00	3091.79	< 0.0001
<i>B</i> -Deflection	29.11	1	29.11	3750.07	< 0.0001
<i>C</i> -Temperature	27.35	1	27.35	3523.59	< 0.0001
<i>D</i> -Wire Geometry	510.72	1	510.72	65787.26	< 0.0001
<i>AB</i>	0.18	1	0.18	23.27	< 0.0001
<i>AC</i>	0.051	1	0.051	6.52	0.0162
<i>AD</i>	3.59	1	3.59	462.05	< 0.0001
<i>BC</i>	0.073	1	0.073	9.39	0.0047
<i>BD</i>	4.66	1	4.66	599.76	< 0.0001
<i>CD</i>	3.60	1	3.60	464.24	< 0.0001
Residual	0.23	29	7.763×10 ⁻³		
Lack of Fit	0.23	19	0.012		
Pure Error	0.000	10	0.000		
Std. Dev.	0.088			R-Squared	0.9996
Mean	9.59			C.V. %	0.92
ANOVA for Response Surface Quadratic model					
Model (Y_2)	42.37	8	5.30	223.58	< 0.0001
<i>A</i> -Inter-bracket Distance	0.050	1	0.050	2.11	0.1563
<i>B</i> -Deflection	21.30	1	21.30	899.10	< 0.0001
<i>C</i> -Temperature	2.81	1	2.81	118.40	< 0.0001
<i>D</i> -Wire Geometry	12.51	1	12.51	528.07	< 0.0001
<i>AB</i>	1.25	1	1.25	52.95	< 0.0001
<i>BC</i>	0.80	1	0.80	33.81	0.0001
<i>BD</i>	3.31	1	3.31	139.84	< 0.0001
<i>CD</i>	0.34	1	0.34	14.38	0.0006
Residual	0.73	29	0.024		
Lack of Fit	0.73	19	0.035		
Pure Error	0.00	10	0.00		
Std. Dev.	0.15			R-Squared	0.9830
Mean	1.63			C.V. %	9.42

Table 3 (continue)

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
ANOVA for Response Surface Quadratic model					
Model (Y_3)	14.25	10	1.43	112.21	< 0.0001
<i>A</i> -Inter-bracket Distance	1.94	1	1.94	152.77	< 0.0001
<i>B</i> -Deflection	4.90	1	4.90	385.76	< 0.0001
<i>C</i> -Temperature	0.70	1	0.70	54.76	< 0.0001
<i>D</i> -Wire Geometry	5.08	1	5.08	400.18	< 0.0001
<i>AB</i>	0.11	1	0.11	8.70	0.0062
<i>AC</i>	0.064	1	0.064	5.02	< 0.0001
<i>AD</i>	0.29	1	0.29	22.86	< 0.0001
<i>BD</i>	0.65	1	0.65	51.01	< 0.0001
<i>CD</i>	0.097	1	0.097	7.60	0.0100
B^2	0.42	1	0.42	33.41	< 0.0001
Residual	0.37	29	0.013		
Lack of Fit	0.37	19	0.019		
Pure Error	0.000	10	0.000		
Std. Dev.	0.11			R-Squared	0.9748
Mean	0.90			C.V. %	12.58

the model correctly predicts the archwire force within the range of factors studied. The ANOVA analysis reveals that the R-squared values for the loading force (Y_1), the unloading force (Y_2), and the force slope (Y_3) are 0.996, 0.9830, and 0.9748, respectively, proving that the chosen quadratic models are good response predictors and are in agreement with the actual simulation results. Additionally, the probability values (Prob>F) are less than 0.0001, indicating the significance of the developed prediction equations.

Given that three factors influence the force responses, a perturbation plot is used to determine which factors have the greatest effect on the wire forces. Each factor's reference point (coded as 0) is set to the midpoint by default in the program's settings. A line plot with a high inclination or curvature indicates that the response is sensitive to changes in a particular factor. In contrast, a line plot with a relatively flat slope indicates that the response is insensitive to that factor.

The loading force perturbation plots for the rectangular and round archwire are shown in Figures 3(a) and 3(b), respectively. The coded values for each factor correspond to the actual values, as stated in Table 1. In brief, the loading force is highly dependent on the wire's geometry, with rectangular wire always releasing greater force magnitude. For example, the rectangular wire's forces ranged from 11.30 N to 14.93 N. However, the values decrease dramatically as soon as the round wire is used, now ranging from 5.22 N to 6.80 N. Bear in mind that while both archwires release greater forces than the suggested

values, their severe impact on dentition may be of some concern, as the force strength decrease abruptly to a lower magnitude of 0.55 N as soon as small tooth movement takes place by 0.5 mm as seen in Figure 2.

Figures 3(a) and 3(b) demonstrate that all considered factors have similar significance on the loading force magnitude. The loading force increased linearly with increasing wire deflection and temperature while linearly decreasing as the inter-bracket distance increased. It is seen that similar plot trends are observed for both wire geometries, with the rectangular wire exhibiting the steepest slope. It demonstrates that the rectangular wire magnifies the impact of bending factors on the wire force. For comparison, at a given temperature of 36°C and deflection of 3.0 mm, the change of inter-bracket distance from 7.0 mm to 8.0 mm significantly decreased the rectangular wire’s loading force by 3.20 N (from 14.93 N to 11.73 N). Meanwhile, a smaller loading force reduction of 1.45 N (from 6.80 N to 5.35 N) was recorded when the round wire was considered for the same bending settings.

The unloading force perturbation plots for the rectangular and round archwire are shown in Figures 4(a) and 4(b), respectively. The rectangular wire’s unloading forces ranged between 1.11 N and 3.61 N. In contrast, the round wire’s unloading forces ranged between 0.57 N and 1.71 N. The plots demonstrate that the unloading force is the most responsive to variations in wire deflection, followed by changes in temperature and inter-bracket distance. It is noticed that both wire geometries exhibit similar plot patterns, with the rectangular wire demonstrating the biggest variations in force magnitudes as a function of bending factor variation. The unloading force magnitude decreased significantly with increasing wire deflection and mildly increased as the temperature and the inter-bracket distance increased. For instance, at a given temperature of 36°C and inter-bracket distance

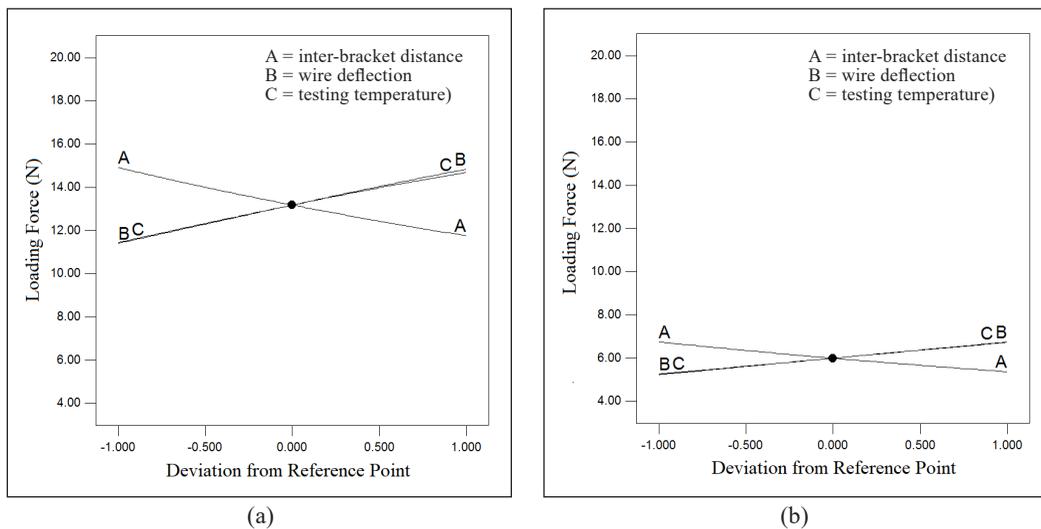


Figure 3. Perturbation plot showing the effect of factors on the loading force of (a) rectangular; and (b) round archwire

of 7.5 mm, the change of wire deflection from 2.0 mm to 4.0 mm had considerably reduced the unloading force of the rectangular wire by 2.50 N (from 3.61 N to 1.11 N). Meanwhile, a smaller unloading force reduction of 1.14 N (from 1.71 N to 0.57 N) was recorded using round wire for the same bending settings.

The wire force slope perturbation plots for the rectangular and round archwire are shown in Figures 5(a) and 5(b), respectively. The rectangular wire had a force slope of

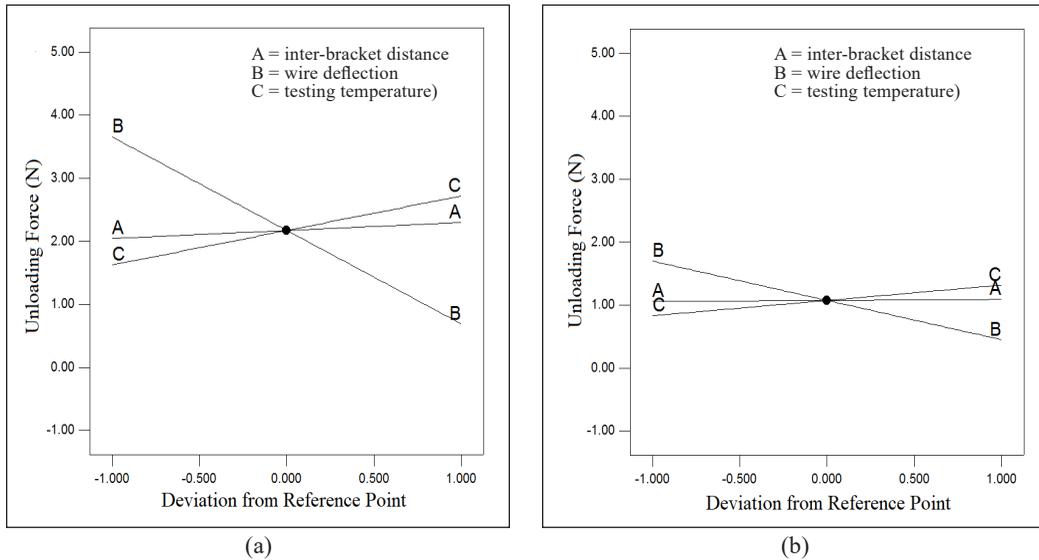


Figure 4. Perturbation plots showing the effect of factors on the unloading force of (a) rectangular; and (b) round archwire

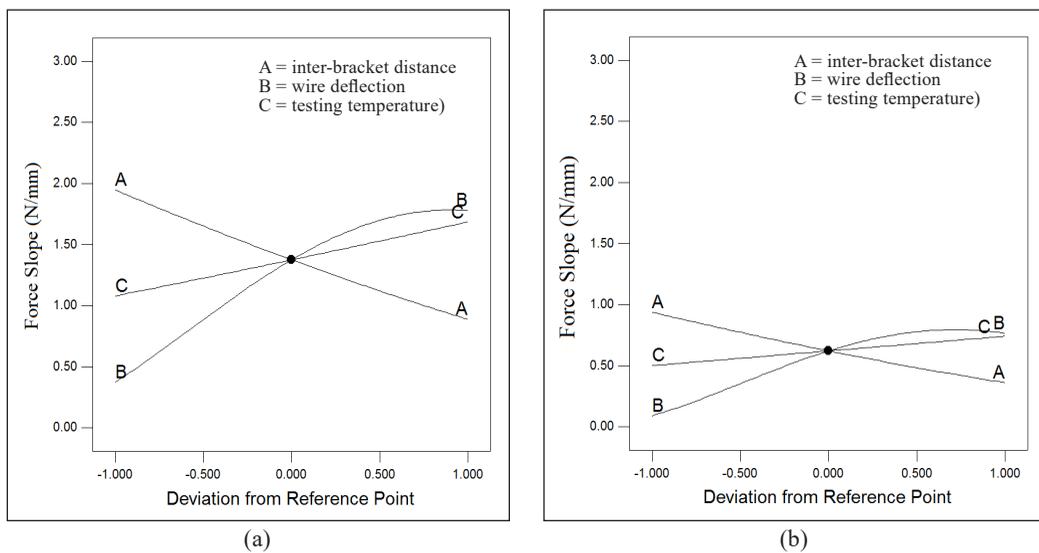


Figure 5. Perturbation plots showing the effect of factors on the force slope of (a) rectangular; and (b) round archwire

0.44 N/mm to 1.71 N/mm, whereas the round wire had a force slope of 0.13 N/mm to 0.72 N/mm. In general, the force slope is the most sensitive to changes in the amount of wire deflection, followed by the changes in inter-bracket distance and temperature. The plots also demonstrate that the force slope is directly proportional to the wire deflection and the bending temperature but inversely proportional to the distance set between brackets. It is interesting to highlight that the force slope's sensitivity to the deflection is reduced as the wire deflects more than 3.0 mm. Both wire geometries demonstrate similar plot trends, with the rectangular wire having the highest change in force slope values when bending settings are varied. For example, at a given temperature of 36°C and deflection of 3.0 mm, the change of inter-bracket distance from 7.0 mm to 8.0 mm had reduced the force slope of the rectangular wire by 0.59 N/mm (from 1.66 N/mm to 1.07 N/mm). Meanwhile, a smaller force slope reduction of 0.25 N/mm (from 0.72 N/mm to 0.47 N/mm) was recorded using round wire for the same bending settings.

DISCUSSION

It is difficult to precisely estimate the archwire force during orthodontic treatment due to many uncontrolled variables affecting the force transmission, such as the pressure from the lips and tongue and the biting force (Proffit et al., 2018). Therefore, while utilizing the established regression models, readers should remember that the estimated biomechanical forces are only valid for the wire-bracket combination and settings addressed in this work. In addition, the choice of ligation methods may result in varying degrees of frictional resistance, thus further complicating the force prediction. For instance, when the elastomeric ligatures are considered during the orthodontic treatment, the overall resistance experienced during the sliding is increased. As a result, the loading and unloading forces measured during the bending would rise and drop in magnitudes greater than those observed in the current study.

The goal of using a rectangular wire as soon as feasible is critical to improving tooth movement as it promotes torque control to the malposed tooth. Unfortunately, the force values summarized in Table 2 called into question the usefulness of considering the rectangular wire at the early stage of leveling since the force magnitudes vary between 1.44 N and 4.72 N for all the considered bending settings, exceeding the ideal force range. Based on this observation, orthodontists are encouraged to avoid using rectangular wires in scenarios involving persistently significant tooth level discrepancy. On the contrary, using the 0.016-inch round archwire would be the optimal choice, as the force varied between 0.13 N and 0.73 N for the large 4.0 mm deflection condition. Notably, this force range is within the reported range of 0.10 N to 1.20 N for optimal tooth movement (Wu et al., 2018; Theodorou et al., 2019).

The relationship between the archwire force and the amount of bending deflection applied to the wire may provide an advantage during orthodontic treatment, specifically during the reactivation process. The present finding of increased wire force at smaller wire deflection suggests that the NiTi archwire may be momentarily removed and then reinstalled during the clinical visit to release more force than at initial activation. For example, if a tooth has traveled away from its original vertical discrepancy of 4.0 mm, reactivating the same wire at a lesser discrepancy, say 2.0 mm, may result in a higher unloading force being delivered. This greater force may be advantageous for leveling a tooth of bigger size, such as a premolar or molar teeth. Additionally, the force data in Table 2 also demonstrate that the unloading force increases as the oral temperature increases. The NiTi wire stiffens when a patient consumes warm drinks, and the unloading force magnitude increases abruptly.

For healthier tooth movement, the force released by the archwire to the dentition must be light and steady throughout the therapy. Therefore, maintaining the periodontal ligament's present cellular activity is necessary for connection to an optimal tooth movement experience. From Table 2, there was only one configuration in which the NiTi wire in the bracket system produced consistent force magnitude during the unloading cycle. This configuration, which incorporates a 2.0 mm wire deflection in 8.0 mm inter-bracket distance condition, results in a force slope of practically zero throughout the unloading cycle. In comparison, at the critical bending conditions, the round and rectangular wire exhibited a steeper force slope of 1.27 N/mm and 2.72 N/mm, respectively.

Given that leveling treatment is usually associated with significant tooth misalignment, the optimal strategy would be to utilize a springier archwire to reduce the effect of sliding friction on the amount and slope of the unloading force. The tiniest round archwire currently available on the market measures 0.012 inches in diameter. According to the expanded simulation work utilizing a comparable three-dimensional finite element model, the unloading of the 0.012-inch superelastic NiTi wire from 4.0 mm deflection in a 36°C environment demonstrated an unloading force slope of 0.3 N/mm (Razali, 2018). This force changing rate is significantly smaller than the rate exhibited by the 0.016 × 0.022-inch and 0.016-inch wires considered in this study, which had a force slope of 1.48 N/mm and 0.68 N/mm, respectively.

CONCLUSION

Three regression models for predicting the loading, unloading, and slope of force exerted by a superelastic NiTi archwire in an orthodontic bracket system were successfully established using response surface methods. The developed regression models exhibit high R-squared values of 0.996, 0.9830, and 0.9748 for the loading, unloading force, and force slope, respectively. The impact of the considered bending settings on wire force is substantially stronger when the rectangular wire is used. The magnitude and slope of the unloading force

of superelastic NiTi wire decrease linearly as the amount of wire deflection and the distance between brackets increase, respectively. Therefore, the 0.016-inch diameter archwire is more appropriate for use during the early stages of orthodontic treatment due to its lighter force of 0.57 N to 1.71 N and a lower force slope of 0.13 N/mm to 0.72 N/mm. On the other hand, the use of 0.016 × 0.022-inch rectangular wire may cause patient discomfort, as the force transferred to induce tooth movement may surpass 3.61 N.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support of the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2020/TK0/USM/03/5).

REFERENCES

- Alavi, S., & Hosseini, N. (2012). Load-deflection and surface properties of coated and conventional superelastic orthodontic archwires in conventional and metal-insert ceramic brackets. *Dental Research Journal*, 9(2), 133-138. <https://doi.org/10.4103/1735-3327.95225>
- Arreghini, A., Lombardo, L., Mollica, F., & Siciliani, G. (2016). Load deflection characteristics of square and rectangular archwires. *International Orthodontics*, 14(1), 1-14. <https://doi.org/10.1016/j.ortho.2015.12.011>
- Auricchio, F., & Taylor, R. L. (1997). Shape-memory alloys: Modelling and numerical simulations of the finite-strain superelastic behavior. *Computer Methods in Applied Mechanics and Engineering*, 143(1-2), 175-194. [https://doi.org/10.1016/S0045-7825\(96\)01147-4](https://doi.org/10.1016/S0045-7825(96)01147-4)
- Badawi, H. M., Toogood, R. W., Carey J. P. R., Heo, G., & Major, P. W. (2009). Three-dimensional orthodontic force measurements. *American Journal of Orthodontics and Dentofacial Orthopedics*, 136(4), 518-528. <https://doi.org/10.1016/j.ajodo.2009.02.025>
- Elayyan, F., Silikas, N., & Bearn, D. (2010). Mechanical properties of coated superelastic archwires in conventional and self-ligating orthodontic brackets. *American Journal of Orthodontics and Dentofacial Orthopedics*, 137(2), 213-217. <https://doi.org/10.1016/j.ajodo.2008.01.026>
- Franchi, L., Baccetti, T., Camporesi, M., & Giuntini, V. (2009). Forces released by nonconventional bracket or ligature systems during alignment of buccally displaced teeth. *American Journal of Orthodontics and Dentofacial Orthopedics*, 136(3), 316.e1-316.e6. <https://doi.org/10.1016/j.ajodo.2009.02.016>
- Higa, R. H., Henriques, J. F. C., Janson, G., Matias, M., Freitas, K. M. S., Henriques, F. P., & Francisconi, M. F. (2017). Force level of small diameter nickel-titanium orthodontic wires ligated with different methods. *Progress in Orthodontics*, 18(1), Article 21. <https://doi.org/10.1186/s40510-017-0175-z>
- Kusy, R. P., & Whitley, J. Q. (2000). Resistance to sliding of orthodontic appliances in the dry and wet states: Influence of archwire alloy, interbracket distance, and bracket engagement. *Journal of Biomedical Materials Research*, 52(4), 797-811. [https://doi.org/10.1002/1097-4636\(20001215\)52:4%3C797::AID-JBM25%3E3.0.CO;2-9](https://doi.org/10.1002/1097-4636(20001215)52:4%3C797::AID-JBM25%3E3.0.CO;2-9)
- Nucera, R., Gatto, E., Borsellino, C., Aceto, P., Fabiano, F., Matarese, G., Perillo, L., & Cordasco, G. (2014). Influence of bracket-slot design on the forces released by superelastic nickel-titanium alignment

- wires in different deflection configurations. *The Angle Orthodontist*, 84(3), 541-547. <https://doi.org/10.2319/060213-416.1>
- Papageorgiou, S. N., Höchli, D., & Eliades, T. (2017). Outcomes of comprehensive fixed appliance orthodontic treatment: A systematic review with meta-analysis and methodological overview. *The Korean Journal of Orthodontics*, 47(6), 401-413. <https://doi.org/10.4041/kjod.2017.47.6.401>
- Proffit, W., Fields, H., Larson, B., & Sarver, D. (2018). *Contemporary orthodontics*. Elsevier.
- Razali, M. F. (2018). *Phase transformation and force-deflection responses of NiTi archwire for bracket assembly in orthodontic treatment* [Unpublished Doctoral dissertation]. Universiti Sains Malaysia, Malaysia.
- Razali, M. F., Mahmud, A. S., & Mokhtar, N. (2018). Force delivery of NiTi orthodontic archwire at different magnitude of deflections and temperatures: A finite element study. *Journal of the Mechanical Behavior of Biomedical Materials*, 77, 234-241. <https://doi.org/10.1016/j.jmbbm.2017.09.021>
- Thalman, T. D. (2008). *Unloading behavior and potential binding of superelastic orthodontic leveling wires* [Unpublished Doctoral dissertation]. Saint Louis University, USA.
- Theodorou, C. I., Kujipers-Jagtman, A. M., Bronkhorst, E. M., & Wagener, F. A. D. T. G. (2019). Optimal force magnitude for bodily orthodontic tooth movement with fixed appliances: A systematic review. *American Journal of Orthodontics and Dentofacial Orthopedics*, 156(5), 582-592. <https://doi.org/10.1016/j.ajodo.2019.05.011>
- Wilkinson, P. D., Dysart, P. S., Hood, J. A. A., & Herbison, G. P. (2002). Load-deflection characteristics of superelastic nickel-titanium orthodontic wires. *American Journal of Orthodontics and Dentofacial Orthopedics*, 121(5), 483-495. <https://doi.org/10.1067/mod.2002.121819>
- Wu, J., Liu, Y., Peng, W., Dong, H., & Zhang, J. (2018). A biomechanical case study on the optimal orthodontic force on the maxillary canine tooth based on finite element analysis. *Journal of Zhejiang University Science B*, 19(7), 535-546. <https://doi.org/10.1631/jzus.B1700195>



Robust Linear Discriminant Rule Using Double Trimming Location Estimator with Robust Mahalanobis Squared Distance

Yik Siong Pang*, Nor Aishah Ahad and Sharipah Soaad Syed Yahaya

School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM, Sintok, Kedah, Malaysia

ABSTRACT

The commonly employed classical linear discriminant rule, based on classical mean and covariance, are highly sensitive to outliers. Therefore, outlier influence on location and scale estimation will affect the accuracy of a discriminant rule and lead to high misclassification rates. The past studies used classical Mahalanobis Squared Distance (MSD) to alleviate the problem. However, the highly sensitive mean and covariance shortcoming can still affect the distance computation, causing masking and swamping effects. In a previous study, researchers proposed a double trimming procedure that adopted MSD-based α -trimmed mean into MSD-based α -trimmed median to construct a robust classifier. However, the proposed procedure has an overlooked flaw because the procedure employed the MSD in the computation. Thus, this study proposed to employ a robust MSD for the distance-based trimmed median procedure. The improvised trimmed median was then used to construct a robust linear discriminant rule and compared with the classical and existing robust rules using a simulation study. The results show that this study's proposed robust linear discriminant rule has better accuracy and consistent performance than the classical linear discriminant rule and two other robust linear discriminant rules.

Keywords: Discriminant analysis, distance-based trimmed median, robust Mahalanobis squared distance

ARTICLE INFO

Article history:

Received: 22 January 2022

Accepted: 20 April 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.05>

E-mail addresses:

ys_pang@ahsgs.uum.edu.my (Yik Siong Pang)

aishah@uum.edu.my (Nor Aishah Ahad)

sharipah@uum.edu.my (Sharipah Soaad Syed Yahaya)

* Corresponding author

INTRODUCTION

The classical linear discriminant analysis (CLDA) is a widely employed multivariate classification technique to allocate new observations into one of the dichotomous categories. CLDA is derived by using population mean (μ_1, μ_2) with homoscedasticity assumptions ($\Sigma_1 = \Sigma_2 = \Sigma$).

Since actual population parameters are usually unattainable, the Classical Linear Discriminant Rule (CLDR) will be constructed based on estimated mean vectors ($\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$) and scatter matrices (\mathbf{S}_1 and \mathbf{S}_2). For example, suppose that a set of d -dimensional training data consists of n_1 and n_2 observations corresponding to populations π_1 and π_2 , the allocation rule of new unknown observation (\mathbf{x}_0) is defined as Equation 1 (Johnson & Wichern, 2013):

If

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}_{pooled}^{-1} \left\{ \mathbf{x}_0 - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right\} \geq \ln \left(\frac{p_2}{p_1} \right), \tag{1}$$

then $\mathbf{x}_0 \in \pi_1$, otherwise, $\mathbf{x}_0 \in \pi_2$.

where the threshold for the decision rule is decided by the prior probability of the original population of an observation, p_1 , and p_2 .

Notably, the classical mean is extremely sensitive to outliers. Even one outlier distorts the location estimation and influences (co)variance accuracy (Erceg-hurn et al., 2013). Therefore, CLDR’s affected mean and covariance will result in a high misclassification rate. Pang et al. (2019) adopted and modified Alloway and Raghavachari’s (1990) Mahalanobis Squared Distance (MSD) based α -Trimmed Mean into an MSD-based α -Trimmed Median to construct a robust classifier and obtained practical results compared to the former and classical estimations.

This paper further explores the Distance-based α -Trimmed Median by Pang et al. (2019). The double trimming procedure employed in Pang et al. (2019) used MSD-based trimming as the first layer and Median act as the second trimming has an overlooked flaw, such that the MSD computation itself, which is also based on the outlier sensitive mean vector ($\bar{\mathbf{x}}$) and covariance matrix (\mathbf{S}). The notation for MSD for each d -dimensional observation x_i for $i = 1, \dots, n$ is written in Equation 2 (Hadi et al., 2009; Rousseeuw & Van Zomeren, 1990).

$$MSD(x_i) = (x_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (x_i - \bar{\mathbf{x}}) \tag{2}$$

Suppose outlier(s) exist in the data. It will result in an inaccurate mean vector and covariance matrix and cause classical MSD to suffer from outliers’ secondary effects, *masking*, and *swamping* effects. *Masking* effect or outlier false negative which higher distance measurement outliers masked the supposedly less-far outliers. The *swamping* effect or false positive outlier refers to the shift of the MSD’s centroid, causing good observations to appear as outliers. While the high-distance observations were deemed contaminated and trimmed, there are possibilities that outliers may remain in the multivariate data set (Pang et al., 2021). Hence, such a situation may cause any estimation from the trimmed data to be inaccurate. Previous researchers had robustified the distance measurement by using other robust estimators such as M-estimator, Minimum Volume Ellipsoid (MVE), and Minimum Covariance Determinant (MCD) to overcome the masking and swamp

problems (Campbell, 1980; Penny & Jolliffe, 2001; Rousseeuw, 1985; Rousseeuw & Van Driessen, 1999; Rousseeuw & Van Zomeren, 1990). Therefore, this study robustified the classical MSD with a robust location estimator, median and robust scale estimator, and Robust Covariance (S_R) to improve the performance of the distance-based procedure in Pang et al. (2019).

The rest of this paper follows: The materials and methods section discusses the formulae, methods, and simulation settings, while the results and discussions section highlights and interprets the simulation results and ends with a conclusion.

MATERIALS AND METHODS

The classical estimators \bar{x} and S_{pooled} in CLDR were substituted with robust estimators to construct Robust Linear Discriminant Rules (RLDR). Three robust location estimators were employed in this study, namely (i) α -Trimmed Mean based on classical MSD ($\bar{X}_{(MSD,\alpha)}$), (ii) α -Trimmed Median based on classical MSD ($\hat{M}_{(MSD_R,\alpha)}$), and (iii) α -Trimmed Median based on robust MSD ($\hat{M}_{(MSD,\alpha)}$). All three robust location estimators were paired with the same robust scale estimator, Winsorized Robust Covariance (S_{RW}), to reduce unconformity other than the location estimations when comparing the performance of the robust classifiers. The three pairs of robust location and scale estimators were then substituted into Equation 1 to construct $RLDR_M$, $RLDR_T$, and $RLDR_R$. A total of four classifiers (one classical and three robust) were constructed and listed in Table 1. The other formulae and procedures involved in this study are discussed in the subsequent subsections.

Table 1
Classifiers' notation with paired location and scale estimator

	Classifier	Notation	Location Estimator	Scale Estimator
i)	CLDR	CLDR	\bar{x}	S
ii)	$RLDR_{R(\alpha)}$	$R(\alpha)$	$\hat{M}_{(MSD_R,\alpha)}$	S_{RW}
iii)	$RLDR_{T(\alpha)}$	$T(\alpha)$	$\hat{M}_{(MSD,\alpha)}$	S_{RW}
iv)	$RLDR_{M(\alpha)}$	$M(\alpha)$	$\bar{X}_{(MSD,\alpha)}$	S_{RW}

Robust Covariance

The Robust Covariance (S_R) is a direct substitution of robust components, Rescaled Median Absolute Deviation (MAD_n) and Spearman Rho (ρ_s), into the components of classical variance-covariance matrix, σ and ρ , as shown in Equations 3 and 4 (Abu-Shawiesh & Abdullah, 2001; Lim et al., 2016; Pang et al. 2019; Yahaya et al., 2016):

$$S_{ij} = \rho_{ij} \times \sigma_i \times \sigma_j \tag{3}$$

$$S_{Rij} = \rho_{s_{ij}} \times MAD_{n_i} \times MAD_{n_j} \tag{4}$$

Robust Mahalanobis Squared Distance

As mentioned earlier, to alleviate outliers' *masking* and *to swamp* the effect on the distance-based trimming algorithm, Median ($\hat{\mathbf{M}}$) was chosen as it is a well-known classical location estimator with high outliers tolerance. The classical MSD was robustified as \mathbf{MSD}_R using ($\hat{\mathbf{M}}$) and \mathbf{S}_R in Equation 5:

$$MSD_R(x_i) = (x_i - \hat{\mathbf{M}})^T \mathbf{S}_R^{-1} (x_i - \hat{\mathbf{M}}) \tag{5}$$

Distance-Based Trimming Algorithm

In order to examine the effect of different trimming percentages on robust classifiers' ability to handle outliers, α is set to be 20% and 40%. The selection of these trimming percentages illustrates the difference in using either moderate or high trimming percentages to lower the misclassification rates. The following algorithm was carried out to compute the three robust location and scale estimations:

- Step 1: Compute the distance measurement (MSD or \mathbf{MSD}_R);
- Step 2: Arrange the distance measurement in ascending order;
- Step 3: Trim $\alpha\%$ of the highest distance for each population ($\alpha = 20\%$ or 40%);
- Step 4: Compute the mean or median based on the trimmed sample;
- Step 5: Winsorize all the trimmed observations ($\alpha\%$) by replacing them with the next $\alpha\%$ highest MSD observations in the remaining samples;
- Step 6: Compute \mathbf{S}_{RW} based on the winsorized observations.

Simulation Settings

The performance of a classifier depends on classification accuracy, i.e., the lower the misclassification rates, the better the classifier. Therefore, the classical and robust discriminant rules (CLDR, RLDR_R, RLDR_T, and RLDR_M) were examined via simulation study using the Tukey-Huber Contamination Model (THCM), as shown in Equation 6 (Huber, 1964; Tukey, 1962). THCM was employed in various previous robust discriminant studies to generate data for classifier validation (Croux & Dehon, 2001; He & Fung, 2000; Hubert & Van Driessen, 2004; Lim et al., 2016; Pang et al., 2019; Todorov & Pires, 2007; Yahaya et al., 2016).

$$\begin{aligned} \pi_1: (1 - \varepsilon)n_1N_d(0, I_d) + \varepsilon n_1N_d(0 + \mu, \omega I_d) \\ \pi_2: (1 - \varepsilon)n_2N_d(1, I_d) + \varepsilon n_2N_d(1 - \mu, \omega I_d) \end{aligned} \tag{6}$$

According to THCM, the underlying distribution for populations π_1 and π_2 are d -dimensional Normal Distributions with locations centered at 0 and 1, respectively, and scaled with d -dimensional Identity Matrices (I_d). The simulation settings used in this study are shown in Table 2.

Table 2
Simulation parameters settings

Contamination Parameters	Controlled Setting
Data Dimensions (d)	2; 6
Contamination Percentage (ε)	0%; 10%; 20%; 40%
Training Sample Size (n_1, n_2)	(20,20); (50,50); (100,100)
Contaminants Location Shift (μ)	0; 3; 5
Contaminants Covariance Amplification (ω)	1; 9; 25; 100

The procedure for the simulation study is as follows:

- Step 1: Generate training data according to one of each parameter setting in Table 2;
- Step 2: Construct a discriminant rule based on the training data;
- Step 3: Generate 2,000 uncontaminated testing samples for each population and record the misclassification rate;
- Step 4: Repeat Step 1 to Step 3 for 2,000 iterations;
- Step 5: Calculate the mean misclassification rate of the 2,000 iterations.

RESULTS AND DISCUSSION

The average misclassifications of the seven classifiers on the 204 data distributions from 2,000 iterations were recorded and organized according to training samples' size and data dimension in Tables 3, 5, 7, 9, 11, and 13. The performance of the classifiers was compared separately as two categories for different trimming percentages (α), i.e., i) CLDR vs. RLDRs ($\alpha=40\%$) and ii) CLDR vs. RLDRs ($\alpha=20\%$). The lowest misclassification rates were bolded with grey highlights for each category. Additionally, descriptive statistics (Tables 4, 6, 8, 10, 12, and 14) were conducted on the simulation results to illustrate the classifiers' performance better. Similar to the performance comparison, the lowest mean and standard deviation for each category were bolded with grey highlights.

Based on the results in Tables 3, 5, 7, 9, 11, and 13, CLDR undoubtedly performed the best amongst the classifiers when dealing with non-contaminated training data ($\varepsilon=0, \mu=0, \omega=1$) for all sample size-dimension combinations. Contrarily, when the classifiers deal with contaminated training data, the robust classifiers perform better than the classical classifier as CLDR constantly remains at high misclassification rates. However, throughout the 198 contaminated data distributions, neither robust classifier illustrated an overall dominant performance over one another regardless of trimming percentage.

Firstly, comparing $RLDR_R$ and $RLDR_T$ performance in handling contaminated data (198 distributions), there were 149 and 153 occasions where $RLDR_R$ was better than $RLDR_T$, at a trimming percentage of 40% and 20%, respectively. At the data distributions where $RLDR_R$ outperforms $RLDR_T$, the maximum difference may increase up to 0.2669 (Table 13, $n=100, d=6, \varepsilon=40\%, \mu=5, \omega=1, \alpha=40\%$) and 0.2164 (Table 5, $n=20, d=6, \varepsilon=20\%, \mu=5,$

Table 3
Misclassification rate for sample size (20,20) at d=2

ϵ	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.2511	0.2767	0.2689	0.2641	0.2689	0.2656	0.2614
10	0	9	0.3178	0.2765	0.2747	0.2704	0.2671	0.2653	0.2611
10	0	25	0.4205	0.2765	0.2816	0.2798	0.2666	0.2651	0.2612
10	0	100	0.4903	0.2764	0.2903	0.2919	0.2665	0.2650	0.2629
10	3	1	0.3389	0.2898	0.2936	0.2880	0.2723	0.2740	0.2687
10	3	9	0.3884	0.2781	0.2798	0.2756	0.2677	0.2663	0.2619
10	3	25	0.4527	0.2768	0.2838	0.2831	0.2669	0.2653	0.2618
10	3	100	0.4937	0.2764	0.2919	0.2932	0.2666	0.2653	0.2630
10	5	1	0.4987	0.2902	0.3004	0.2944	0.2698	0.2693	0.2639
10	5	9	0.4548	0.2799	0.2827	0.2783	0.2680	0.2676	0.2634
10	5	25	0.4755	0.2774	0.2850	0.2847	0.2669	0.2664	0.2627
10	5	100	0.4961	0.2764	0.2922	0.2934	0.2667	0.2655	0.2623
20	0	9	0.3624	0.2777	0.2761	0.2735	0.2628	0.2639	0.2604
20	0	25	0.4637	0.2777	0.2804	0.2822	0.2615	0.2631	0.2603
20	0	100	0.4995	0.2779	0.2838	0.2802	0.2615	0.2629	0.2638
20	3	1	0.5770	0.3004	0.3102	0.3141	0.2918	0.3152	0.3454
20	3	9	0.5083	0.2796	0.2822	0.2827	0.2627	0.2649	0.2613
20	3	25	0.5041	0.2781	0.2840	0.2863	0.2617	0.2633	0.2604
20	3	100	0.5027	0.2778	0.2847	0.2914	0.2615	0.2629	0.2638
20	5	1	0.6530	0.2931	0.2955	0.2929	0.2607	0.3010	0.3871
20	5	9	0.6039	0.2826	0.2855	0.2864	0.2629	0.2655	0.2630
20	5	25	0.5310	0.2788	0.2856	0.2882	0.2617	0.2638	0.2613
20	5	100	0.5053	0.2778	0.2855	0.2921	0.2616	0.2630	0.2640
40	0	9	0.4100	0.2714	0.2732	0.2730	0.2774	0.2769	0.3465
40	0	25	0.4804	0.2694	0.2722	0.2830	0.2775	0.2774	0.4523
40	0	100	0.4975	0.2689	0.2714	0.3160	0.2781	0.2777	0.4971
40	3	1	0.7061	0.4763	0.5428	0.6530	0.5432	0.5828	0.6756
40	3	9	0.6106	0.2730	0.2815	0.3014	0.2908	0.2954	0.4685
40	3	25	0.5174	0.2692	0.2737	0.2955	0.2821	0.2837	0.4872
40	3	100	0.5005	0.2689	0.2719	0.3179	0.2790	0.2792	0.4995
40	5	1	0.6955	0.3039	0.5061	0.6753	0.4716	0.5642	0.6941
40	5	9	0.6693	0.2735	0.2878	0.3382	0.3059	0.3135	0.5694
40	5	25	0.5446	0.2692	0.2747	0.3053	0.2848	0.2881	0.5099
40	5	100	0.5023	0.2688	0.2723	0.3195	0.2798	0.2801	0.5011

Table 4
Summary of misclassification rate for sample size (20,20) at d=2

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.4978	0.2843	0.2972	0.3131	0.2851	0.2914	0.3514
Std.Dev.	0.0998	0.0344	0.0577	0.0891	0.0572	0.0719	0.1295
Min	0.2511	0.2688	0.2689	0.2641	0.2607	0.2629	0.2603
Max	0.7061	0.4763	0.5428	0.6753	0.5432	0.5828	0.6941
Range	0.4550	0.2075	0.2739	0.4112	0.2825	0.3199	0.4338

Table 5
Misclassification rate for sample size (20,20) at $d=6$

ε	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.1409	0.1948	0.1907	0.1835	0.1748	0.1735	0.1673
10	0	9	0.2108	0.1942	0.1886	0.1819	0.1711	0.1711	0.1645
10	0	25	0.2543	0.1944	0.1884	0.1819	0.1707	0.1712	0.1645
10	0	100	0.2725	0.1944	0.1887	0.1821	0.1711	0.1714	0.1648
10	3	1	0.3915	0.2034	0.2076	0.1994	0.1736	0.2098	0.2046
10	3	9	0.2679	0.1958	0.1923	0.1846	0.1721	0.1715	0.1651
10	3	25	0.2655	0.1943	0.1896	0.1825	0.1718	0.1715	0.1651
10	3	100	0.2733	0.1945	0.1887	0.1820	0.1710	0.1713	0.1648
10	5	1	0.4998	0.2032	0.2045	0.1964	0.1725	0.1889	0.1884
10	5	9	0.3253	0.1967	0.1938	0.1861	0.1724	0.1718	0.1656
10	5	25	0.2783	0.1946	0.1900	0.1827	0.1723	0.1714	0.1650
10	5	100	0.2742	0.1945	0.1888	0.1822	0.1713	0.1713	0.1649
20	0	9	0.2514	0.1924	0.1870	0.1789	0.1649	0.1678	0.1613
20	0	25	0.3613	0.1916	0.1862	0.1780	0.1649	0.1678	0.1613
20	0	100	0.4694	0.1923	0.1867	0.1784	0.1651	0.1676	0.1610
20	3	1	0.5365	0.2139	0.3438	0.3825	0.2549	0.3836	0.4601
20	3	9	0.3933	0.1945	0.1898	0.1817	0.1653	0.1666	0.1606
20	3	25	0.4204	0.1923	0.1871	0.1789	0.1646	0.1676	0.1611
20	3	100	0.4780	0.1919	0.1869	0.1789	0.1652	0.1677	0.1610
20	5	1	0.5668	0.1994	0.3348	0.4499	0.1678	0.3842	0.5837
20	5	9	0.4956	0.1949	0.1935	0.1855	0.1636	0.1652	0.1596
20	5	25	0.4625	0.1941	0.1886	0.1804	0.1650	0.1670	0.1607
20	5	100	0.4846	0.1920	0.1869	0.1791	0.1651	0.1676	0.1610
40	0	9	0.3240	0.1833	0.1845	0.1780	0.1912	0.1920	0.2773
40	0	25	0.4563	0.1828	0.1837	0.1763	0.1930	0.1948	0.4285
40	0	100	0.4991	0.1834	0.1829	0.1749	0.1930	0.1948	0.4933
40	3	1	0.6433	0.4665	0.5349	0.6055	0.5181	0.5511	0.6275
40	3	9	0.6382	0.1817	0.1833	0.1775	0.2229	0.2303	0.4610
40	3	25	0.5355	0.1832	0.1831	0.1753	0.1989	0.2010	0.4833
40	3	100	0.5035	0.1834	0.1832	0.1751	0.1942	0.1951	0.4970
40	5	1	0.6137	0.2644	0.5186	0.6406	0.4709	0.5303	0.6589
40	5	9	0.7232	0.1809	0.1838	0.1800	0.2623	0.2805	0.5876
40	5	25	0.5805	0.1829	0.1828	0.1745	0.2069	0.2119	0.5185
40	5	100	0.5070	0.1832	0.1832	0.1748	0.1959	0.1955	0.4992

Table 6
Summary on misclassification rate for sample size (20,20) at $d=6$

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.4235	0.2023	0.2173	0.2209	0.2005	0.2166	0.2961
Std.Dev.	0.1413	0.0481	0.0853	0.1152	0.0774	0.0965	0.1751
Min	0.1409	0.1809	0.1828	0.1745	0.1636	0.1652	0.1596
Max	0.7232	0.4665	0.5349	0.6406	0.5181	0.5511	0.6589
Range	0.5823	0.2856	0.3521	0.4661	0.3545	0.3859	0.4993

Table 7
Misclassification rate for sample size (50,50) at d=2

ε	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.2442	0.2550	0.2515	0.2496	0.2516	0.2499	0.2483
10	0	9	0.2759	0.2548	0.2558	0.2547	0.2507	0.2503	0.2488
10	0	25	0.3863	0.2548	0.2641	0.2664	0.2506	0.2509	0.2500
10	0	100	0.4842	0.2549	0.2765	0.2833	0.2505	0.2511	0.2506
10	3	1	0.2960	0.2628	0.2658	0.2623	0.2531	0.2545	0.2519
10	3	9	0.3610	0.2556	0.2585	0.2576	0.2509	0.2507	0.2492
10	3	25	0.4441	0.2550	0.2662	0.2696	0.2506	0.2509	0.2502
10	3	100	0.4916	0.2549	0.2774	0.2848	0.2505	0.2513	0.2508
10	5	1	0.4986	0.2633	0.2717	0.2677	0.2524	0.2523	0.2502
10	5	9	0.4732	0.2568	0.2615	0.2610	0.2511	0.2512	0.2494
10	5	25	0.4870	0.2552	0.2672	0.2716	0.2506	0.2510	0.2502
10	5	100	0.4963	0.2549	0.2779	0.2856	0.2505	0.2513	0.2510
20	0	9	0.3055	0.2552	0.2561	0.2559	0.2491	0.2503	0.2483
20	0	25	0.4277	0.2551	0.2623	0.2651	0.2487	0.2500	0.2479
20	0	100	0.4911	0.2551	0.2671	0.2731	0.2486	0.2499	0.2487
20	3	1	0.6202	0.2699	0.2770	0.2769	0.2634	0.2814	0.2967
20	3	9	0.5334	0.2566	0.2588	0.2591	0.2493	0.2508	0.2485
20	3	25	0.5062	0.2554	0.2629	0.2662	0.2487	0.2502	0.2479
20	3	100	0.4993	0.2552	0.2675	0.2740	0.2486	0.2500	0.2487
20	5	1	0.6911	0.2669	0.2693	0.2656	0.2484	0.2705	0.3282
20	5	9	0.6795	0.2578	0.2590	0.2608	0.2492	0.2508	0.2487
20	5	25	0.5590	0.2559	0.2629	0.2662	0.2487	0.2505	0.2482
20	5	100	0.5048	0.2554	0.2680	0.2746	0.2487	0.2500	0.2486
40	0	9	0.3491	0.2528	0.2536	0.2523	0.2548	0.2545	0.2897
40	0	25	0.4571	0.2521	0.2534	0.2527	0.2551	0.2550	0.4075
40	0	100	0.4965	0.2520	0.2533	0.2640	0.2549	0.2551	0.4925
40	3	1	0.7328	0.4645	0.5662	0.6920	0.5686	0.6159	0.7046
40	3	9	0.6767	0.2537	0.2550	0.2580	0.2603	0.2638	0.4598
40	3	25	0.5499	0.2522	0.2537	0.2546	0.2568	0.2579	0.4854
40	3	100	0.5035	0.2520	0.2534	0.2658	0.2552	0.2557	0.4999
40	5	1	0.7252	0.2759	0.5070	0.7049	0.4599	0.5954	0.7175
40	5	9	0.7172	0.2537	0.2576	0.2822	0.2688	0.2773	0.6317
40	5	25	0.5992	0.2522	0.2536	0.2582	0.2581	0.2599	0.5379
40	5	100	0.5080	0.2520	0.2534	0.2677	0.2556	0.2563	0.5042

Table 8
Summary on misclassification rate for sample size (50,50) at d=2

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.5021	0.2626	0.2784	0.2913	0.2680	0.2755	0.3409
Std.Dev.	0.1266	0.0355	0.0654	0.1022	0.0631	0.0829	0.1427
Min	0.2442	0.2520	0.2515	0.2496	0.2484	0.2499	0.2479
Max	0.7328	0.4645	0.5662	0.7049	0.5686	0.6159	0.7175
Range	0.4886	0.2125	0.3147	0.4553	0.3202	0.3660	0.4696

Table 9
Misclassification rate for sample size (50,50) at $d=6$

ε	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.1214	0.1449	0.1420	0.1385	0.1364	0.1359	0.1329
10	0	9	0.1812	0.1448	0.1423	0.1385	0.1355	0.1343	0.1311
10	0	25	0.2696	0.1450	0.1430	0.1393	0.1354	0.1342	0.1309
10	0	100	0.4413	0.1448	0.1434	0.1397	0.1355	0.1336	0.1305
10	3	1	0.3286	0.1522	0.1523	0.1475	0.1353	0.1524	0.1487
10	3	9	0.2757	0.1460	0.1450	0.1408	0.1355	0.1340	0.1308
10	3	25	0.3288	0.1449	0.1435	0.1398	0.1356	0.1341	0.1308
10	3	100	0.4572	0.1449	0.1433	0.1397	0.1354	0.1337	0.1305
10	5	1	0.5004	0.1524	0.1499	0.1451	0.1348	0.1419	0.1402
10	5	9	0.3809	0.1470	0.1463	0.1421	0.1354	0.1340	0.1308
10	5	25	0.3812	0.1451	0.1440	0.1401	0.1355	0.1339	0.1307
10	5	100	0.4675	0.1449	0.1433	0.1398	0.1354	0.1337	0.1305
20	0	9	0.1980	0.1442	0.1421	0.1388	0.1316	0.1335	0.1304
20	0	25	0.3534	0.1440	0.1454	0.1426	0.1316	0.1333	0.1302
20	0	100	0.4871	0.1442	0.1505	0.1484	0.1316	0.1329	0.1298
20	3	1	0.5611	0.1596	0.2614	0.2981	0.2079	0.3228	0.4302
20	3	9	0.4948	0.1455	0.1437	0.1398	0.1312	0.1326	0.1296
20	3	25	0.4977	0.1444	0.1449	0.1419	0.1315	0.1334	0.1301
20	3	100	0.5036	0.1440	0.1499	0.1479	0.1317	0.1330	0.1299
20	5	1	0.6101	0.1486	0.2497	0.3910	0.1325	0.3268	0.6372
20	5	9	0.6776	0.1464	0.1444	0.1403	0.1308	0.1319	0.1290
20	5	25	0.5911	0.1447	0.1444	0.1411	0.1314	0.1333	0.1300
20	5	100	0.5146	0.1443	0.1494	0.1472	0.1317	0.1330	0.1298
40	0	9	0.2487	0.1391	0.1400	0.1353	0.1476	0.1465	0.2029
40	0	25	0.4247	0.1390	0.1397	0.1351	0.1483	0.1475	0.3767
40	0	100	0.4949	0.1391	0.1398	0.1352	0.1487	0.1479	0.4909
40	3	1	0.7165	0.4905	0.5410	0.6529	0.5432	0.5722	0.6814
40	3	9	0.7623	0.1383	0.1382	0.1334	0.1733	0.1798	0.5195
40	3	25	0.5995	0.1387	0.1395	0.1345	0.1540	0.1546	0.4967
40	3	100	0.5101	0.1389	0.1398	0.1351	0.1495	0.1488	0.5015
40	5	1	0.6793	0.2704	0.5172	0.6895	0.4747	0.5486	0.7132
40	5	9	0.8173	0.1376	0.1377	0.1329	0.2155	0.2345	0.7144
40	5	25	0.6701	0.1385	0.1388	0.1337	0.1599	0.1625	0.5800
40	5	100	0.5195	0.1388	0.1398	0.1350	0.1503	0.1501	0.5089

Table 10
Summary on misclassification rate for sample size (50,50) at $d=6$

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.4725	0.1581	0.1728	0.1830	0.1651	0.1787	0.2830
Std.Dev.	0.1664	0.0618	0.0930	0.1318	0.0885	0.1062	0.2104
Min	0.1214	0.1376	0.1377	0.1329	0.1308	0.1319	0.1290
Max	0.8173	0.4905	0.5410	0.6895	0.5432	0.5722	0.7144
Range	0.6959	0.3529	0.4033	0.5566	0.4124	0.4403	0.5854

Table 11
Misclassification rate for sample size (100,100) at $d=2$

ε	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.2420	0.2482	0.2462	0.2452	0.2461	0.2452	0.2442
10	0	9	0.2587	0.2482	0.2492	0.2488	0.2456	0.2454	0.2446
10	0	25	0.3447	0.2481	0.2563	0.2588	0.2455	0.2460	0.2459
10	0	100	0.4800	0.2481	0.2689	0.2773	0.2455	0.2469	0.2475
10	3	1	0.2741	0.2531	0.2548	0.2522	0.2471	0.2478	0.2464
10	3	9	0.3270	0.2484	0.2511	0.2506	0.2457	0.2458	0.2450
10	3	25	0.4234	0.2482	0.2569	0.2607	0.2456	0.2460	0.2460
10	3	100	0.4929	0.2481	0.2696	0.2790	0.2455	0.2469	0.2474
10	5	1	0.5010	0.2533	0.2594	0.2563	0.2467	0.2467	0.2454
10	5	9	0.4804	0.2490	0.2523	0.2522	0.2458	0.2460	0.2451
10	5	25	0.4917	0.2483	0.2567	0.2609	0.2456	0.2460	0.2459
10	5	100	0.5012	0.2482	0.2698	0.2797	0.2455	0.2469	0.2474
20	0	9	0.2745	0.2479	0.2490	0.2493	0.2446	0.2453	0.2442
20	0	25	0.3929	0.2479	0.2549	0.2580	0.2443	0.2452	0.2439
20	0	100	0.4896	0.2480	0.2617	0.2668	0.2443	0.2451	0.2439
20	3	1	0.6542	0.2573	0.2628	0.2615	0.2528	0.2650	0.2732
20	3	9	0.5678	0.2488	0.2498	0.2500	0.2446	0.2454	0.2442
20	3	25	0.5237	0.2481	0.2537	0.2575	0.2443	0.2453	0.2440
20	3	100	0.5042	0.2481	0.2617	0.2678	0.2443	0.2452	0.2439
20	5	1	0.7124	0.2554	0.2577	0.2548	0.2442	0.2584	0.2931
20	5	9	0.7158	0.2499	0.2493	0.2497	0.2446	0.2454	0.2442
20	5	25	0.6061	0.2483	0.2521	0.2557	0.2443	0.2454	0.2439
20	5	100	0.5144	0.2481	0.2613	0.2678	0.2443	0.2453	0.2440
40	0	9	0.3063	0.2468	0.2470	0.2461	0.2476	0.2476	0.2645
40	0	25	0.4346	0.2463	0.2471	0.2464	0.2476	0.2478	0.3655
40	0	100	0.4940	0.2462	0.2469	0.2479	0.2474	0.2476	0.4788
40	3	1	0.7442	0.4503	0.5969	0.7147	0.5946	0.6475	0.7236
40	3	9	0.7162	0.2471	0.2472	0.2471	0.2503	0.2523	0.4543
40	3	25	0.5867	0.2463	0.2472	0.2463	0.2483	0.2490	0.4791
40	3	100	0.5076	0.2462	0.2470	0.2480	0.2475	0.2479	0.4949
40	5	1	0.7389	0.2605	0.5148	0.7231	0.4514	0.6276	0.7312
40	5	9	0.7372	0.2470	0.2481	0.2561	0.2554	0.2611	0.6729
40	5	25	0.6453	0.2464	0.2472	0.2468	0.2491	0.2500	0.5635
40	5	100	0.5159	0.2462	0.2469	0.2485	0.2478	0.2481	0.5065

Table 12
Summary on misclassification rate for sample size (100,100) at $d=2$

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.5059	0.2549	0.2718	0.2833	0.2628	0.2710	0.3367
Std.Dev.	0.1456	0.0342	0.0720	0.1093	0.0674	0.0918	0.1501
Min	0.2420	0.2462	0.2462	0.2452	0.2442	0.2451	0.2439
Max	0.7442	0.4503	0.5969	0.7231	0.5946	0.6475	0.7312
Range	0.5022	0.2041	0.3507	0.4779	0.3504	0.4024	0.4873

Table 13
Misclassification rate for sample size (100,100) at $d=6$

ε	μ	ω	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
0	0	1	0.1157	0.1279	0.1261	0.1243	0.1235	0.1228	0.1213
10	0	9	0.1505	0.1274	0.1270	0.1254	0.1222	0.1220	0.1204
10	0	25	0.2252	0.1275	0.1294	0.1282	0.1223	0.1221	0.1207
10	0	100	0.4310	0.1275	0.1382	0.1384	0.1223	0.1228	0.1217
10	3	1	0.2740	0.1315	0.1316	0.1287	0.1224	0.1313	0.1288
10	3	9	0.2414	0.1279	0.1282	0.1263	0.1223	0.1223	0.1206
10	3	25	0.3142	0.1274	0.1295	0.1281	0.1223	0.1223	0.1208
10	3	100	0.4562	0.1274	0.1380	0.1383	0.1223	0.1229	0.1218
10	5	1	0.4991	0.1316	0.1301	0.1274	0.1222	0.1256	0.1244
10	5	9	0.4000	0.1285	0.1290	0.1267	0.1223	0.1221	0.1204
10	5	25	0.4072	0.1275	0.1294	0.1278	0.1223	0.1223	0.1208
10	5	100	0.4736	0.1274	0.1377	0.1380	0.1224	0.1229	0.1217
20	0	9	0.1587	0.1273	0.1263	0.1247	0.1213	0.1225	0.1207
20	0	25	0.2921	0.1273	0.1298	0.1292	0.1213	0.1223	0.1204
20	0	100	0.4684	0.1272	0.1357	0.1355	0.1212	0.1219	0.1202
20	3	1	0.5866	0.1367	0.2081	0.2341	0.1783	0.2737	0.3957
20	3	9	0.5381	0.1280	0.1266	0.1248	0.1211	0.1221	0.1202
20	3	25	0.5044	0.1274	0.1282	0.1271	0.1213	0.1223	0.1204
20	3	100	0.4960	0.1274	0.1351	0.1348	0.1212	0.1220	0.1202
20	5	1	0.6526	0.1298	0.1968	0.3281	0.1216	0.2800	0.6842
20	5	9	0.7669	0.1288	0.1268	0.1249	0.1208	0.1217	0.1198
20	5	25	0.6490	0.1276	0.1267	0.1253	0.1212	0.1224	0.1203
20	5	100	0.5147	0.1274	0.1343	0.1339	0.1213	0.1220	0.1202
40	0	9	0.1893	0.1250	0.1255	0.1232	0.1285	0.1281	0.1585
40	0	25	0.3682	0.1250	0.1255	0.1232	0.1290	0.1287	0.3053
40	0	100	0.4853	0.1250	0.1253	0.1229	0.1291	0.1288	0.4745
40	3	1	0.7677	0.5069	0.5553	0.6932	0.5673	0.5954	0.7254
40	3	9	0.8194	0.1245	0.1238	0.1216	0.1445	0.1484	0.5534
40	3	25	0.6495	0.1249	0.1249	0.1225	0.1323	0.1326	0.5001
40	3	100	0.5128	0.1250	0.1253	0.1230	0.1296	0.1295	0.4960
40	5	1	0.7300	0.2669	0.5338	0.7277	0.4822	0.5756	0.7511
40	5	9	0.8526	0.1240	0.1235	0.1213	0.1781	0.1905	0.7862
40	5	25	0.7379	0.1248	0.1243	0.1219	0.1356	0.1373	0.6380
40	5	100	0.5306	0.1250	0.1253	0.1229	0.1300	0.1300	0.5119

Table 14
Summary on misclassification rate for sample size (100,100) at $d=6$

	CLDR	$R_{(\alpha=0.4)}$	$T_{(\alpha=0.4)}$	$M_{(\alpha=0.4)}$	$R_{(\alpha=0.2)}$	$T_{(\alpha=0.2)}$	$M_{(\alpha=0.2)}$
Mean	0.4782	0.1427	0.1577	0.1707	0.1513	0.1635	0.2802
Std.Dev.	0.1985	0.0677	0.0984	0.1403	0.0949	0.1118	0.2293
Min	0.1157	0.1240	0.1235	0.1213	0.1208	0.1217	0.1198
Max	0.8526	0.5069	0.5553	0.7277	0.5673	0.5954	0.7862
Range	0.7369	0.3829	0.4318	0.6064	0.4465	0.4737	0.6664

$\omega=1$, $\alpha=20\%$). On the contrary, in the 49 and 45 instances where $RLDR_T$ outperforms $RLDR_R$, only a small difference margin with maximum of 0.0060 (Table 5, $n=20$, $d=6$, $\varepsilon=10\%$, $\mu=0$, $\omega=25$, $\alpha=40\%$) and 0.0019 (Table 9, $n=50$, $d=6$, $\varepsilon=10\%$, $\mu=0$, $\omega=100$, $\alpha=20\%$) are observed. It implied that employing robust MSD yields better results than the classical MSD during the distance-based trimming procedure.

Moving on to the comparison between $RLDR_R$ and $RLDR_M$, the performance of both robust classifiers in terms of occasions are comparable in both $\alpha=40\%$ and $\alpha=20\%$. A total of 88 and 99 per 198 occasions were found where $RLDR_M$ outperforms $RLDR_R$ for trimming percentages of 40% and 20%, respectively. When 40% trimming was applied, the maximum difference in misclassification rates when $RLDR_M$ achieved lower misclassification was 0.0139 (Table 5, $n=20$, $d=6$, $\varepsilon=20\%$, $\mu=0$, $\omega=100$), while $RLDR_R$ outperforms with a maximum difference of 0.4626 (Table 11, $n=100$, $d=2$, $\varepsilon=40\%$, $\mu=5$, $\omega=1$). On the other hand, in 20% trimming category, the maximum differences in misclassification rates are 0.0073 ($RLDR_M < RLDR_R$; Table 5, $n=20$, $d=6$, $\varepsilon=10\%$, $\mu=5$, $\omega=25$) and 0.6081 ($RLDR_M > RLDR_R$; Table 13, $n=100$, $d=6$, $\varepsilon=40\%$, $\mu=5$, $\omega=9$). The comparison between these two robust classifiers shows that despite the high number of occasions where $RLDR_M$ is better, the misclassification rates yielded are not substantially lower than $RLDR_R$, whereas $RLDR_R$ yielded noticeably better performance.

In addition to evaluating the classifiers' performance, the descriptive statistics in Tables 4, 6, 8, 10, 12, and 14 can help the classifiers' assessment. Even though CLDR was the best classifier in non-contaminated data distribution, poor overall descriptive statistics are expected. CLDR had the highest mean of misclassification rates. Notably, regardless of trimming percentages, $RLDR_R$ achieved the lowest mean, standard deviation, and range compared to the other robust classifiers. The comparisons of descriptive statistics between the three classifiers in both trimming categories showed that $RLDR_R$ yielded averagely lower and relatively consistent misclassifications.

In summary, applying Robust MSD in distance-based trimming procedure to the method in Pang et al. (2019) showed improvement with a much lower misclassification rate and stable performance presented in $RLDR_R$.

CONCLUSION

The established THCM considered contamination from both location and covariance aspects. Based on the simulation results in this study, it is noticeable that the increment of location and/or covariance contaminations led to an increment in misclassification rates in the classical rule. Despite CLDR showing optimal results when dealing with non-contaminated data, the CLDR should be applied with caution as non-contaminated data are hardly attainable in real-life applications. On the other hand, the simulation study has shown promising outcomes from $RLDR_R$ as it outperforms CLDR, $RLDR_T$, and $RLDR_M$ when

dealing with outliers in 2- and 6-dimensional contaminated distributions where $n=20, 50,$ and 100. One should note that the performance of a classifier should not be determined by the number of occasions it bests the others but should also consider the overall performance. Based on the performance of $RLDR_R$ in the simulation study, researchers could consider employing $RLDR_R$ if data contamination is suspected.

The main goal of this study is to determine whether employing robust MSD in distance-based trimmed median ($\widehat{M}_{(d,\alpha)}$) can yield a lower misclassification rate in constructing a robust classifier. It is shown that the robust classifier employed $\widehat{M}_{(MSD_R,\alpha)}$ noticeably outperformed $\widehat{M}_{(MSD,\alpha)}$ with a lower and consistent misclassification rate. However, the study was limited to comparing classical MSD (with \bar{x} and S) and robust MSD (with \widehat{M} and S_R) in $\widehat{M}_{(d,\alpha)}$. Therefore, future studies can explore ways to robustify MSD in the distance-based trimmed median.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Higher Education (MoHE) of Malaysia through Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UUM/02/8). The authors also want to express gratitude toward Universiti Utara Malaysia for the support and facilities provided to conduct this study.

REFERENCES

- Abu-Shawiesh, M. O., & Abdullah, M. B. (2001). A new robust bivariate control chart for location. *Communications in Statistics-Simulation and Computation*, 30(3), 513-529. <https://doi.org/10.1081/SAC-100105076>
- Alloway, J. J., & Raghavachari, M. (1990). Multivariate control charts based on trimmed means. In *ASQC Quality Congress Transactions* (Vol. 44, pp. 449-453). American Society for Quality Control.
- Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(3), 231-237. <https://doi.org/10.2307/2346896>
- Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *Canadian Journal of Statistics*, 29(3), 473-493. <https://doi.org/10.2307/3316042>
- Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). Robust statistical estimation. In T. D. Little (Ed), *The Oxford Handbook of Quantitative Methods: Foundations* (Vol. 1, pp. 388-406). Oxford University Press.
- Hadi, A. S., Imon, A. H. M., & Werner, M. (2009). Detection of outliers. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 57-70. <https://doi.org/10.1002/wics.6>
- He, X., & Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis*, 72(2), 151-162. <https://doi.org/10.1006/jmva.1999.1857>

- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73-101.
- Hubert, M., & Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics & Data Analysis*, 45(2), 301-320. [https://doi.org/10.1016/S0167-9473\(02\)00299-2](https://doi.org/10.1016/S0167-9473(02)00299-2)
- Johnson, R. A., & Wichern, D. W. (2013). *Applied Multivariate Statistical Analysis* (6th Ed.). Pearson.
- Lim, Y. F., Yahaya, S. S. S., & Ali, H. (2016). Winsorization on linear discriminant analysis. In *AIP Conference Proceedings* (Vol. 1782, No. 1, p. 050010). AIP Publishing LLC. <https://doi.org/10.1063/1.4966100>
- Pang, Y. S., Ahad, N. A., & Yahaya, S. S. S. (2021). Robust multivariate location estimation in the existence of casewise and cellwise outliers. *Mathematics and Statistics*, 9(5), 653-663. <https://doi.org/10.13189/ms.2021.090505>
- Pang, Y. S., Ahad, N. A., Yahaya, S. S. S., & Lim, Y. F. (2019). Robust linear discriminant rule using novel distance-based trimming procedure. *Journal of Advanced Research in Dynamical and Control Systems*, 11(05-Special Issue), 969-978.
- Penny, K. I., & Jolliffe, I. T. (2001). A comparison of multivariate outlier detection methods for clinical laboratory safety data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(3), 295-307. <https://doi.org/10.1111/1467-9884.00279>
- Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, 8, 283-297.
- Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212-223. <https://doi.org/10.1080/00401706.1999.10485670>
- Rousseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633-639. <https://doi.org/10.1080/01621459.1990.10474920>
- Todorov, V., & Pires, A. M. (2007). Comparative performance of several robust linear discriminant analysis methods. *REVSTAT Statistical Journal*, 5, 63-83. <https://doi.org/10.2307/3316042>
- Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.
- Yahaya, S. S. S., Lim, Y. F., Ali, H., & Omar, Z. (2016). Robust linear discriminant analysis with automatic trimmed mean. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(10), 1-3.

Design Guidelines for Vegetative Approach Application to Mitigate Riverbank Landslides in Mangkang Wetan and Mangunharjo Villages, Semarang Coastal Area, Indonesia

Santy Paulla Dewi* and Rina Kurniati

Department of Urban and Regional Planning, Faculty of Engineering, Universitas Diponegoro, Jl. Prof. Soedarto, SH, Tembalang, Semarang, 50275, Central Java, Indonesia

ABSTRACT

The Semarang coastal area is vulnerable to floods, land subsidence, and landslides, including Mangkang Wetan and Mangunharjo villages. The community villages struggle with flooding every year because of the riverbank landslides. The government's efforts to improve the critical embankments have not yet reached these villages. The community prevents landslides by a conventional method, such as using sacks filled with sand. However, this effort did not make it significantly effective. Therefore, this study proposes a vegetative approach as an alternative effort to reduce the landslide possibility and provide directions for riverbank management. This approach emphasizes soil conservation by applying a suitable plant type to reduce the landslides possibility. Determining the design guidelines for the vegetative approach starts by analyzing the physical condition of the village and the river to specify zone types; then continues by analysis of the appropriate vegetation types. Afterward, determine the design direction for each zone and recommend a suitable type of vegetation. The analysis results show two strategies related to the possible application of a vegetative approach. The do-nothing strategy is for locations that cannot physically intervene because there is no river border. In contrast, the do something strategy

is applied to locations that physically allow for a vegetative approach at the river border. However, no single effort can prevent landslides, so it needs to combine mechanical and vegetative approaches that adjust to the physical characteristic of the river and the soil type.

ARTICLE INFO

Article history:

Received: 05 February 2022

Accepted: 21 April 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.06>

E-mail addresses:

santy.paulla.dewi@pwk.undip.ac.id (Santy Paulla Dewi)

rina.kurniati@pwk.undip.ac.id (Rina Kurniati)

* Corresponding author

Keywords: Design, landslide, riverbank, vegetative approach

INTRODUCTION

The increasing population and their activities have led to escalated land conversion in urban areas, particularly from rice fields to build areas (Ducey et al., 2018; Oliver & Thomas, 2014), which further raises the risk of disasters, especially hydro-geomorphological disasters (Barros et al., 2021; Sidle & Onda, 2004; Su et al., 2021). Hydrological processes with natural landscapes and geomorphic processes with surface and subsurface waters in temporal and spatial aspects define hydro-geomorphology (Barros et al., 2021; Sidle & Onda, 2004). One of the hydro-geomorphological disasters is riverbank landslides caused by water erosion from upstream areas.

Riverbank landslides include shallow landslides that occur due to rainfall, where the duration of the rain is moderate to high, which is between 1–15 days with moderate to long rain intensity, as well as in an area with soil types that are difficult to hold water (Persichillo et al., 2017; Zêzere et al., 2005). Shallow landslides are difficult to detect early but can be mitigated through nature-based solutions, such as using the vegetation (de Jesús Arce-Mojica et al., 2019; Kalsnes & Capobianco, 2022; Persichillo et al., 2017). Shallow landslides are defined as landslides caused by the process of saturation of the soil layer due to heavy rains of short duration, involving the soil layer near the surface, move quickly, can flow, and usually occurs in flow landslides, slides, and collapses (the Indonesian Research Institute-LIPI). However, there are several efforts to overcome riverbank landslides, and the most common strategy is to develop an embankment made of a mixture of river stone, cement, and sand. These strategies are considered adequate for protecting and maintaining soil stability and preventing landslides, but it requires a high cost and does not have a minimal environmental impact (Bariteau et al., 2013).

Handling shallow landslides that have a significant long-term impact is applying suitable plants and involving the community in the process (Hostettler et al., 2019). Likewise, the President of the Republic of Indonesia, Joko Widodo, has encouraged and facilitated the local government and community to adopt the vegetative approach to mitigate. The vegetative approach has some benefits such as increasing water absorption using plant roots, maintaining soil stability, low cost in implementation, allowing the community to be involved in the process and sustaining to empower the community. In addition, this approach is in line with the recommendations of the Sendai Declaration related to Disaster Risk Reduction, which prioritizes ecosystem-based solutions. The Indonesian government introduced and promoted one type of plant to overcome shallow erosions: vetiver grass. This plant is one of the most widely developed plants, which can mitigate landslides by covering, protecting, and stabilizing the environment (Ghosh & Bhattacharya, 2018).

The Semarang City Government also obeyed the central government's instructions to use vetiver grass to reduce landslides. However, Semarang is highly vulnerable to disasters (Li et al., 2016; Sherly et al., 2015), including a high potential danger of landslides (National

Disaster Management Agency of the Republic of Indonesia, 2015). Indeed, the most common disaster in Semarang City in 2020 was landslides in hilly and steeper slope areas (southern and left part of Semarang) and in the downstream area (coastal areas or north part of Semarang). This study focuses on shallow landslides in coastal areas, particularly in Mangkang Wetan and Mangunharjo villages, where the riverbank landslides were followed by flash floods that inundated community settlements. The existing disaster management strategies emphasize protecting coastal areas by developing mangrove conservation (Semarang City Government, 2016), whereas other efforts concerning shallow landslides handling have not been further exposed. Conventionally, the community independently makes deposits from sand-filled sacks as a prevention effort.

However, vetiver system implementation faces challenges such as lacking community support because they do not understand the vegetation's role in mitigating disaster risk. Likewise, the community comprehending the riverbank's role as river protection is lacking, so the river border is used as a garbage disposal, even as a place to live (permanent housing). Hence, it is essential to identify the possibility of a vegetative approach implementation to mitigate landslides in the coastal area of Semarang, especially in Mangkang Wetan and Mangunharjo Villages. Furthermore, the operationalization of the vegetative approach is also crucial; it discusses the plant types and riverbank arrangement, including the application of the vegetative approach and the appropriate design.

MATERIALS AND METHOD

Research Study Area

The research location is in the coastal landslides-prone villages, Mangkang Wetan and Mangunharjo Villages, Tugu District (Figure 1). Geographically, the Mangunharjo Village is passed by the Beringin River, which reaches 0.03 km², whereas the Beringin River traverses the Mangkang Wetan Village, which reaches 0.26 km² (Semarang City Government, 2014). The width of the Beringin River in the upstream area is 20 meters, or more extensive than in the downstream area (Mangunharjo Village), and reaches 9.5 meters (Indrayati & Hikmah, 2018). Therefore, the Beringin River has a distinct classification; it is classified as a big river in the upstream area, while in the estuary area, it is classified as a small river (Maryono, 2009).

Every rainy season, the water of the Beringin River always overflows to inundate hundreds of houses of residents of these two villages. The flood triggered by high rainfall prevents the river from accommodating the water discharge and causes embankment landslides. The most significant impact of flooding occurs in the settlement located in the riverbank area or at 1.5–2 meters from the river. Moreover, the riverbank has no buffers that are left overgrown with weeds. Regarding the physical characteristics, this river has a flat slope and approaches sea level elevation, so it is difficult to drain water during high tide conditions (Figure 2).

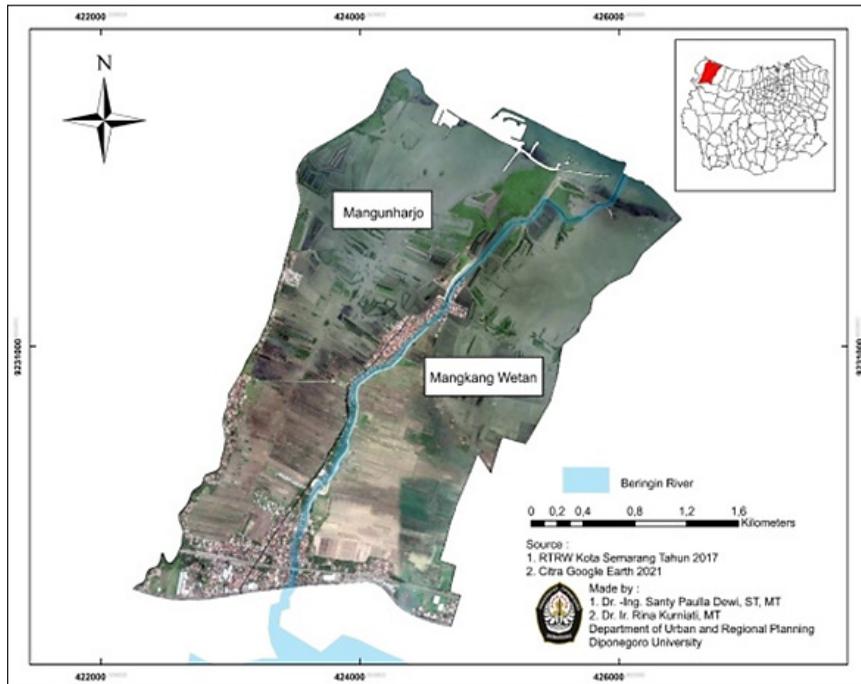


Figure 1. Research location

Source. Semarang City Drainage System Master Plan 2011-2031 (Semarang City Government, 2014)

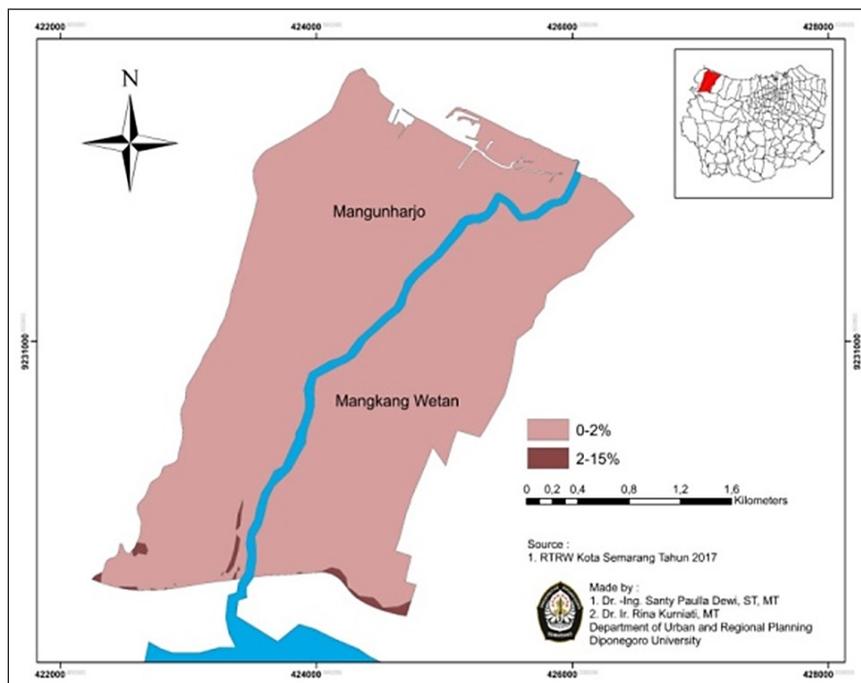


Figure 2. Slope of research location

Source. Spatial plan of Semarang City 2011-2031, (Semarang City Government, 2017)

Research Method

This study reveals vegetative approach implementation that appropriates the physical characteristic of the river to reduce the risk of riverbank landslides. The research method used is spatial by overlaying data related to the area's physical characteristics, such as soil type, slope, and land use, to determine land-use changes (size and the distribution). In addition, the vegetative approach design is formulated by comparing the government regulation on border and riverbank arrangement, green open space alignment, and river normalization.

The data used in this study are mainly collected from field observations, especially those related to the physical condition of the research area, and secondary data from government agencies such as data related to spatial planning and landslide management policies obtained from the Semarang Central Bureau of Statistics, Semarang Development Planning Agency, and Regional Disaster Management Agency (BPBD) of Semarang City. Following are the research process stages.

Identification of Physical Characteristics of the Research Area

The physical characteristics revealed from the land use, soil type, rainfall, slope, disaster mitigation reports, settlement, and dynamics data gained from government institution websites such as Semarang Central Bureau of Statistics reports or planning or development documents from Semarang Development Planning Agency and BPBD of Semarang City. The analysis technique is a spatial method that produces a map of the landslides-prone area and its level.

Analysis of Vegetative Approach

This analysis aims to determine the appropriate vegetation types for mitigating riverbank landslides. This stage was conducted by identifying various kinds of vegetation which can withstand landslides based on previous research (Adhitya et al., 2016; Herman et al., 2013; Izzul et al., 2020; Mentari et al., 2018; Risdiyanto, 2019). The vegetation analyzed is the type of vegetation that can grow in Indonesia, which means compatible to plant in the research area. Afterward, reviewing the kind of vegetation with the research area's soil type, slopes, and fertility. Besides, other economic considerations have also been considered, such as the vegetation price, maintenance aspect (whether the vegetation needs extra care), and availability (which relates to how the community gets plant seeds).

Analysis of Riverbank Design Guidelines

Determination of riverbank design guidelines is defined from the government regulations on riverbank arrangement and urban design standards, which are compared with the existing

situation. The current condition of the research area showed that the riverbank was left with no specific vegetation. Meanwhile, the government’s plan to normalize the river is still in progress due to land acquisition constraints; unclear the fixed realization of river normalization or to what extent the normalization is; whether all along the river or certain parts of the land that has been successfully acquired.

Therefore, according to public policy theory, there are two scenarios: do nothing and do something. The do-nothing scenario is also referred to as the status quo when the government has not yet successfully implemented the Beringin River normalization plan. Meanwhile, the do-something scenario represents that the design is adjusted to the government’s plan combined with the study results of appropriate vegetation types.

RESULTS

Physical Characteristics

The Beringin River, which crosses Mangkang Wetan and Mangunharjo Villages, includes a river with an embankment in urban areas. Therefore, according to the Minister of Public Works and Public Housing Regulation, the border at the research location is at least three meters from the outer edge of the embankment border. However, there are some points in the Beringin watersheds where the width of the riverbank is less than three meters (Figure 3, the red number represents the area with a range of fewer than three meters). These locations are dangerous and experienced landslides because of the high-water runoff from upstream. The landslides include the type C zone, with the soil type being the Mediterranean and

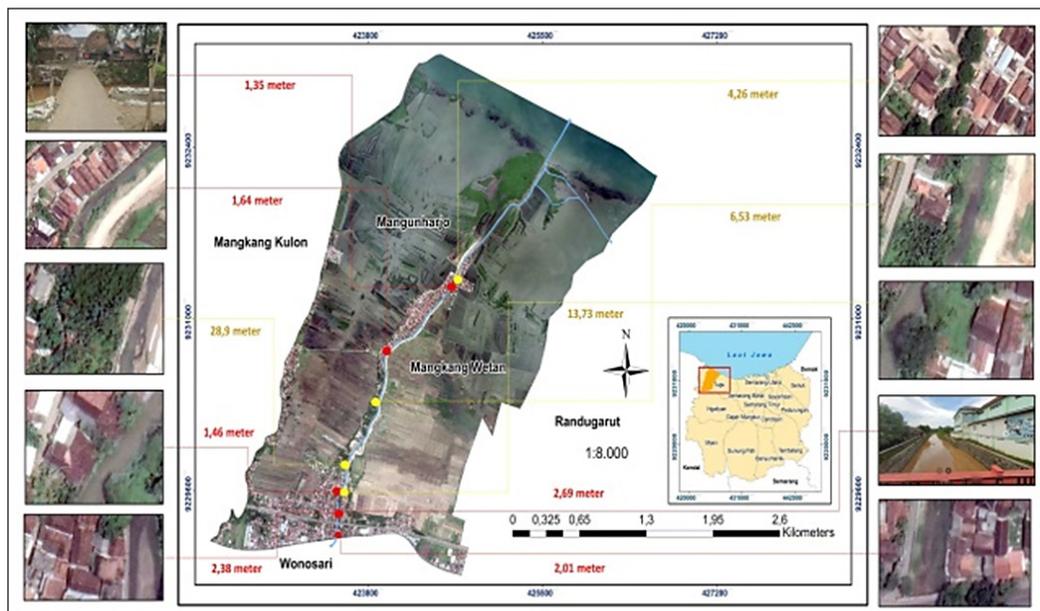


Figure 3. Riverbank existing situation (range between river and building or other functions)

Another problem in the Beringin watershed is sedimentation, where the total sediment deposition reaches more than 20,000 m³/year, making the river shallow and narrow (Semarang City Government, 2014). Likewise, the growth of wild plants causes the river wide narrower. A high-intensity rain makes runoff that flows downstream cannot accommodate and leads to flooding and erosion. Based on the Semarang City BPBD report in 2020, the Beringin riverbank landslide was caused by the medium intensity of rainfall which reached 22.10 mm/month to 393.20 mm/month. It increased the volume of the Beringin River water, reaching 60 -70 cm, and caused the water to overflow into the settlements.

Vegetative Approach Application

The vegetative approach is one of the solutions to overcome riverbank landslides which is considered more accessible and economically than others, such as the concrete method (Leknoi & Likitlersuang, 2020). Likewise, this approach aims to reduce or delay water runoff and serve as a barrier to overcoming landslides. In addition, this method can also hold back sedimentation and reduce pollution carried by water to maintain water quality and the environment (Lobo & Bonilla, 2017). There are several techniques in the vegetation approach: reforestation, agroforestry consisting of alley cropping, strip cropping, grass strips, rows of crop residues, cover crops, and the application of cropping patterns, including crop rotation intercropping and relay cropping (Subagyo et al., 2003).

The Beringin riverbank was overgrown by wild plants that did not function optimally. There has been no reforestation effort on the river border, so it has not been planted with particular vegetation that can mitigate landslides. Only weeds, bushes, and banana trees are planted randomly by the community. Hence, vegetation on the riverbank in the coastal areas is vulnerable to contamination by mud and garbage carried by water runoff during floods. The selection of greenery in a vegetative approach may consider the plants' capacity to withstand the cliff structure from landslides, which refers to the following criteria:

- Vegetation that can protect embankments and riverbanks from cliff erosion and landslides
- Cover crops that can restrain the water runoff rate and can be a barrier
- Plants that have the robustness and produce leaf litter that easily falls off
- The selection of vegetation types considers the natural physical characteristics (soil type, water flow, microclimate temperature, and sunlight), economic aspects (affordable prices, easy to obtain, provide benefits), and social aspects (plants that are familiar to the community and easy to cultivate).

The following are vegetation that can protect the soil and mitigate landslides that differ from the characteristic (Table 1).

Table 1
Characteristics of vegetation

No	Vegetation	Strength	Weakness
1	Candlenut (<i>Aleurites moluccana</i>)	<ul style="list-style-type: none"> • Deeply penetrating taproot with multi-branched roots • Can thrive on Latosol, Podzolic, and Andosol soil types 	<ul style="list-style-type: none"> • It needs three years to grow optimal • Not suitable for the type of soil in the study area (alluvial)
2	Vetiver grass (<i>Chrysopogon zizanioides</i>)	<ul style="list-style-type: none"> • It stabilizes the soil because the roots reach three m deep • Reducing pollution (chemical and medical waste, mercury) • Easy to maintain 	<ul style="list-style-type: none"> • Short lifetime • Need to combine with other plants, mainly with land cover
3	Mindi (<i>Melia azedarach</i>)	<ul style="list-style-type: none"> • Deeply penetrating taproot with multi-branched roots 	<ul style="list-style-type: none"> • Hard to obtain seeds, considering the wood has a high economic value, so that very demanding • Vulnerability to insects and plant pests
4	Yellow bamboo (<i>Bambusa vulgaris</i>)	<ul style="list-style-type: none"> • Its fibrous roots stabilize the soil • Easy to maintain • Can grow on all types of soil • Low-cost investment 	<ul style="list-style-type: none"> • It needs five years of planting • It requires a large planting medium
5	Pandanus (<i>Pandanus amaryllifolius</i>)	<ul style="list-style-type: none"> • It has a large and long roots • The sea pandanus can break the waves and become a nesting place for turtles 	<ul style="list-style-type: none"> • Suitable if applied to the beach or the sea
6	Banana tree (<i>Mussa sp</i>)	<ul style="list-style-type: none"> • Retain the soil moisture • Fibrous root system (rhizome roots) 	<ul style="list-style-type: none"> • Requires care and fertilization to grow optimally • It is easy to collapse because the root does not penetrate the soil
7	Guava tree (<i>Psidium guajava l.</i>)	<ul style="list-style-type: none"> • It has a strong root system (taproot and fibrous) • It can grow in various soil types 	<ul style="list-style-type: none"> • Requires a large planting medium

Determining vegetation characteristics needs to consider the plants growing irregularly along the riverbank. Therefore, it concludes that the vegetation types can be divided into conservation plants and riparian vegetation. Conservation plants play a role in revegetating river areas to be planted, especially in border areas. Meanwhile, riparian vegetation is the majority of plants that can survive in a water-submerged area. The following is an example of flexible riparian vegetation with a robust root system, such as vetiver grass, yellow bamboo, and guava. This vegetation can stabilize the soil and protect the riverbanks by reducing surface flow velocity when runoff and reducing pollution. In addition, the soil type has also determined the implementation of the vegetative approach, such as alluvial soil, which is included as a fertile soil with a flat slope of 0–2%.

The vegetative approach is addressed to give long-term protection for riverbanks against landslides and provide economic benefits for the community without damaging the environment. For instance, planting guava trees on riverbanks offers several advantages,

such as being a conservation plant to revegetate borders and a commodity plant with an economic value to the community.

Based on the Drainage System Plan, one of the efforts to conserve soil in landslide-prone areas is the reforestation of river borders. However, the recommended plants are primarily trees that take a relatively long time to grow and impact. Therefore, the government currently encourages greening riverbanks by planting vetiver grass or *akar wangi* (the local name of vetiver grass), or *Chrysophogon zizaionide* to mitigate landslides. In Indonesia, the application of vetiver grass as a barrier to cliff landslides has only been introduced and encouraged by the government since 2007, although it was still on a small scale; even until 2010, initial studies or trials were still being conducted in several pilot locations (Kusnandar & Kusminingrum, 2011).

However, *akar wangi*, used as a craft material, has been around for a long time, especially in Bali and Garut. The root fibers are processed into bags, paintings, lampshades, prayer rugs, and others. The processing of vetiver into these crafts has been recognized and recorded as intangible cultural heritages in the domain of traditional crafts skills and proficiency by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia in 2013.

Research from the Agency for the Assessment and Application of Technology showed that shoot and root growth and the ability to absorb pollution of vetiver grass were better when planted with river water media than groundwater media (Komarawidjaja & Garno, 2016). Therefore, the utilization of vetiver as an implementation of a vegetative approach to preventing cliff landslides is appropriate. In addition, with the government's support through various programs and assistance from the private sector in providing vetiver seeds, implementation in the research location is more likely. Therefore, ecological design is appropriate to mitigate the Beringin River's riverbank landslides (Sittadewi, 2016).

However, vetiver grass has a short lifetime (only reaches one year and needs to be planted the following year again); it requires a continuous and consistent commitment from the government and other parties to provide and facilitate the provision of vetiver seeds (Leknoi & Likitlersuang, 2020; Mondal & Patel, 2020). Furthermore, the growth pattern of vetiver grass tends to be vertical, so it needs to be combined (intercropped) with ground cover plants such as elephant grass and vegetables or bamboo and jute geogrids (Mondal & Patel, 2020). Hence, the local community's role is necessary to realize riverbank conservation. Meanwhile, the plants that economically have great value and can reduce landslide risks are more in the form of trees. However, it needs several years for the tree to absorb water and resist erosions. For instance, Candlenut, especially the sultan candlenut species, can produce oil as an alternative fuel, but it is not suitable for research areas with Mediterranean soil types. In addition, the vegetative approach cannot rely on a single vegetation type; it requires mixing with other plants or other methods.

Moreover, the vegetative approach in the research area is primarily realized in the form of mangrove conservation, whereas reforestation to prevent erosions is still focused on hilly steep slope areas. The vegetative approach is one of the soil conservation techniques besides mechanical and chemical techniques, by utilizing the root system of plants to absorb water, reduce runoff, and prevent soil structure damage. Hence, prevention of planting solid roots and growing fast in landslides is necessary, including in hilly steep slope areas and the lowland (Ministry of Public Works, 2007).

DISCUSSION

Realizing three meters of width for the Beringin riverbank is challenging; on the left side of the river or directly adjacent to the river embankment is a permanent building, whereas on the right side of the river is a rice field area (Figure 3). The landscape arrangement of the riverbank intends to maintain the river sustainability, both ecosystems, and prevent river damage caused by floods and riverbank erosions. The landscape arrangement of the riverbank intends to maintain the river sustainability, both ecosystems, and prevent river damage caused by floods and riverbank landslides.

Floods always occur in coastal areas, inundate settlements and paralyze accessibility to and from Semarang City. Therefore, the Semarang City Government seeks to overcome this by coordinating with other parties. One of them is the normalization of the Beringin River, which in its implementation is handled by the central government through the Pemali Juwana River Area Center for two years starting in 2020 along a length of 7.18 km. This river normalization plan has been around for a long time; starting in 2003, that emphasized land acquisition activities for settlements located in river border areas. The central government funds the river normalization, whereas the Semarang City government takes care of the land acquisition. Therefore, the land acquisition process started in 2003–2020, when 10 ha of land had been completed, and the remaining 8 ha were still being discussed with the community, especially in Mangkang Wetan and Mangunharjo Villages (Semarang City Government, 2021).

Therefore, the normalization of the Beringin River cannot be carried out in the two sub-districts before the land acquisition is finished. Hence, the vegetative approach implementation design at the research site refers to two strategies: do nothing and do something. The do-noting strategy is a situation where the existing alternative solutions have a detrimental impact (Patton et al., 2016). It means that the normalization of the Beringin River cause people will be evicted from their homes that have been occupied for years. In addition, there is a possibility that they will also lose their jobs or increase their expenses to pay for transportation costs from their new place to their workplace. On the other hand, if it remains today, the community will always experience flooding yearly. Meanwhile, the do something strategy refers to river normalization efforts as a solution to overcome flooding and improve environmental quality.

The guidelines for implementing vetiver as landslides barrier issued by the Ministry of Public Works are more intended for road slopes. Meanwhile, there is no guide to applying the vegetative approach on river slopes. Therefore, the formulation of these designs refers to the river's existing condition and the use of the surrounding land and government policies, especially the Law on Provision of Green Spatial Planning and the Semarang City Drainage System Master Plan.

Following are the explanation of do-nothing and do-something strategies:

1. A do-nothing strategy means that there is no intervention in the river. This strategy is applied when considering the land acquisition process, which takes a long time (17 years) but leaves 400 parcels of land unfinished. In this strategy, the design for the vegetative approach follows the existing condition where there are several points where the river border is free from buildings, but there are buildings in some places.

- Typology 1: the segment of the border area's left and right side of Beringin River in the garden, while the distance between the border and the building is equal to or more than three meters (Figure 6). It means that without a river normalization process, the width of the river border in this typology is in accordance with the standard for embankment rivers in urban areas. Although these areas have met

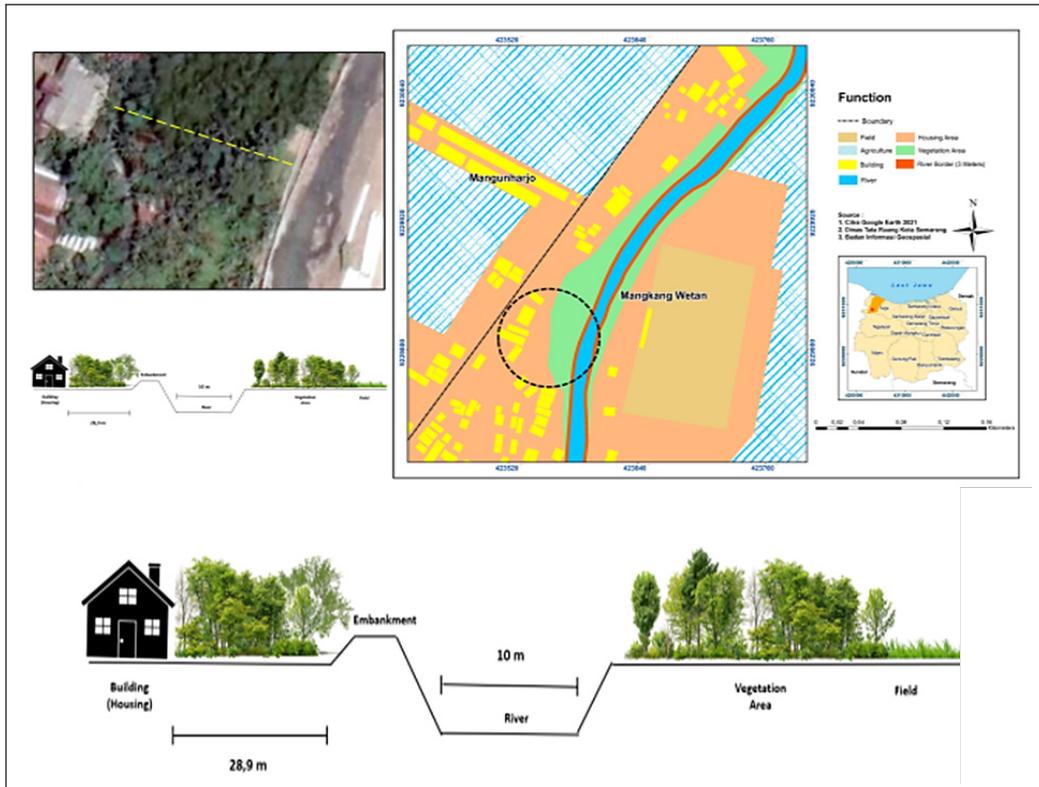


Figure 6. Design guideline for Typology 1

the requirements of the river border standard, it still faces flooding every year due to silting and narrowing of the river width and lack of vegetation buffer. The vegetative approach in this segment aims to mitigate flooding and conserve land, especially riverbanks. Based on the analysis of vegetative approach application, the plant that is possible to apply in this segment is the method of intercropping vetiver grass with elephant grass or bamboo (short term) or kinds of trees (long term). Another tree that can be combined with vetiver grass is the tamarind tree. Historically, this tree has been found in Semarang City, and even then, the toponym of the city's name is also taken from the tamarind tree (the local name is *pohon asem*). In addition, the tamarind tree is one of the plants recommended by the Research and Development Center for Watershed Management Technology as vegetation that can prevent riverbank landslides.

- Typology 2: the left and right segments of the Beringin River border area are adjacent to the buildings (construction), with a distance between the borders of less than three meters (Figure 7). This area is the most affected by floods and landslides. Ideally, buildings on the riverbank need to be displaced, yet it is not easy to do because the community rejected the relocation scheme. Therefore, applying the vegetative approach in this segment adjusts to the existing conditions, where the design direction is combined with a technical method by hardening the embankment with cement. After the cement embankment, it plants vetiver grass and elephant grass as a cover crop. This pavement and composition are plotted in fewer than three meters border areas. However, the construction of river embankments with cement is quite expensive for local people with economic limitations (some are traditional fishers and industrial laborers), so it needs assistance from the government and other parties. If there are no supports, the minimal solution is combining vetiver grass with traditional embankment from sand sacks.

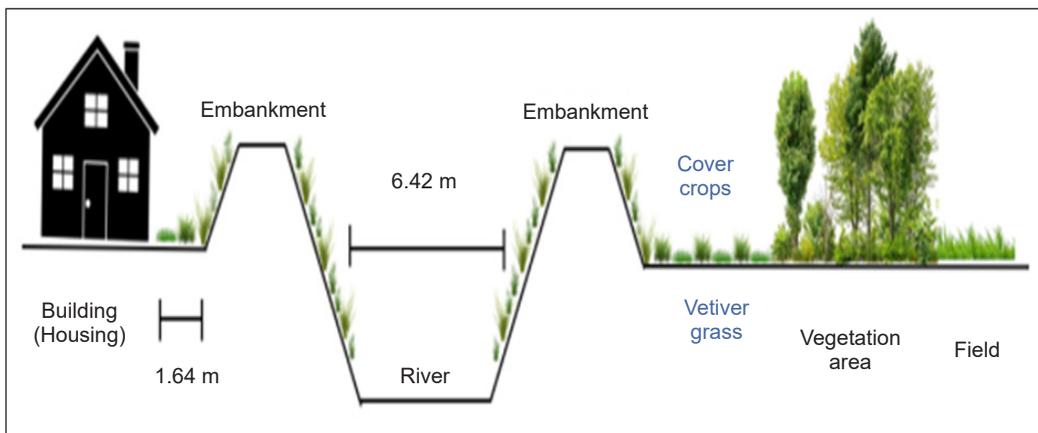


Figure 7. Design guideline for Typology 2

2. Do-something strategy

This strategy means that there is intervention from the government to overcome riverbank landslides by normalizing the river. Therefore, the design of the vegetative approach adjusts to the rules of normalization and reforestation of river borders.

- Typology 3: is a segment of the Beringin Riverbank normalized following government regulations. Riverbank rehabilitation, especially for the small rivers category with a watershed area that is less than or equal to 500 km², is carried out at least 50 meters from the left and right banks of the riverbed along the river channel (The Minister of Environment and Forestry Regulation No. 105 of 2018). The river normalization standard is ideal condition for rehabilitating riverbanks and mitigating landslides and floods. This normalization has been planned since 2020, yet due to the COVID-19 pandemic, this activity has not been carried out.

Hence, the design guideline in this segment considers an ideal pattern, particularly after the river normalization by building a barrier (Figure 8). The vegetative approach implementation utilizes vetiver grass planted after the barrier, intercropped with trees, such as guava and banana trees. Eventually, the community benefits economically from planting these trees by consuming or selling the fruits. River normalization means widening

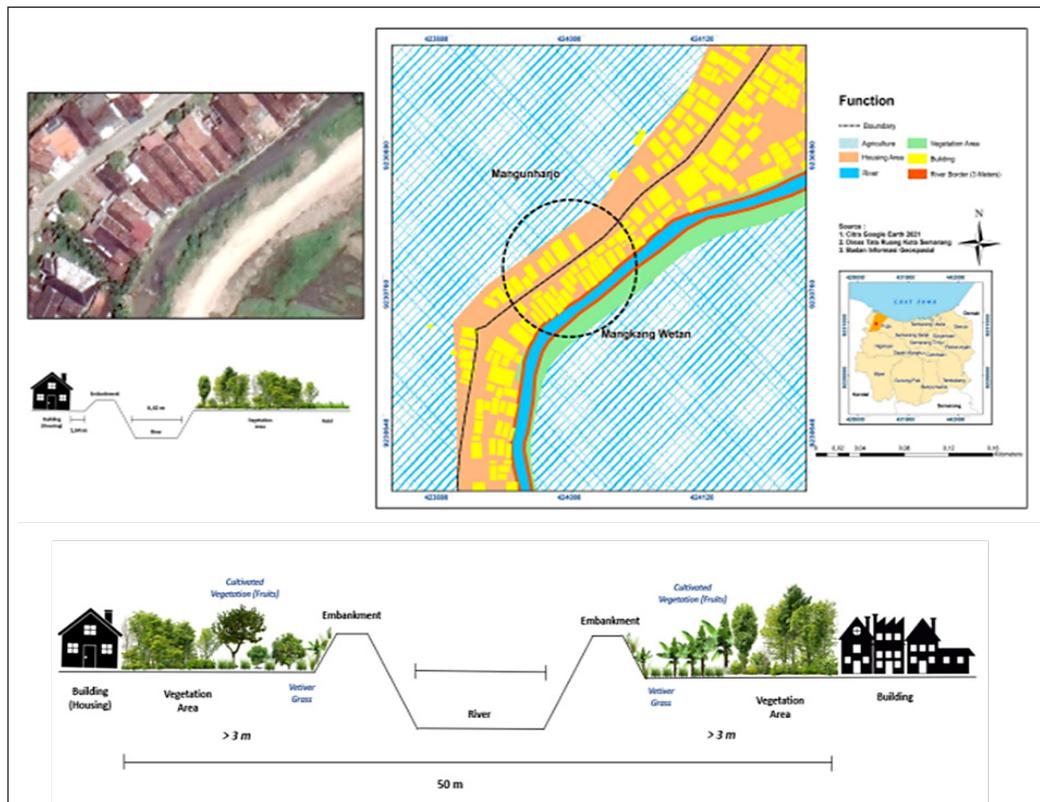


Figure 8. Design guideline for Typology 3

and restoring the cross-section of the river according to its class, improving the slope of the river channel, and strengthening the slope by making embankments, for example, with concrete blocks. The widening of the river is under government regulations that can accommodate the planned flood discharge for a ten-year return period of 180.59/second (Putra et al., 2014).

The essential point when implementing typologies 1–3 is that introducing vetiver grass to the community is challenging. For example, it can be seen from the previous study that the application of vetiver grass in Central Java Province was considered a failure due to: the lack of experts who understood the function of vetiver in preventing landslides, lack of community participation due to lack of understanding, and no economic benefits obtained by the community, making them less interested on vetiver development (Budi & Wariyanto, 2003). In contrast, the application of vetiver grass in Thailand, where there is good leadership in the community and a strong commitment from the government, the vetiver program has succeeded in improving the quality of the environment as well as the socio-economic status of the community (Leknoi & Likitlersuang, 2020).

CONCLUSION

The design guidelines for the application of the vegetative approach, both typologies 1, 2, and 3, can be applied in other locations, primarily on riverbanks that are prone to landslides, if they have the same characteristics, which are included in the category of dammed rivers in urban areas. Likewise, the selected vegetation types can adopt in other locations because the vegetation can be found and does not require specific maintenance. Hence, the community can participate in implementing this vegetative approach easily. In addition, if a do-something strategy has been implemented, the government not only focuses on this technical solution but also combines it with vegetative approaches such as tree planting so that the results are more optimal and benefit the community.

ACKNOWLEDGMENT

This study is part of the Research of Application and Development (RPP) funded from non-State Budget funds (non-APBN) Universitas Diponegoro 2021.

REFERENCES

- Adhitya, F., Rusdiana, O., & Saleh, M. B. (2016). Penentuan jenis tumbuhan lokasi dalam upaya mitigasi longsor dan teknik budidayanya pada areal rawan longsor di KPH Lawu DS: Studi kasus di RPH Cepoko [Determination of plant types in landslide mitigation efforts and their cultivation techniques in landslide prone areas in the Lawu Forest Management Unit (KPH) DS: A case study at the Cepoko Forest Management Resort (RPH)]. *Jurnal Silvikultur Tropika*, 8(1), 9-19.

- Bariteau, L., Bouchard, D., Gagnon, G., Levasseur, M., Lapointe, S., & Bérubé, M. (2013). A riverbank erosion control method with environmental value. *Ecological Engineering*, 58, 384-392. <https://doi.org/10.1016/j.ecoleng.2013.06.004>
- Barros, J. L., Tavares, A. O., & Santos, P. P. (2021). Land use and land cover dynamics in Leiria City: Relation between peri-urbanization processes and hydro-geomorphologic disasters. *Natural Hazards*, 106(1), 757-784. <https://doi.org/10.1007/s11069-020-04490-y>
- Budi, P. K., & Wariyanto, A. (2003, October 6-9). Why vetiver grass fails to show its miracles in Central Java, Indonesia? In *International Conference on Vetiver and Exhibition* (pp. 1-6). Guangzhou, China.
- Dariah, A., Sutono, S., Nurida, N. L., Hartatik, W., & Pratiwi, E. (2015). Pembenh tanah untuk meningkatkan produktivitas lahan pertanian [Soil Improvement to increase agricultural land productivity]. *Jurnal Sumberdaya Lahan*, 9(2), 67-84.
- de Jesús Arce-Mojica, T., Nehren, U., Sudmeier-Rieux, K., Miranda, P. J., & Anhuf, D. (2019). Nature-based solutions (NbS) for reducing the risk of shallow landslides: Where do we stand? *International Journal of Disaster Risk Reduction*, 41(April), Article 101293. <https://doi.org/10.1016/j.ijdr.2019.101293>
- Ducey, M. J., Johnson, K. M., Belair, E. P., & Cook, B. D. (2018). The influence of human demography on land cover change in the Great Lakes States, USA. *Environmental Management*, 62(6), 1089-1107. <https://doi.org/10.1007/s00267-018-1102-x>
- Ghosh, C., & Bhattacharya, S. (2018). Landslides and erosion control measures by vetiver system. In R. S. I. Pal (Ed.), *Disaster Risk Governance in India and Cross Cutting Issues* (pp. 387-413). Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-10-3310-0_19
- Herman, M., Syakir, M., Pranowo, D., & Saefudin, S. (2013). *Kemiri sunan (reutealis trisperma (blanco) airy shaw) tanaman penghasil minyak nabati dan konservasi lahan* [Kemiri Sunan (Reutealis trisperma (Blanco) Airy Shaw) vegetable oil-producing crops and land conversion]. Indonesian Agency for Agricultural Research and Development.
- Hostettler, S., Jöhr, A., Montes, C., & D'Acunzi, A. (2019). Community-based landslide risk reduction: A review of a red cross soil bioengineering for resilience program in Honduras. *Landslides*, 16, 1779-1791. <https://doi.org/https://doi.org/10.1007/s10346-019-01161-3>
- Indrayati, A., & Hikmah, N. I. (2018). Pembelajaran bencana banjir Bandang dengan pendekatan geospasial di DAS Beringin Kota Semarang [Flash flood disaster lessons learned with geospatial approaches in the Beringin Watershed, Semarang City]. *Prosiding Seminar Nasional Geotik 2018*, 70-81. Universitas Muhammadiyah Surakarta.
- Izzul, M., Noor, F., Bakhtiar, Y., & Saleh, A. (2020). Pemanfaatan tanaman sela pada lahan budidaya jambu kristal (*Psidium guajava* L.) di Desa Neglasari [Utilization of alley cropping on crystal guava (*Psidium guajava* L.) Cultivation in Neglasari Village]. *Jurnal Pusat Inovasi Masyarakat*, 2(5), 763-770.
- Kalsnes, B., & Capobianco, V. (2022). Use of vegetation for landslide risk mitigation. In C. Kondrup, P. Mercogliano, F. Bosello, J. Mysiak, E. Scoccimarro, A. Rizzo, R. Ebrey, M. de Ruiter, A. Jeuken & P. Watkiss (Eds.), *Climate Adaptation Modelling* (pp. 77-86). Springer Nature. <https://doi.org/https://doi.org/10.1007/978-3-030-86211-4>

- Komarawidjaja, W., & Garno, Y. S. (2016). Peran Rumput vetiver (*Chrysopogon zizanioides*) dalam fitoremediasi pencemaran perairan sungai [The role of vetiver grass (*Chrysopogon zizanioides*) in phytoremediation of river water pollution]. *Jurnal Teknologi Lingkungan*, 17(1), 7-14. <https://doi.org/10.29122/jtl.v17i1.1459>
- Kusnandar, A., & Kusminingrum, N. (2011). *Reviewing Development of Vetiver Grass in Preventing Erosion and Shallow Landslides on Road Slopes* (1st Ed.). Ministry of Public Works Republic of Indonesia.
- Leknoi, U., & Likitlersuang, S. (2020). Good practice and lesson learned in promoting vetiver as solution for slope stabilisation and erosion control in Thailand. *Land Use Policy*, 99(July), Article 105008. <https://doi.org/10.1016/j.landusepol.2020.105008>
- Li, Y., Zhang, X., Zhao, X., Ma, S., Cao, H., & Cao, J. (2016). Assessing spatial vulnerability from rapid urbanization to inform coastal urban regional planning. *Ocean and Coastal Management*, 123, 53-65. <https://doi.org/10.1016/j.ocecoaman.2016.01.010>
- Lobo, G. P., & Bonilla, C. A. (2017). A modeling approach to determining the relationship between vegetative filter strip design and sediment composition. *Agriculture, Ecosystems and Environment*, 237, 45-54. <https://doi.org/10.1016/j.agee.2016.12.027>
- Maryono, A. (2009). Kajian lebar sempadan sungai (studi kasus sungai-sungai di provinsi daerah istimewa Yogyakarta) [A study of stream buffer width (Case study of rivers in Daerah Istimewa Yogyakarta Province)]. *Dinamika Teknik Sipil*, 9(Januari), 55-66.
- Mentari, M., Mulyaningsih, T., & Aryanti, E. (2018). Identifikasi bambu di Sub Daerah Aliran Sungai Kedome Lombok Timur dan alternatif manfaat untuk konservasi sempadan sungai [Identification of bamboo in the Kedome River Basin, East Lombok and alternative benefits for river border conservation]. *Jurnal Penelitian Pengelolaan Daerah Aliran Sungai*, 2(2), 111-122.
- Ministry of Public Works. (2007). *Guidelines for spatial planning, landslide disaster area*. Indonesia.
- Mondal, S., & Patel, P. P. (2020). Implementing vetiver grass-based riverbank protection programmes in rural West Bengal, India. *Natural Hazards*, 103(1), 1051-1076. <https://doi.org/10.1007/s11069-020-04025-5>
- National Disaster Management Agency of Republic of Indonesia. (2015). *Disaster Risk Study Document, Central Java Province, 2016 - 2020*. National Disaster Management Agency of Republic of Indonesia.
- Oliver, R. D., & Thomas, V. A. (2014). Micropolitan areas: Exploring the linkages between demography and land-cover change in the United States cities. *Cities*, 38, 84-94. <https://doi.org/10.1016/j.cities.2014.01.002>
- Patton, C. V., Sawicki, D. S., & Clark, J. J. (2016). *Basic Methods of Policy Analysis and Planning* (3rd Ed.). Routledge.
- Persichillo, M. G., Bordoni, M., Meisina, C., Bartelletti, C., Barsanti, M., Giannecchini, R., D'Amato Avanzi, G., Galanti, Y., Cevasco, A., Brandolini, P., & Galve, J. P. (2017). Shallow landslides susceptibility assessment in different environments. *Geomatics, Natural Hazards and Risk*, 8(2), 748-771. <https://doi.org/10.1080/19475705.2016.1265011>
- Putra, D. P., Suharyanto, & Nugroho, H. (2014). Perencanaan normalisasi sungai beringin di kota semarang [Planning for normalization of the Beringin River in the city of Semarang]. *Karya Teknik Sipil*, 3, 1083-1097.

- Risdiyanto, I. (2016). *Mapping high conservation value area for ecosystem services at streams and riparian buffer*. Bogor Agricultural University, Indonesia.
- Semarang City Government. (2014). *Semarang City Drainage System Master Plan 2011-2031*. Indonesia.
- Semarang City Government. (2016). *Mid-term Development Plan Semarang 2016-2021*. Indonesia.
- Semarang City Government. (2017). *Spatial Plan of Semarang City*. Indonesia.
- Semarang City Government. (2021). *The Semarang City Government Socializes the Land Acquisition of Beringin River*. https://semarangkota.go.id/p/2836/pemkot_sosialisasi_pembebasan_lahan_kali_beringin
- Sherly, M. A., Karmakar, S., Parthasarathy, D., Chan, T., & Rau, C. (2015). Disaster vulnerability mapping for a densely populated coastal urban area: An application to Mumbai, India. *Annals of the Association of American Geographers*, 105(6), 1198-1220. <https://doi.org/10.1080/00045608.2015.1072792>
- Sidle, R. C., & Onda, Y. (2004). Hydrogeomorphology: Overview of an emerging science. *Hydrological Processes*, 18(4), 597-602. <https://doi.org/10.1002/hyp.1360>
- Sittadewi, E. H. (2016). Penentuan jenis vegetasi lokal untuk perlindungan tebing Sungai Siak dengan desain eko - Engineering tanpa turap [Determination of local vegetation types for protection of Siak River Cliffs with eco design - Engineering without sheeting]. *Jurnal Teknologi Lingkungan*, 11(2), 139-322. <https://doi.org/10.29122/jtl.v11i2.1202>
- Su, Q., Chen, K., & Liao, L. (2021). The impact of land use change on disaster risk from the perspective of efficiency. *Sustainability*, 13(6), 1-14. <https://doi.org/10.3390/su13063151>
- Subagyono, K., Marwanto, S., & Kurnia, U. (2003). Teknik konservasi tanah secara vegetatif. In *Sesi Monograf No.1 Sumber Daya Tanah Indonesia* (1st Ed.) [Vegetative soil conservation techniques. In Monograph Session No.1 Land Resources Indonesia (1st Ed.)]. Balai Penelitian Tanah.
- Sutopo, I. G., Teguh, M., & Pradipta, N. W. (2019). *Mitigasi Risiko Banjir Sungai Bringin Wilayah Hilir, Kabupaten Semarang Barat* [Mitigation of the Beringin River Flood Risk in the downstream area, West Semarang Regency]. Universitas Islam Indonesia.
- Zêzere, J. L., Trigo, R. M., & Trigo, I. F. (2005). Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): Assessment of relationships with the North Atlantic Oscillation. *Natural Hazards and Earth System Science*, 5(3), 331-344. <https://doi.org/10.5194/nhess-5-331-2005>



Quality Evaluation of Color and Texture of the Dabai Fruit (*Canarium odontophyllum* Miq.) at Different Temperatures and Times of Blanching

Rosnah Shamsudin^{1,2*}, Siti Hajar Ariffin¹, Wan Nor Zanariah Zainol @ Abdullah³, Nazatul Shima Azmi¹ and Arinah Adila Abdul Halim¹

¹Department of Process and Food Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Laboratory of Halal Services, Halal Products Research Institute, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

³Department of Basic Science and Engineering, Faculty of Agriculture and Food Sciences, Universiti Putra Malaysia Bintulu Sarawak Campus, Nyabau Road, 97008, Bintulu Sarawak, Malaysia

ABSTRACT

The Dabai fruit (*Canarium odontophyllum* Miq.) is seasonal in Sarawak. The flesh is often blanched in hot or lukewarm water to make it creamier and softer before being served as a snack or side dish. The fruit was blanched at temperatures ranging from 60°C to 100°C, with 10°C increments, for up to 10 minutes, whereby the quality changes in color and texture were investigated. Notably, the L parameter has remained unchanged during the blanching process, indicating that the dabai fruit's dark color is retained. The parameters a*, b*, chroma (C), and total color difference (TCD), on the other hand, have increased significantly as the temperature and time of blanching are increased. However, the change in

firmness was not substantial due to the slight variations in firmness as the temperature and time increased. From these findings, it can be inferred that the temperature and duration of blanching have significantly changed the dabai fruit's color except for the firmness of the fruit. These findings would be useful in designing thermal processes and related calculations for the dabai fruit.

ARTICLE INFO

Article history:

Received: 6 February 2022

Accepted: 7 April 2022

Published: 21 July 2022

DOI: <https://doi.org/10.47836/pjst.30.4.07>

E-mail addresses:

rosnabs@upm.edu.my (Rosnah Shamsudin)

hajarariffin@upm.edu.my (Siti Hajar Ariffin)

wnzz@upm.edu.my (Wan Nor Zanariah Zainol @ Abdullah)

nazatul_shima89@yahoo.com (Nazatul Shima Azmi)

arina.adila@gmail.com (Arinah Adila Abdul Halim)

*Corresponding author

Keywords: Blanching, color, dabai, firmness, quality evaluation

INTRODUCTION

The dabai fruit (*Canarium odontophyllum* Miq.) is in the family of Burseraceae of Sapindales order, where it belongs to the Eudicotyledonae class. It is categorized in the drupe fruit family, where in botany, it is often known as 'stone fruit,' which includes olive oil and palm oil. It is an indigenous seasonal fruit that can only be found in Borneo Island, mainly in the Sibuhay and Kapit regions of Sarawak (Rukayah, 2009; Chua & Nicholas, 2009; Chua et al., 2015). The most fruitful annual seasons for dabai are from May to June and December to January. The dabai fruits were collected using a sickle and a net to gather the falling fruits and branches owing to the cutting of terminal branches of the fruit panicles when the immature white fruit became purplish-black (Ding, 2011). Significantly, dabai fruits are known for their high nutritional value, antioxidant, and mineral content, including potassium, phosphorus, calcium, and magnesium, all of which are present in the skin (Shakirin et al., 2010; Ariffin et al., 2020). The dabai fruit is always eaten by dipping it in soy sauce or is used as an ingredient in a dish such as 'nasi goreng dabai.' Furthermore, a few studies have reported that the skin of dabai fruits has high phytochemical contents such as anthocyanin and flavanoids (Pin et al., 2014; Khoo et al., 2016; Ariffin et al., 2020). According to Ding (2011), one of the most common methods to prepare the dabai fruit is by blanching it for 15 minutes using lukewarm water.

Blanching is rapidly heating vegetables and fruits to a certain temperature and holding them for 1 to 10 minutes. The process is a thermal treatment used to enhance the taste or texture of fruits and vegetables. The convectional water method is the common technique for blanching dabai fruit. It is easier to be operated as the dabai fruit is immersed in hot or lukewarm water for several minutes. Then, the water will be drained, and the blanched products will be cooled before the next process is continued. As aforementioned in the previous study, the dabai fruits are served after soaking in lukewarm water at 50°C for about 15 to 20 minutes to soften the fruit's flesh (Ding, 2011). As a result, the blanched dabai will have a smooth, creamy texture and enhance the flavor. Usually, the locals will eat the blanched dabai directly after the treatment. Thus, there is no processing operation after the blanching process. However, the blanching process by the locals varies as they do not have a predetermined temperature and duration. In the traditional method, most of them blanched the dabai fruits judging merely by their appearances.

Quality products were defined as the products that meet the consumer's perception, whereby the quality of the products was judged based on the product attributions that the consumer needed. The attributes or characteristics of the products can be measurable or based on the consumer's liking, the characteristics of the environment of the products, the reliability, and more. Generally, quality components consist of appearance, such as freshness, ripeness, color, flavor, and firmness of the measured fruits. For example, Abu-Ghannam and Jaiswal (2015) observed the effects of blanching on the Irish York cabbage

and found that the texture and color of the cabbage were reduced. Color is one of the important parameters for the consumers' judgment as it contributes to the appearance attributes; thereby, if the product's color is undesirable, it will influence the consumers' acceptance, thus, affecting the market value of the product.

The color of the products can be related to other quality properties (Xiao et al., 2017). Color can be used as the indicator of quality change of the products when they undergo the blanching process and to determine the quality degradation caused by the thermal blanching process. Moreover, color is also important in determining the shelf life of the dabai fruits and indicates the fruits' maturity in the duration of storage. When fully matured, the skin of the dabai fruit is blue-black. The pigment color of the skin is mainly attributed to anthocyanin (cyanine-3-glucoside) (Deylami et al., 2016; Shakirin et al., 2010). During the blanching process, a Maillard reaction might affect the dabai fruit's color changes (Razak et al., 2021). The texture is also one of the important attributes to ensure that the products are of good quality and value for the consumers. Therefore, the texture is often used as an indicator in the blanching process. It will determine the physical-chemical properties of the cell wall and indicate what has happened to the texture of the products during processing (Xiao et al., 2017). According to Jaiswal et al. (2012), during the blanching process, a loss of texture occurs that continues up to 14 minutes.

The process of chemicals changes the texture of processed fruit and vegetables. Furthermore, during the blanching or soaking process, it can escalate moisture diffusivity, which leads to the significance in hydration frequency, thus, catalyzing the water absorption and changing the textural (Shafaei et al., 2016; Ehiem et al., 2019; Razak et al., 2021). Subsequently, the process will affect the dabai fruit's nutritional content, such as its crude fat that provides the texture and flavor (Razak et al., 2021). The color and texture were regarded as the qualitative features of food items to guarantee product acceptance. After the blanching process, it is important to observe the quality of the dabai fruit in terms of color and firmness to avoid under-blanching and over-blanching. It is because under-blanching can speed-up enzyme activity, while over-blanching can cause degradation of texture and color (Ruiz-Ojeda & Peñas, 2013; Shamsudin et al., 2021). There has been little study on the effect of blanching conditions, including temperature and the duration of blanching that can retain the quality of the dabai fruits, which have yet to be determined.

In this paper, the researchers attempt to understand and determine the effects of blanching conditions on the quality of the dabai fruit that is blanched under a range of time-temperature conditions. This data will benefit in designing the future thermal processes and related calculations for dabai fruit.

METHODOLOGY

Plant Materials

The dabai fruits were supplied by a supplier from Sibul, Sarawak, in Malaysia. The fruit was packed in an icebox and delivered within 24 hours to Universiti Putra Malaysia, Serdang, Selangor, in Malaysia. Upon arrival, they were stored in the freezer (SJC218, Sharp, Selangor, Malaysia) at a temperature of -4°C . Dabai fruits that were free of damage and pests were used as the raw material for the blanching assessment (in triplicate). The Dabai fruit was defrosted at room temperature for a few minutes before being treated, prior to blanching.

Water Blanching

The dabai fruit was immersed in a beaker filled with distilled water for each trial. It was then heated using a water bath (JSSB, Laft Technologies, Melbourne, Australia) at a temperature between 60°C and 100°C , with an increment of 10°C . Samples were blanched for up to 10 minutes at each temperature, at 2 minutes intervals. After that, the samples were analyzed for color and texture properties.

Color Analysis

A colorimeter (Miniscan EZ Spectrophotometer, HunterLab, Virginia, USA) was used to examine the visual color of the blanched dabai samples. The color parameters measured were L^* for whiteness or brightness, a^* for redness or greenness, and b^* for yellowness or blueness. The quantitative attribute of colorfulness, also known as chroma, C^* , and total color difference, TCD, are calculated using Equations 1 and 2, respectively (Cruz et al., 2007):

$$C = \sqrt{a^{*2} + b^{*2}} \quad (1)$$

$$\text{TCD} = \sqrt{(L^* - L_o^*)^2 + (a^* - a_o^*)^2 + (b^* - b_o^*)^2} \quad (2)$$

Texture Analysis

The texture of the blanched dabai fruit was determined using a Texture Analyzer (TA-XT plus, Stable Micro Systems, Surrey, UK), where the parameter measured was firmness (N). For energy and firmness measurements, a 500 g load cell with a 5 mm diameter probe was employed (Gonçalves et al., 2007). A single puncture measurement of 10 mm penetration depth on each sample was performed at a velocity of 1.0 mm s^{-1} . In addition, the area under the curves and the slope newton (N/mm) were measured to determine the hardness of the force formation curves.

RESULTS AND DISCUSSION

Effect of Blanching on the Color of Dabai Fruit

Color is one of the most important quality parameters attributed to food materials. Table 1 shows the color characteristics of the dabai fruit before blanching. The dabai fruit has dark purplish color, as indicated in Table 1. Before blanching, the dabai fruit has a value of 0.000 for the L parameter, which is translated to the blackish region. Furthermore, it has a negative value of a* parameter, which is -0.275, which indicates a shift towards greenness. On the other hand, the positive value of the b* parameter, which is 5.923, suggests a shift towards yellow. Therefore, the combination of a* parameter (-0.275) and b* parameter (5.923) is identified as neutral color or gray.

Table 1

Color characteristics of dabai fruit 'Ngemah' variety before blanching

Parameter	Initial value
L	0.000
a*	-0.275
b*	5.923
C	5.213
TCD	0.000

The changes of color in the dabai fruit after blanching were studied at 60°C, 70°C, 80°C, 90°C, and 100°C at different durations which were 2, 4, 6, 8, and 10 minutes as shown in Figure 1(a-d). From the observations, the L parameter remained unchanged even after blanching at a high temperature and for a prolonged blanching time. The robust stability of the anthocyanins pigment responsible for the dark purple color of dabai may have been why the L

parameter does not change (Khoo et al., 2017). Since the L value indicates the sample's darkness, the sample will not turn darker due to its initial color, which is already a dark purplish color (Ganjloo et al., 2009).

In Figure 1(a), the a* value increases as the duration time increases from 2 to 10 minutes for all the temperatures except the blanching temperatures, which is at 60°C. The a* value at 60°C does not show any significant changes ($p > 0.05$) at any blanching time, suggesting that the heat supplied has not been sufficient to change the a* value. On the other hand, for blanching at a higher temperature, 100°C, the value of a* at 2 min was 5.876, which showed a high shift towards the redness color. A similar trend can be seen at temperatures of 70°C, 80°C, and 90°C. The trend for the temperature at 70°C shows a value that has increased gradually from 2 minutes until 10 minutes as time increased. However, the trend for the temperatures 80°C, 90°C, and 100°C shows that the value of a* has increased significantly from 4 to 10 minutes. The significant increase ($p < 0.05$) value of a* as the processing time and the temperature increased indicated that the dabai fruit gained a red color. This finding agrees with the study by Krokida et al. (2001), where the a* parameter has increased significantly due to the drying process on the potato strips indicated in the product's browning.

The b^* parameter for the blanched dabai fruit also showed some changes in values, as shown in Figure 1(b). At 60°C at 2 minutes b^* value was 6.063 and increased to 7.480 for 10 minutes of blanching. For 100°C, the value of b^* after blanching at 2 minutes was 8.574 and rose to 11.563 at 10 minutes. A similar trend can also be seen in 70°C, 80°C, and 90°C. The result showed that the b^* value increased significantly ($p < 0.05$) as the temperature and time duration increased. At all temperatures, the b^* values gradually increased except for the temperature of 100°C, where the b^* value had increased significantly from 9.431 (6 min) to 11.447 (8 min). However, the temperature of 60°C and 70°C increased moderately throughout the 10 minutes. Overall, because of blanching on the b^* parameter, it can be concluded that as the temperature and time of blanching increase, the b^* parameter will increase too, resulting in the yellowness of the dabai fruit after blanching. This result is also in agreement with the drying and dehydration effect on the b^* parameter color, where a significant increase of yellowness can be seen on the carrots after both of the heat treatments (Zielinska & Markowski, 2012).

Based on Figure 1(c), the value of chroma after blanching at 60°C for 2 minutes is 6.066 and has increased to 7.577 at 10 minutes. For 100°C, the value of chroma at 2 minutes is 10.394 and has increased to 22.662 at 10 minutes. The temperature at 70°C, 80°C, and 90°C has also shown the same trend. The result showed that the degree of chroma during blanching increased significantly ($p < 0.05$) as the time increased for all temperatures. Figure 1(c) describes that the temperatures 80°C, 90°C, and 100°C for 60°C and 70°C have gradually increased, where the chroma value rises at 6 minutes of blanching. However, the temperature of 60°C and 70°C increased moderately throughout the 10 minutes. Thus, the temperature and time of blanching affect the chroma value of the dabai fruit. This finding was in agreement with Brewer et al. (1995), where blanching treatments such as conventional boiling water, steam blanching, and microwave blanching of the frozen broccoli increased the broccoli's chroma color florets and stems. The increment value of chroma (C) suggested that the blanched dabai fruit gained vividness or saturation of color. The chroma parameter is related to the proportion of the grey component responsible for characterizing the color (Olivera et al., 2008). As C increases, the intensity of the color has also increased.

The TCD for all temperatures in Figure 1(d) displayed increasing values of TCD as the blanching time increased. The temperature of 80°C, 90°C, and 100°C showed that the value of TCD soared after 4 minutes. However, the temperature of 60°C and 70°C increased moderately throughout the 10 minutes. For example, at 60°C at 4 minutes, the value of TCD is 0.653 and has increased to 1.045 at 6 minutes which is a 37.47% difference. As for 100°C, the TCD value at 4 minutes is 7.795, which soars to 16.671 at 6 minutes which is 46.76%. The result shows that the TCD has increased significantly ($p < 0.05$) as the blanching duration increases. The increasing value of the total color difference of the

dabai fruit caused by the blanching suggested that the color had changed due to the thermal blanching process. According to Jaiswal et al. (2012), the TCD increases during blanching due to the different reactions that may have resulted in the color changes, such as thermal degradation of the pigments and the oxidation of ascorbic acid enzymatic browning and non-enzymatic browning.

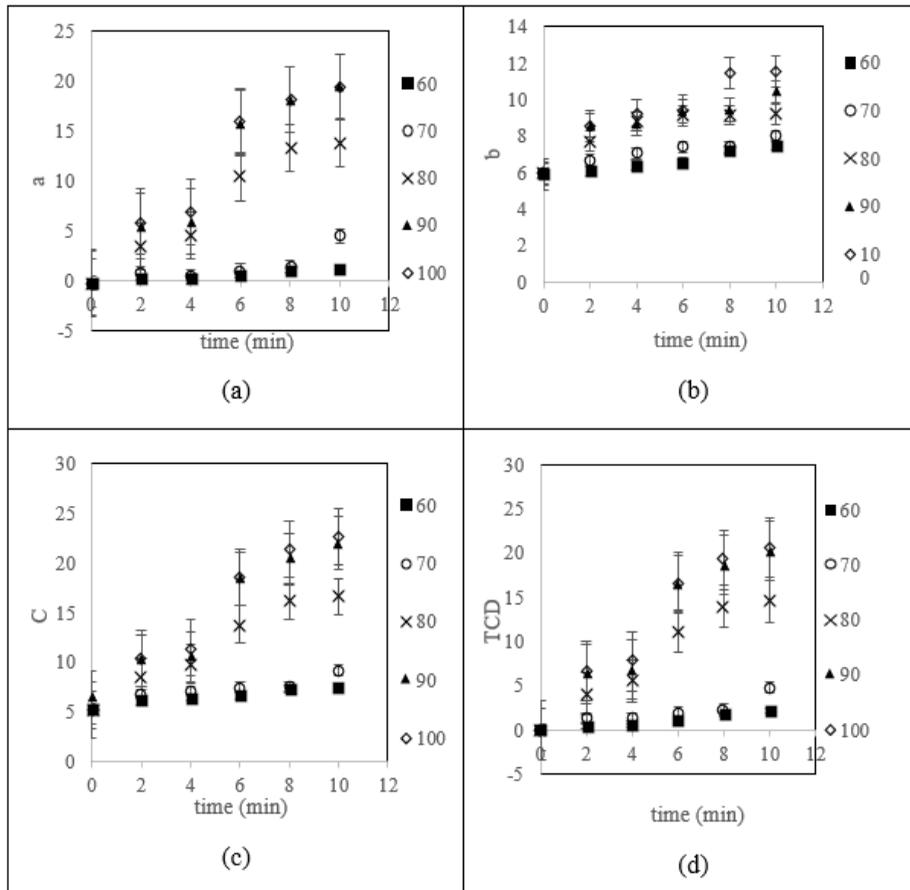


Figure 1. The effect of blanching (60°C, 70°C, 80°C, 90°C, and 100°C) on the color parameters of dabai fruit: (a) a*; (b) b*; (c) C; and (d) TCD

Effect of Blanching on the Texture of Dabai Fruit

The effect of blanching on the dabai fruit in terms of texture was measured using the firmness parameter. For 60°C, at 2 minutes of blanching, the firmness value was 3.616 N and increased to 4.065 at 10 minutes, as shown in Figure 2. For 100°C, the firmness of the dabai fruit after blanching at 2 minutes decreased to 2.034N from 5.304N (0 min) and increased again at 10 minutes to 2.075N. A similar trend can be seen at temperatures 70°C,

80°C, and 90°C. According to the Figure 2, the firmness of the dabai fruit decreased as the blanching time increased. However, at 8 minutes, the degree of firmness started to grow for all the temperatures. It may be due to the pectin methylesterase enzyme in some vegetables and fruits. The enzyme pectin methylesterase (PME) was found to have a firming effect on fruits or vegetables if activated by blanching (Anthon & Barrett, 2006). The decrease in the firmness of the dabai fruit may be caused by the changes in the pectin strengths caused by the thermal softening and enzyme degradation (Kim, 2006). The phenomena that started at 8 minutes can be explained by the PME enzyme activity that increases when the pectic substances inside the dabai fruit are demethoxylated, which causes the formation of calcium and magnesium pectate (Inheiro et al., 2009). However, most researchers have stated that the firming effect of the pectin methylesterase enzyme is often around the temperature range of 50°C to 70°C (Kim, 2006; Ni et al., 2005). In addition, according to the research by Abu-Ghannam and Crowley (2006), pre-treatment using 65°C before the blanching process, which ranges from 95°C to 100°C, causes the firmness of the potato to increase compared with the blanching process without any pre-treatment. However, Table 2 shows that the firmness value has no significant ($p>0.05$) difference after blanching at 60°C to 100°C between 2 to 10 minutes. It is due to the small value increment and decrement of firmness in all temperatures and durations. It shows that the blanching time does not affect the firmness of the dabai fruit.

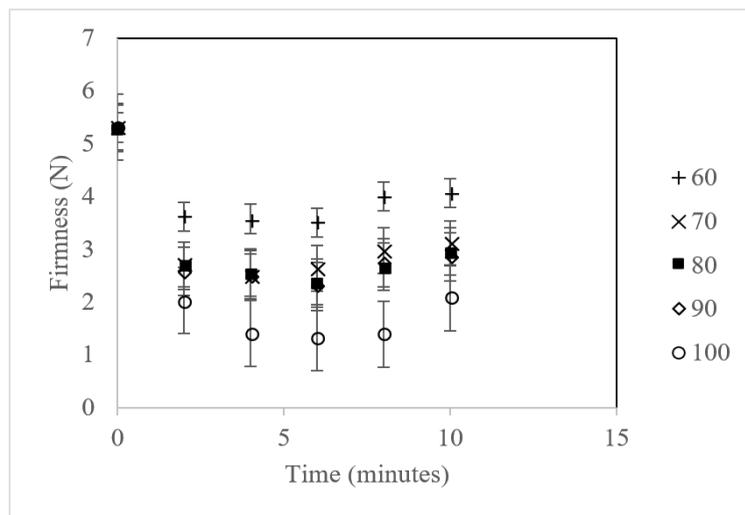


Figure 2. The effect of blanching (60°C, 70°C, 80°C, 90°C, and 100°C) on the firmness of dabai fruit

Table 2
Effect of different temperature and duration of blanching on the firmness dabai fruit

Temperature (°C)	Duration of blanching (min)					
	0	2	4	6	8	10
60	5.300±3.250 ^{Aa}	3.616±1.660 ^{Aa}	3.570±1.930 ^{Aa}	3.503±1.5018 ^{Aa}	3.994±0.533 ^{Aa}	4.070±2.960 ^{Aa}
70	5.300±3.250 ^{Aa}	2.709±0.880 ^{Aa}	2.485±0.648 ^{Aa}	2.631±1.033 ^{Aa}	2.974±0.574 ^{Aa}	3.110±2.860 ^{Aa}
80	5.300±3.250 ^{Aa}	2.684±0.767 ^{Aa}	2.553±0.696 ^{Aa}	2.357±0.473 ^{Aa}	2.664±0.244 ^{Aa}	2.952±0.107 ^{Aa}
90	5.300±3.250 ^{Aa}	2.580±2.480 ^{Aa}	2.510±2.070 ^{Aa}	2.300±1.870 ^{Aa}	2.738±1.179 ^{Aa}	2.848±1.303 ^{Aa}
100	5.300±3.250 ^{Aa}	2.034±1.227 ^{Aa}	1.4074±0.1143 ^{Aa}	1.325±0.547 ^{Aa}	1.382±0.328 ^{Aa}	2.075±0.604 ^{Aa}

*Mean with capital letter A represents same temperature of blanching with different minutes of blanching are not significantly differ ($p>0.05$)

**Mean with different small letter a represents same duration of blanching with different temperature of blanching are not significantly differ ($p>0.05$)

CONCLUSION

The findings in the present study confirmed that blanching has a deleterious effect on the color and texture of the dabai fruit. For the color analysis, parameters a^* , b^* , C , and TCD for the blanched dabai fruit showed significant changes for each temperature and time of blanching, except for the L parameter. However, the firmness of the blanched dabai fruit at different temperatures and blanching times was insignificant as there were only minor changes. This information will add to the knowledge that contributes as an advantage and an impact on consumer food selection and enhances the ability to find the right conditions for blanching the dabai fruit. Furthermore, the information would also be useful in designing thermal processes and related calculations for the dabai fruit.

ACKNOWLEDGMENT

The authors thank the Ministry of Higher Education of Malaysia for providing financial support under the Fundamental Research Grants Scheme (FRGS) (project number: FRGS/1/2019/WAB01/UPM/02/30) and the Sarawak Biodiversity Centre, who approved and issued the R&D permit.

REFERENCES

- Abu-Ghannam, N., & Jaiswal, A. K. (2015). Blanching as a treatment process: effect on polyphenol and antioxidant capacity of cabbage. In *Processing and impact on active components in food* (pp. 35-43). Academic Press. <https://doi.org/10.1016/B978-0-12-404699-3.00005-6>
- Anthon, G. E., & Barrett, D. M. (2006). Characterization of the temperature activation of pectin methylesterase in green beans and tomatoes. *Journal of Agricultural and Food Chemistry*, 54(1), 204-211. <https://doi.org/10.1021/jf051877q>
- Ariffin, S. H., Shamsudin, R., & Tawakkal, I. S. M. A. (2020). Dabai fruit: Postharvest handling and storage. *Advances in Agricultural and Food Research Journal*, 1(2), 1-12. <https://doi.org/10.36877/aafmj.a0000126>
- Brewer, M. S., Begum, S., & Bozeman, A. V. A. (1995). Microwave and conventional blanching effects on chemical, sensory, and color characteristics of frozen broccoli. *Journal of Food Quality*, 18(6), 479-493. <https://doi.org/10.1111/j.1745-4557.1995.tb00398.x>
- Chua, H. P., & Nicholas, D. (2009). Dabai - Speciality fruit of Sarawak. *Agromedia*, 30, 28-30.
- Chua, H. P., Nicholas, D., & Yahya, M. N. A. (2015). Physical properties and nutritional values of dabai fruit (*Canarium odontophyllum*) of different genotypes. *Journal Tropical Agriculture and Food Science*, 43(1), 1-10.
- Cruz, R. M. S., Vieira, M. C., & Silva, C. L. M. (2007). Modelling kinetics of watercress (*Nasturtium officinale*) colour changes due to heat and thermosonication treatments. *Innovative Food Science & Emerging Technologies*, 8(2), 244-252, <https://doi.org/10.1016/j.ifset.2007.01.003>

- Deylami, M. Z., Rahman, R. A., Tan, C. P., Bakar, J., & Olusegun, L. (2016). Effect of blanching on enzyme activity, color changes, anthocyanin stability and extractability of mangosteen pericarp: A kinetic study. *Journal of Food Engineering*, 178, 12-19. <https://doi.org/10.1016/j.jfoodeng.2016.01.001>
- Ding, P. (2011). *Dabai (Canarium odontophyllum Miq.), Postharvest biology and technology of tropical and subtropical fruits: Volume 3: Cocona to mango*. Woodhead Publishing Limited. <https://doi.org/10.1016/B978-1-84569-735-8.50003-6>
- Ehiem, J. C., Ndirika, V. I. O., Onwuka, U. N., Garipey, Y., & Raghavan, V. (2019). Water absorption characteristics of *Canarium Schweinfurthii* fruits. *Information Processing in Agriculture*, 6(3), 386-395. <https://doi.org/10.1016/j.inpa.2018.12.002>
- Ganjloo, A., Rahman, R. A., Bakar, J., Osman, A., & Bimakr, M. (2009). Modelling the kinetics of peroxidase inactivation and colour changes of seedless guava (*Psidium guajava* L.) during thermal treatments. *Applied Sciences*, 7(1), 105-112.
- Gonçalves, E. M., Pinheiro, J., Abreu, M., Brandão, T. R. S., & Silva, C. L. M. (2007). Modelling the kinetics of peroxidase inactivation, colour and texture changes of pumpkin (*Cucurbita maxima* L.) during blanching. *Journal of Food Engineering*, 81, 693-701. <https://doi.org/10.1016/j.jfoodeng.2007.01.011>
- Inheiro, J. O. P., Legria, C. A. A., & Breu, M. A. A. (2009). Degradation kinetics of peroxidase enzyme, phenolic content, and physical and sensorial characteristics in broccoli (*Brassica oleracea* L. ssp. *Italica*) during blanching. *Journal of Agricultural and Food Chemistry*, 57(12), 5370-5375. <https://doi.org/10.1021/jf900314x>
- Jaiswal, A. K., Gupta, S., & Abu-Ghannam, N. (2012). Kinetic evaluation of colour, texture, polyphenols and antioxidant capacity of Irish York cabbage after blanching treatment. *Food Chemistry*, 131(1), 63-72. <https://doi.org/10.1016/j.foodchem.2011.08.032>
- Khoo, H. E., Azlan, A., Kong, K. W., & Ismail, A. (2016). Phytochemicals and medicinal properties of indigenous tropical fruits with potential for commercial development. *Evidence-Based Complementary and Alternative Medicine*, 2016, Article 7591951. <https://doi.org/10.1155/2016/7591951>
- Khoo, H. E., Azlan, A., Tang, S. T., & Lim, S. M. (2017). Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food and Nutrition Research*, 61(1), 0-21. <https://doi.org/10.1080/16546628.2017.1361779>
- Kim, J. A. E. C. (2006). Firmness of thermal processes onion as affected by blanching. *Journal of Food Processing and Preservation*, 30(6), 659-669. <https://doi.org/10.1111/j.1745-4549.2006.00096.x>
- Krokida, M. K., Oreopoulou, V., Maroulis, Z. B., & Marinos-Kouris, D. (2001). Deep fat frying of potato strips-quality issues. *Drying Technology*, 19(5), 879-935. <https://doi.org/10.1081/DRT-100103773>
- Ni, L., Lin, D., & Barrett, D. M. (2005). Pectin methylesterase catalyzed firming effects on low temperature blanched vegetables. *Journal of Food Engineering*, 70(4), 546-556. <https://doi.org/10.1016/j.jfoodeng.2004.10.009>
- Olivera, D. F., Viña, S. Z., Marani, C. M., Ferreyra, R. M., Mugridge, A., Chaves, A. R., & Mascheroni, R. H. (2008). Effect of blanching on the quality of brussels sprouts (*Brassica oleracea* L. gemmifera DC) after frozen storage. *Journal of Food Engineering*, 84(1), 148-155. <https://doi.org/10.1016/j.jfoodeng.2007.05.005>

- Pin, C. H., Daniel, N., & Mos, S. (2014). Phenolic and flavonoid contents and antioxidant activities of selected dabai (*Canarium odontophyllum*) genotypes. *Journal of Tropical Agriculture and Food Science*, 42(2), 105-114.
- Razak, A. F. A., Abidin, M. Z., Hassan, N. A., Edwin, A. J., Abdullah, M. S., Razak, A., Salleh, M. H., & Hamim, N. A. (2021). The impact of (*Canarium Odontophyllum* Miq.) dabai optimum soaking condition towards the development of dabai peanut spread physicochemical properties and sensory evaluation. *Journal of Agrobiotechnology*, 12(2), 56-67.
- Ruiz-Ojeda, L. M., & Peñas, F. J. (2013). Comparison study of conventional hot-water and microwave blanching on quality of green beans. *Innovative Food Science and Emerging Technologies*, 20, 191-197. <https://doi.org/10.1016/j.ifset.2013.09.009>
- Rukayah, A. (2009). Buah-buahan eksotik dari Sabah dan Sarawak di Mardi, Serdang [Exotic fruits from Sabah and Sarawak in Mardi, Serdang]. *Agromedia*, 22, 28-34.
- Shafaei, S. M., Masoumi, A. A., & Roshan, H. (2016). Analysis of water absorption of bean and chickpea during soaking using Peleg model. *Journal of the Saudi Society of Agricultural Sciences*, 15(2), 135-144. <https://doi.org/10.1016/j.jssas.2014.08.003>
- Shakirin, F. H., Prasad, K. N., Ismail, A., Yuon, L. C., & Azlan, A. (2010). Antioxidant capacity of underutilized Malaysian *Canarium odontophyllum* (dabai) Miq. fruit. *Journal of Food Composition and Analysis*, 23(8), 777-781. <https://doi.org/10.1016/j.jfca.2010.04.008>
- Shamsudin, R., Ariffin, S. H., Zainol@Abdullah, W. N. Z., Azmi, N. S., & Halim, A. A. A. (2021). Modelling the kinetics of color and texture changes of dabai (*Canarium odontophyllum* Miq.) during blanching. *Agronomy*, 11(11), Article 2185. <https://doi.org/10.3390/agronomy11112185>
- Xiao, H. W., Pan, Z., Deng, L. Z., El-Mashad, H. M., Yang, X. H., Mujumdar, A. S., Gao, Z. J., & Zhang, Q. (2017). Recent developments and trends in thermal blanching - A comprehensive review. *Information Processing in Agriculture*, 4(2), 101-127. <https://doi.org/10.1016/j.inpa.2017.02.001>
- Zielinska, M., & Markowski, M. (2012). Color characteristics of carrots: Effect of drying and rehydration. *International Journal of Food Properties*, 15(2), 450-466. <https://doi.org/10.1080/10942912.2010.489209>

Review Article

Computed Tomography and Other Imaging Modalities in Pediatric Congenital Heart Disease

Hongying Chen¹, Norafida Bahari^{1,2*}, Suraini Mohamad Saini^{1,3} and Noramaliza Mohd Noor^{1,2}

¹Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Department of Radiology, Teaching Hospital Universiti Putra Malaysia, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

³Centre of Diagnostic Nuclear Imaging, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Congenital heart defects (CHD) are the most common congenital disabilities. Early and accurate diagnosis of coronary heart disease is very important for patients to get timely and effective treatment. In recent years, the accuracy of coronary heart disease diagnosis has been greatly improved with the development of medical imaging equipment and technology. At present, the clinical application of echocardiogram (echo), cardiovascular magnetic resonance (CMR) and computed tomography angiography (CTA) in coronary heart disease anatomy and function has increased significantly, which plays an important role in preoperative diagnosis, intraoperative monitoring, and postoperative recovery evaluation. However, each imaging technique has its indications. Providing the best examination plan for patients requires clinicians and radiologists' close cooperation. Therefore, this study reviewed the imaging techniques for diagnosing coronary heart disease.

Keywords: Cardiovascular magnetic resonance, computed tomography, congenital heart defects, echocardiography, imaging diagnosis, radiation dose

ARTICLE INFO

Article history:

Received: 04 February 2021

Accepted: 05 July 2021

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.08>

E-mail addresses:

hedychenhy@163.com (Hongying Chen)

afidabahari@upm.edu.my (Norafida Bahari)

surainims@upm.edu.my (Suraini Mohamad Saini)

noramaliza@upm.edu.my (Noramaliza Mohd Noor)

* Corresponding author

INTRODUCTION

Congenital heart defect (CHD) is a structural anomaly of the cardiovascular system which occurs during fetal development. It can be classified as cyanosis or acyanosis (left-to-right shunts or obstructive lesions), and the

presence or absence of cyanosis depends on the direction of blood flow to the heart and the great arteries. The most common is Atrial Septal Defect (ASD) (Figure 1), Tetralogy of Fallot (TOF) (Figure 2), Ventricular Septal Defect (VSD) (Figure 2), Patent Ductus Arteriosus (PDA), Pulmonary Stenosis, and Aortic Stenosis et al. For example, in TOF, since the blood from the right heart is shunted to the left through the defect, the arterial blood is mixed with venous blood, and cyanosis occurs when the reduced hemoglobin level in the bloodstream is elevated above 5%. On the other hand, in the case of simple pulmonary stenosis, ASD, or VSD, blood from the left heart flows back to the right heart, and there are no venous and arterial blood mixes, so cyanosis is not present. There are many types of CHD, and several defects may occur in combination, also known as compound malformations.

According to the global burden of disease study (GBD) 2017, congenital heart disease (CHD) causes 261247 deaths worldwide, of which 180624 are infants under one year old (Zimmerman et al., 2020). The mortality rate of CHD in different countries and regions is quite different. Among them, the incidence rate of coronary heart disease is highest in sub-Saharan Africa, central and eastern China, Central Asia and Southeast Asia, and some inland provinces in China and India (Zimmerman et al., 2020). Coronary heart disease accounts for about 1% of the total number of newborns or nearly 40000 cases in the United States annually (Hoffman & Kaplan, 2002). In 2016, circulation published a study that estimated that about 1 million children in the United States had congenital heart defects in 2010 (Gilboa et al., 2016). The incidence rate of China's coronary heart disease is about 2.4–10.4 per thousand, according to a 2019 report on cardiovascular health and disease in China. It is a major congenital malformation in China (Hu, 2020). According to the health indicators of statistics in Malaysia in 2007, nearly 1/100 infants (nearly 1 or 5000 infants) in Malaysia suffered from coronary heart disease every year (MOH, 2007; MOH, 2008a; MOH, 2008b).

CHDs are a high incidence of congenital disabilities, the main cause of neonatal and infant deaths. Therefore, it is very important to diagnose CHD early and accurately so that patients can get timely and effective treatment. With

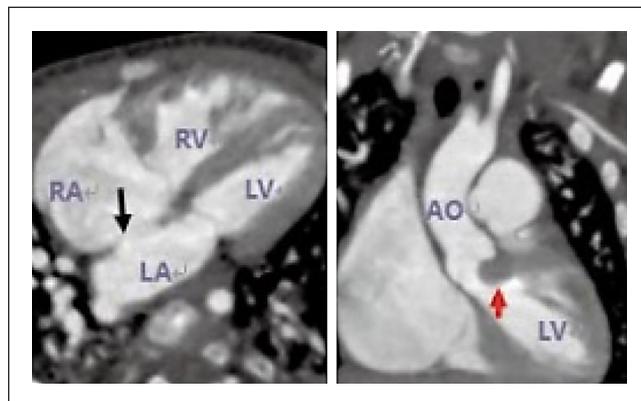


Figure 1. A 4-month-old girl with ASD. The diagram shows subaortic stenosis, ventricular septal projection on one side, left ventricular outflow tract (red arrow), and ASD (black arrow). RV = right ventricle, LV = left ventricle, RA = right atrium, LA = left atrium, AO = aorta. (Source: Miao et al., 2016)

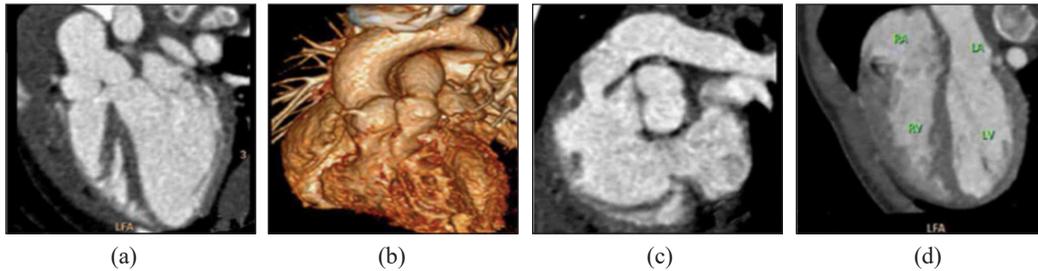


Figure 2. A 13-month-old girl with TOF: (a) The degree of aortic override was 50%; (b) Aortic root directly above the septal defect; (c) Narrowing of the right ventricle outflow tract, which was indicative of VSD.; and (d) Atrium and ventricle were a normal connection. (Source: D. Chen et al., 2017)

the development of medical imaging equipment and technology, there are more options for diagnosing CHD, and the accuracy has been greatly improved. At present, there are many diagnostic imaging methods, including Echocardiography (Echo), Cardiovascular Magnetic Resonance (CMR) and computed tomography angiography (CTA), which play an important role in the preoperative diagnosis, intraoperative monitoring, and postoperative recovery efficacy evaluation of CHD, and provide comprehensive and reliable information on cardiovascular anatomical structure and function of CHD.

CT scan can evaluate both anatomy and function of CHD morphologically. With the development of scanning and post-processing methods, the application of CTA in cardiovascular diseases is becoming more and more mature. However, due to the possibility of radiation damage in CT, different CT vendors have developed various techniques and systems to reduce the radiation dose to the patients.

The following problems exist in the imaging examination of CHD: (1) A single examination often cannot obtain all the lesion details and diagnostic information of CHD. (2) Traditional two-dimensional imaging is not ideal for treating CHD because the image display is incomplete. (3) Some examination methods are invasive and risky to patients. (4) Repeated imaging examination is a common phenomenon at present.

This review evaluates the role of imaging in treating patients with CHD; in particular, the recent advances in diagnosing CHD are described in detail. As much as possible, we will use published research evidence to support our discussion, while those lacking sufficient research evidence will adopt the expert consensus.

DIAGNOSTIC IMAGING OF CHD

Plain Chest X-Ray and Fluoroscopy

The advantage of a chest X-ray is that it cannot only show the outline and size of the heart but also fully reflect the outline of the lung, which is irreplaceable by any other imaging methods (Somerville & Grech, 2009). However, an X-ray cannot clearly show the structure of the heart due to the projections of the heart and the overlapping vessels. Therefore, it is

necessary to combine different projection positions to show the edges of the various atria and the vessels and to judge the change in heart size. On the other hand, chest fluoroscopy has the characteristics of an X-ray, and it is dynamic, can arbitrarily change position, direction and angle, and can observe patients when inspiratory and suffocating in the organ change. However, the disadvantage of chest fluoroscopy is that it cannot save the image and can only be observed in real-time. Besides, patients are exposed to higher doses of radiation for longer periods.

Echocardiography (Echo)

An Echo is a simple, rapid, real-time dynamic observation of the cardiac structure and ventricular myocardial wall movement. It can visualize intracardiac malformations, especially for observing the morphology and function of valves and real-time hemodynamic measurements. Therefore, it is the non-invasive medical imaging technique preferred for diagnosing CHD (X. F. Chen et al., 2017). Transesophageal echocardiography (TEE) and real-time three-dimensional echocardiography (RT3DE) have expanded the clinical application of echocardiography in terms of exploration approaches and imaging modes, respectively. Echo has a non-invasive, no ionizing radiation and no radiation hazard (Lai et al., 2006). In patients with CHD, the ultrasound enhancing agent (UEA) improves the visualization of segmental wall motion of the left and right ventricles, allowing for better quantification of ventricular function at rest and during physiological or drug load. The use of UEA-enhanced Doppler signals is beneficial for quantification of right ventricular (RV) systolic blood pressure in patients with CHD, allowing simultaneous measurement of RV and left ventricular (LV) myocardial perfusion while analyzing ventricular wall motion (Kutty et al., 2016; Porter et al., 2018). The echo can be carried out near the bed and repeatedly detected (Marin et al., 2012). Moreover, the price is relatively lower than other imaging examinations.

However, in practical application, since most of the heart and the great vessels are covered by bone and lung tissue, the location and direction of the exploration are still limited, and the whole course of the vessels cannot be tracked. Besides, the spatial resolution is low, and the field of vision is small, so the anatomical observation of some parts is limited to some extent, such as the development of the middle and distal segments of the left and right pulmonary arteries, as well as the determination of the origin and location of the great vessels. At the same time, ultrasound images are all body layer images, which need multi-level observation to construct stereo images, so the accuracy of examination is related to the theoretical basis and experience of the sonographer.

Therefore, according to the American Society of Echocardiography (ASE), each pediatric cardiac sonographer must obtain a certificate of qualification within a certain period by passing the relevant examination prescribed by an industry-recognized

qualification body. Must have passed a formal course of education in the specialty of ultrasound recognized by a professional association or authority and in accordance with the requirements of a qualifying body. Must be able to perform various echocardiographic examinations and write echocardiographic reports. The skill level must be maintained through participation in appropriate continuing medical education courses (Lai et al., 2006).

Transthoracic Echocardiography (TTE). TTE is the most common type of Echo. It is non-invasive and takes place entirely outside the patient's body. The probe is placed on the left, or right edge of the sternum, the apex of the heart, the supra sternal fovea or the subxiphoid region of the ribs for observation of the aortic valve, left ventricular outflow tract and descending aorta. Compared with CMR, TTE has the advantages of low cost and high availability (Koestenberger et al., 2012).

Transesophageal Echocardiography (TEE). TEE is another echo type in which the probe is placed inside the esophagus or stomach, then scans the heart and vessels from the rear and below. TEE provides regular observations of cardiac valve function, cardiovascular structural development, and ventricular function (Sluysmans & Colan, 2005; Gutgesell & Rembold, 1990). At present, micro-TEE can be widely used in children with CHD (Pavithran et al., 2014). However, this technique is limited in the assessment of complex CHD and extracardiac vascular malformations (Lian et al., 2018).

Real-Time Three-Dimensional Echocardiography (RT3DE). RT3DE has a promising clinical application in CHD; compared with the traditional two-dimensional echocardiography (2DE), which can only display the flat picture of the heart structure, RT3DE expands the vision of scanning the heart and blood vessels from the spatial dimension and the time dimension. Studies have shown that RT3DE can more accurately describe atrial septal defects (ASD) and various congenital valve damage, the fetal heart can be analyzed locally, and the degree of aortic dilatation in Marfan's syndrome can be studied quantitatively (Acar et al., 2004). The new high-resolution RT3DE probe and ultra-high-speed real-time ultrasound image processing platform can real-time, quickly collect images and can timely synchronous display the three-dimensional (3D) image of the heart. RT3DE can observe the cardiac valve structure, pathological property, and abnormal function. It is possible to accurately observe and quantitatively evaluate the heart valves, thus more accurately guiding interventional or surgical treatment of various heart valve diseases. RT3DE images can be obtained in less than 5 minutes, and the use of sedation by the examiner is never required. In 2012, the European and the American societies of echocardiography jointly issued the three-dimensional echocardiography (3DE) operating guidelines (Lang et al., 2012).

However, at present, the spatial resolution of RT3DE in quantitative valve imaging still needs to be improved. Furthermore, the temporal and spatial resolution needs to be improved to obtain the best image quality, and wide-angle acquisition is needed to realize distance visualization (Kleijn & Kamp, 2009). These limitations may cause measurement errors and variations in the results.

The Characteristic of the Echo. Compared with MR and CT, echocardiography has the following advantages: (1) Availability. The echo can be found in almost all hospitals and clinics. However, cardiac examination operators need to undergo a long period of training and evaluation. (2) Portability. Due to the development of portable ultrasound machines, the echo can be taken to the patient's bedside (Rehman et al., 2021). (3) Cost. The initial and operational costs of echocardiography are significantly lower than those of CMR and CT. (4) No radiation. Echo uses high-frequency sound waves to get the image of the heart.

Cardiovascular Magnetic Resonance (CMR)

With the development of Magnetic Resonance Imaging (MRI), CMR has been widely accepted in clinical practice because of its safety and accuracy in assessing myocardial function and showing the overall anatomy of the heart in the diagnosis of CHD, especially CHD with abnormal external great vessels, complex CHD, and postoperative CHD. Compared with Echo, CMR is more accurate in quantifying the shape and function of the heart, allowing the image can be observed repeatedly, and there are no strict requirements for the operator's experience and skill level. In addition, CMR can produce high-resolution images to describe the anatomical structures, such as the size of the heart, valve shape and chamber size, and also provide dynamic information such as angiography, myocardial perfusion, shunt, and abnormal morphological manifestations (Glockner et al., 2003). CMR has the additional advantage that it does not involve radiation. Therefore, CMR and Contrast-enhanced MR Angiography (CE MRA) are valuable tools for assessing associated vascular abnormalities and blood flow patterns (Sreedhar et al., 2005).

However, CMR of CHD generally takes more time to perform than other imaging methods, around 30–60 mins. The presence of implants or other metal objects can sometimes make it difficult to get a clear image, but movement during imaging can have the same effect. MRI can be very intimidating for children because it is noisy; to obtain high-quality images, pediatric patients need to lie completely still in an enclosed space and follow the breath-holding instructions while recording the image. In addition, imaging may be difficult to perform if patients are anxious or confused. Some children may be unable to stay still for the MRI scan, so young children (ages 1 to 6) often need sedatives to help them relax or fall asleep during the exam (Barkovich et al., 2018; Arlachov & Ganatra, 2012).

The Preparation of CMR. (1) Introduce the CMR to parents in detail and ask them to tell the child in advance in a way that the child can understand to ensure that the examination is performed successfully. (2) Before entering the examination room, the doctor checks the patient's medical history and checks if the patient has a metal device implanted or wearing a metal object to avoid damage or artifacts. (3) Anesthesia or sedation is often used in younger children because the examination is long, and the patient needs to remain stationary during the examination to avoid movement artifacts. (4) The contrast agent is given intravenously. (5) Patients need to wear a hearing protection device. (6) The temperature of the pediatric patient should be monitored, and a blanket should be added to keep the patient warm if necessary. (7) Patients need to wear a vital signs monitor. (8) By positioning the image to determine the appropriate coil placement and scanning range (Fratz et al., 2013).

Spin Echo Sequence (SE). The most common variant of SE is the Fast spin echo sequence (Siemens, TSE; GE, FSE; Philips, TSE; Toshiba, Fast SE) and Ultrafast spin echo sequence (Ultrafast SE) (Siemens, HASTE; GE, SSFSE; Philips, SSTSE; Toshiba, Fast-DIET). Ultrafast SE has a fast-scanning speed, collects one layer of cross-sectional images in one cardiac cycle, and is insensitive to breathing and motion artifacts so that it can collect images in free breathing (Lu & Zhao, 2010). First, since the sequence is black blood, axial and sagittal zero interval tomography images can roughly understand the relationship between atrioventricular and ventricle aorta, the development of ventricle and aorta, and atrioventricular valve development. Second, it can be positioned for other sequential scans. For example, two-chamber, four-chamber, and ventricular effluent tract films are all based on HASTE location scans in Siemens. However, the black blood sequence cannot provide optimal cine information. Compared with ultrafast gradient echo sequence (Ultrafast GRE), SE has a lower signal-to-noise ratio (SNR) (Ginat et al., 2011).

Gradient Echo Sequence (GRE). The GRE may use the Spoiled gradient echo sequence, Ultrafast GRE, and the Steady-state gradient echo sequence. GRE is an alternative technique to SE, which can identify the abnormal function of the heart wall and evaluate the bilateral ventricular function. The Cine sequence is characterized by high temporal resolution and high sensitivity to flow changes. This sequence is commonly used for preoperative diagnosis and postoperative follow-up of complex CHD, such as double outlet right ventricle (Yao et al., 2016).

Steady-state gradient echo sequence (Siemens, True-FISP; GE, FIESTA; Philips, balance-FFE; Toshiba, True SSFP) can be used to comprehensively observe the anatomical connection and function of the heart and the great vessels, which is conducive to the display of intracardiac malformations. It is characterized by fast scanning speed and high SNR. Among them, Time-adaptive sensitivity encoding (TSENSE) technology is not sensitive

to breathing and mild motion artifacts, so it is very suitable for arrhythmia patients and infants who cannot cooperate with breath-holding (Jing et al., 2017; Lu & Zhao, 2010)

Contrast Enhanced MR Angiography (CE MRA). CE MRA is mainly used to display vascular malformation, including PDA and artery malformation (Lv et al., 2010). However, MRI contrast agents are associated with nephrogenic systemic fibrosis and should be used with caution in patients with renal insufficiency (Zeng et al., 2014; Elmholdt et al., 2011; Davison & Mead, 2010).

Computed Tomography (CT)

CT has good spatial resolution and can provide a three-dimensional reconstruction of the heart. It can clearly and objectively show the heart structure, the great vessels, the relationship between each atrioventricular valve and ventriculoarterial valve, intracardiac shunt, left and right ventricular function, and the volume of the heart chamber. With the development of technology, the CT scanning time is shortened while maintaining high image quality (Larson et al., 2011). Since the introduction of 64-slice CT, multi-slice spiral CT (MSCT) has been widely recognized as a non-invasive examination for the diagnosis of CHD (Crean, 2007; Ropers et al., 2006). In addition, most hospitals have CT equipment, which is usually cheaper than MRI (Fisher & Sawyers, 2020; Davis & Cunha, 2019). Therefore, CT has developed from the original single-slice to 64-slice, 256-slice, 320-slice spiral CT, and even more advanced dual-source CT (DSCT) (Siegel, 2003). A conventional chest CT scan can show the vessels of the heart, and the adjacent relationship between the mediastinal organs and tissues, so it is helpful to display pericardial effusion, thickening and calcification.

The main characteristic of CT is its excellent spatial resolution, which is better than MRI and Echo, adequate temporal resolution, and relatively poor contrast resolution. Compared with MSCT, flat volume CT have excellent spatial resolution and wider coverage, but the contrast resolution is less than MSCT. The high temporal resolution of cardiac CT is mainly achieved by rapid frame rotation time and partial scan reconstruction (Lin & Alessio, 2009).

The consequent disadvantage to obtaining a clear image is that the radiation dose is increased, and the scanning time is prolonged. In addition, infants and children are in the stage of physical development, and their tissues and organs are richer in water than adults. As a result, infants and children are at a significantly higher risk of radiation injury or damage than adults. Therefore, more attention should be paid to the radiation dose received by infants and children during CT examinations (Brenner & Hall, 2007). There are multiple data acquisition and scanning parameters in CT, non-gated acquisition mode, retrospective electrocardiogram (ECG)-gated spiral acquisition mode, and prospective ECG-triggered sequential acquisition mode.

The Technical Characteristics of MSCT. Due to the high spatial resolution and good image quality, MSCT can display the details of cardiac deformities, especially the deformities in the connection of the great blood vessels, collateral circulation, progression and development of coronary arteries, and development of other organs related to the heart (Feng & Xing, 2012; Liu & Wang, 2017). It has high temporal resolution and reduces the inspection time. Continuous-level tomography avoids image overlap, and a variety of post-processing methods, such as volume reconstruction, surface reconstruction, multi-plane reconstruction, and maximum density projection, ensure the accurate expression of lesions. However, MSCT does not provide enough information about the cardiovascular function and hemodynamics (Enaba et al., 2017, Zhang et al., 2012).

The Preparation of MSCT. (1) The iodine contrast agent allergy test negative due to the poor expression ability of infants, respiratory, heart rate, blood pressure and skin mucosal changes should be closely monitored. (2) Establishment of injection route: superficial venous injection through the lower extremities was preferred to minimize the effect of contrast artifact. Secondly, the contrast agent should be injected through the superficial vein of the right upper limb to prevent the left innominate vein from interfering with the observation of the aortic arch. If an abnormal return of the left superior vena cava is suspected, it can be injected through the left upper limb superficial vein. If an abnormal return of the inferior vena cava is suspected, it should be injected through the superficial veins of both legs. The high concentration of contrast agent injected through the upper limb will affect the imaging of the brachiocephalic trunk. (3) Connecting electrocardiogram (ECG): to prevent metal artifacts caused by the electrode, it should be noted that the electrode should be placed far away from the scanning area. Usually, two upper limbs and an abdomen should be selected for young children. (4) Breath-holding training: for children who can cooperate with the examination, breath-holding training should be carried out. (5) Anesthesia is carried out by a professional anesthetic doctor for children who cannot cooperate with the examination. (6) ECG, respiration, and peripheral oxygen saturation should be monitored.

The Contraindications of MSCT. (1) Allergic to the iodine contrast agent. (2) Patients with acute or chronic renal failure. (3) Patients with severe cardiac insufficiency or corrected heart failure. (4) Patients with severe arrhythmias.

The Various Post-processing Technologies. (1) Volume Rendering (VR). VR can display the spatial position relationship and movement of the heart and large blood vessels. However, due to the need to combine with the image processing method of partial cutting, the operation is more complex and experienced operators are needed. (2)

Maximum Intensity Projection (MIP). MIP can display the anatomical structure of the heart and the large blood vessels from multiple directions and angles, and the thickness can be changed arbitrarily, with high measurement accuracy. However, the dense tissue blocks angiography. (3) Curved Planar Reconstruction (CPR). CPR can better show the calcification and distorted and stented vascular lumen. However, due to the high dependence on the accuracy of the probe, a single curve cannot adequately show eccentric lesions. In addition, hyperperfused vessels are difficult to show, so CPR is used relatively infrequently. (4) Multiplanar reconstruction (MPR) (Figure 3). The data obtained from the axial plane are converted into coronal, sagittal, or sloped non-axial planes. However, the reconstruction time is longer, and there is more data to review and archive each case.

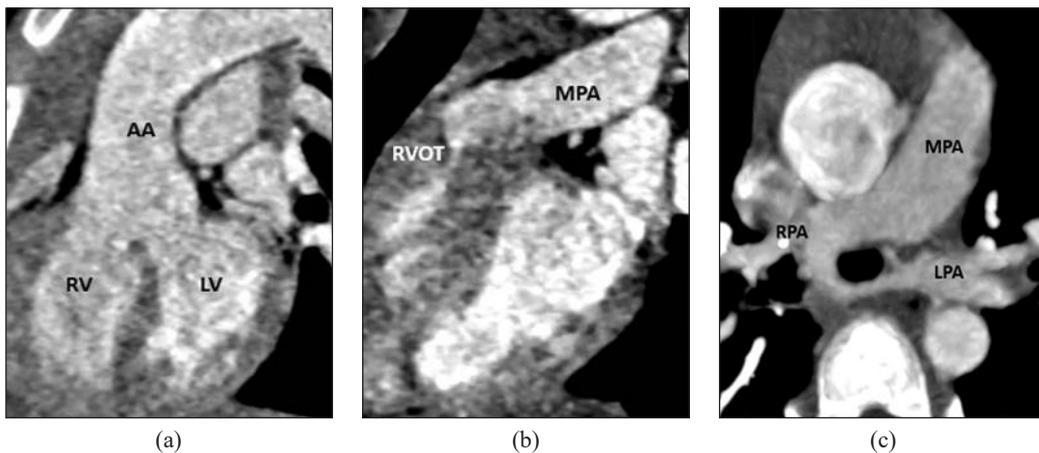


Figure 3. A 5-year girl with VSD (Prospective ECG-triggering DSCT angiography). (a) Oblique coronal MPR image, (b) Oblique sagittal MPR image, (c) axial thin-section MIP image. AA= ascending aorta, MPA = main pulmonary artery, RPA = right pulmonary artery, SVC = superior vena cava. (Source: Nie et al., 2014)

Retrospective ECG-Gated Spiral Acquisition. The retrospective ECG-gated spiral acquisition mode adopts the small-pitch spiral scanning. This model is a whole cardiac cycle scan. The retrospective ECG-gated acquisition is similar to cameras, which can take continuous pictures by pressing the shutter at one time, taking numerous exposures and pictures, whether clear or useful, to select the best or several images in retrospect. ECG data is collected synchronously during CT scan, and the best data is selected for reconstruction according to the time phase of ECG data in retrospect. Therefore, continuous radiation exposure is performed in multiple consecutive heartbeat cycles to obtain clear images, leading to a higher radiation dose in retrospective ECG-gated spiral acquisition compared with that in prospective ECG-triggered sequential acquisition. The advantage of the retrospective ECG-gated protocol is that the image quality of the retrospective ECG-gated spiral mode was better than that of the prospective ECG-triggered sequential mode (Wang et al., 2017).

In addition, the retrospective ECG-gated spiral mode is not limited by heart rate and is applicable to patients with irregular heart rate, premature beat and fast heart rate (>70 beats per minute, bpm) (Stolzmann et al., 2011; Nakagawa et al., 2016).

However, the image quality of DSCT is similar to prospective and retrospective ECG-gated techniques, which can significantly reduce the effective dose of patients, and has important clinical value for subjects who cannot receive a high radiation dose and whose heart rate is lower than 67 bpm (Liu & Wang, 2017; Wang et al., 2017; Pache et al., 2011; Xu et al., 2010; Ünal et al., 2015).

Prospective ECG-Triggered Sequential Acquisition. Prospective ECG-triggered sequential acquisition mode is not a spiral scanning. The mechanism of the prospective ECG-triggered protocol is that the X-ray tube is selectively turned on only when triggered by the ECG signal, and the scan is triggered only when the heart is in diastole for the data collection. The main advantage of the prospective ECG-triggered sequential acquisition mode is that the X-ray exposure time is short. Therefore, the radiation dose is lower than that of the retrospective ECG-gated spiral acquisition mode.

However, as data collection is only carried out at a specific stage of a cardiac cycle and the machine rotation speed is limited, there are higher requirements for heart rate control and stability, so the function of cardiac valve or ventricular wall movement cannot be evaluated (Husmann et al., 2008; Stolzmann et al., 2008). The mode can set the scanning window in the R-R interval, which is very demanding for heart rate, rhythm, and breath hold. It is applicable to patients with a steady heart rate below 70 bpm or 65 bpm, so patients should not have a significant arrhythmia; otherwise, some data may be distorted and affect the accuracy of diagnosis (Roberts et al., 2008; Earls et al., 2008; Gopal et al., 2009).

Non-ECG-Gated Acquisition. Non-ECG-gated scanning is a spiral scan, a simple and feasible technique for CT in patients with CHD. It scans fast and is not limited by the heart rate of examinees. Since the introduction of MSCT, the greatest advantage of DSCT lies in cardiac imaging, where the dose of radiation in cardiac scanning is only 50% of that of conventional CT (Li et al., 2009), so cardiac CT with this scanning protocol has been widely used. In addition, this scanning protocol can be performed during breath-holding or free-breathing (Goo, 2010).

Furthermore, it can display the structure and malformations of the extracardiac vessels, particularly the origin and proximal segment of the coronary artery (Goo et al., 2005). In non-ECG-gated synchronous spiral CT scanning, the large pitch scanning mode is usually adopted for chest scanning, and the coronary arteries can be displayed. It is because the larger the pitch used, the faster the scanning speed and the shorter the scanning time under the same condition. Compared with retrospective ECG-gated imaging, non-ECG-gated

synchronous imaging had a shorter time to obtain the images. However, retrospective ECG-gated protocols generated fewer cardiac motion artifacts in the aortic root (Horehledova et al., 2020).

Dual-Source CT for Cardiac Scanning. Prospective ECG-triggered sequential acquisition in DSCT freely set the width of the acquisition window between 200ms and 380ms during R-R data collection. Retrospective ECG-gated spiral acquisition in DSCT is the whole R-R interval collection. Under the advance guarantee of a full dose acquisition window of > 83ms, the ray dose of other periods in the R-R period can be set arbitrarily from 4% to 100% (Li et al., 2016; X. F. Chen et al., 2017).

320-Slice CT for Cardiac Scanning. The 320-slice CT has a detector of 16cm, which can cover the whole heart for children. One rotation only takes 0.35s, avoiding the influence of heartbeat, breathing and other factors on the image quality. The scan time is short, and the children are not easily wake up during the examination. The operability is better than Echo and MRI, and it is easier to accept by the children's families. Various abnormal shunts of the heart can be identified by dynamic volume scanning. The 320-slice CT eliminates the overlap of spiral CT scans and the redundant data needed for interpolation reconstruction algorithms, and the single cardiac scan dose is 3 to 4 mSv. However, it does not reflect the pressure and oxygen of the degree of pulmonary hypertension, which is a disadvantage of all non-invasive imaging of the heart (Sreedher et al., 2021).

To sum up, imaging diagnosis of CHD has entered a new era of how to apply correctly and efficiently different non-invasive examination methods such as Echo, CMR and CT, which requires closer collaboration between clinicians and radiologists and continuously optimize these methods to make the most of them. In Asian countries, pediatric cardiac CT is more commonly used than pediatric cardiac CMR (Tsai & Goo, 2013). However, compared with CMR and Echo, CT has radiation, and the risk of radiation exposure cannot be ignored. Therefore, the clinicians and technicians must minimize the radiation dose while maintaining adequate spatial, contrast and temporal resolution. The five imaging modalities for diagnosing children were tabulated in Table 1.

RADIATION AWARENESS

Radiation exposure is a method of measuring air ionization due to ionizing radiation of photons, defined as the charge of electricity released by such radiation in a particular air volume divided by the air's mass. All objects in nature emit radiation if the temperature is above zero absolute temperature, and most of the radiation is produced by natural minerals. Even the muscles, bones and tissues of the human body contain naturally occurring radioactive elements. Humans are exposed to natural radiation from the earth and beyond,

Table 1
Strengths and weaknesses of plain chest X-ray, fluoroscopy, CT, CMR, echo in children with CHD

Modality	Plain chest X-ray	Fluoroscopy	CT	CMR	Echo
Cost	+	+	++	+++	++
Examination time	+	+	++	+++	+++
Portability	+++	-	-	-	+++
Monitor*	+++	+++	++	+	+++
Sedative requirement	-	-	+	+++	-
Imaging data save	+	+	+++	+++	++
Need contrast media	-	-	+++	++	+
Limits of image reconstruction	+++	+++	-	-	+++
Contrast resolution	-	-	+	+++	+
Spatial resolution	-	-	+++	+	++
Temporal resolution	-	-	+	++	+++
Radiation dose	+	+	++	-	-

- = not available; + = low; ++ = moderate; +++ = high

*Monitor is for ease of monitoring the patient during the examination.

and radiation obtained from outer space is called cosmic radiation. In addition, other sources of radiation may be artificial, such as X-rays used to diagnose disease and treat cancer. For example, CT scans have higher radiation doses than traditional x-ray imaging procedures, and the ionizing radiation cause side effects (Slovis, 2003; ICRP, 2007).

According to Volume II of the UNSCEAR (2013) Report: Effects of radiation exposure on children, DNA molecules are prime targets for ionizing radiation, which kills cells and causes them to transform. Epidemiological studies have demonstrated that moderate to high doses of ionizing radiation can cause and promote cellular changes that lead to cancer, meaning that individuals who receive a dose of radiation are at risk for cancer (UNSCEAR, 2006). Children grow and metabolize faster than adults, meaning the side effects of radiation doses are more damaging to children. Children with CHD have more opportunities and time to be exposed to radiation, which increases the risk of other diseases. CT scans emit a high degree of ionizing radiation: About 11% of CT scans are pediatric (Mazrani et al., 2007). Studies found that every 500 children under 15 died from cancer caused by abdominal CT scans. Ionizing radiation is associated with brain tumors, and 1 in 1,500 children under 15 died from cancer caused by brain CT radiation (Shrimpton et al., 2005). In 2012, A retrospective cohort study published online by the Lancet medical journal revealed that CT scanning in children with a cumulative dose of about 50mGy might mostly triple the risk of leukemia, and the cumulative dose of around 60mGy were three times more likely to develop brain tumors. Therefore, it is recommended that children should adopt lower scanning conditions (Pearce et al., 2012). With the rapid increase and

advancement of CT examination techniques in medicine, the correct use of radiological examination has become very important.

As a principle of dose control, the ALARA (as low as reasonably achievable) concept (the use of the lowest dose of radiation to obtain the appropriate image quality for accurate diagnosis) is generally accepted in the radiation field. However, blindly reducing the radiation dose will lead to the deterioration of image quality, resulting in misdiagnosis or rescan (Seibert, 2004); therefore, from the perspective of radiological protection, the reasons for radiological examination must be clarified before CT examination. In addition, clinically necessary to provide patient information and suspected diagnosis derived from symptoms and physical examination. This way, protection is optimized to ensure that patients and personnel are not exposed to excessive radiation.

GUIDELINE

Children's imaging examination should follow the principles of non-invasive before invasive, simple before complex, economical before expensive. During the inspection and optimization, measures should be taken according to local conditions, starting from the actual situation of the local area and the children themselves. For example, due to the high cost and lack of availability of MRI, CT is generally more suitable to evaluate pediatric CHD than MRI in Asian countries (Tsai & Goo, 2013). In China, MRI examinations are made by appointment due to the long MRI examinations and the large population base, which is not as convenient and fast as CT. The premise of a successful CT examination of CHD in children and infants is to check and record patient information carefully. The correct positioning of the patient position on the CT gantry and the use of laser beam positioning is to avoid the incorrect position and excessive scanning range, which will lead to the deterioration of image quality and excessive radiation exposure (Booij et al., 2016). In infants and children, prolonged or repeated use of sedatives can lead to neurodevelopmental disorders, particularly in children with CHD, with a higher risk of adverse side effects (Char et al., 2016). Children who do not cooperate may be tested with mild sedation, as the time required for CT examination of CHD is shorter than that for MRI. If sedatives are used after the CT scan, pediatric patients should be monitored and observed until awakened.

It is generally recommended that patients hold their breath during the scan to avoid breathing movement artifacts. Breath-holding scanning can improve image quality (Goo, 2010), but for young children and infants, breathing cannot be controlled; for them, use free breathing during the scan. First, injecting the contrast agent through the superficial vein of the lower extremities is preferable to minimize the influence of the contrast agent's artifacts. Secondly, the contrast agent should be injected through the superficial vein of the right upper limb to prevent the left innominate vein from interfering with the observation of the aortic arch. If an abnormal return of the left superior vena cava is suspected, the injection

can be made through the superficial vein of the left upper limb. If an abnormal return of the inferior vena cava is suspected, injections should be made through the superficial veins of the legs. The high-concentration contrast agent injected through the upper limbs will affect the imaging of the brachiocephalic trunk.

As mentioned above, CT has several scanning techniques that can be used to evaluate children with CHD, including non-ECG spiral scan, retrospective ECG-gated spiral acquisition, and that prospective ECG-triggered sequential acquisition. Different techniques depend on what needs to be observed. For example, the non-ECG spiral scan allows the evaluation of extrathoracic vessels. In addition, in free-breathing children with CHD, a high-pitch dual-source non-ECG spiral scan reduces cardiac and respiratory motion artifacts in thoracic CT (Sriharan et al., 2016).

Many adult studies have shown that prospective ECG-triggered sequential acquisition is recommended for use in patients with stable and low heart rates (≤ 70 bpm). However, this model has not been widely used in young children with a rapid heart rate and inability to control breathing; the main advantage of this pattern is the low radiation dose, and intracardiac structures and coronary arteries can be observed and functional assessments performed (Hausleiter et al., 2012; Earls et al., 2008; Roberts et al., 2008; Gopal et al., 2009). However, although the coronary artery and functional information can be obtained from prospective sequential gating, they are less optimal than retrospective gating.

A retrospective ECG-gated spiral scan provides excellent image quality (Goo, 2010), but its low pitch value results in a high radiation dose. In addition, the breath should be held during scanning to avoid breathing movement artifacts, which observes the intracardiac structure and coronary arteries and assess cardiac function (Jin et al., 2010).

The time has come to obtain good diagnostic images at the minimum cost of radiation exposure. Therefore, we list a guideline, which includes patient preparation, scanning parameters, scanning techniques, radiation dose reduction and optimization in Table 2.

PROBLEMS AND PROSPECTS

Cardiac imaging can provide valuable guidance for treating CHD in all stages. Furthermore, advances in the field of cardiac imaging have greatly improved the prognosis of these patients. Echo can be real-time, dynamic, multi-section two-dimensional imaging, combined with Doppler technology, which has great advantages in displaying CHD intracardiac malformations. However, it is affected by the acoustic window, the display of extracardiac malformations is not good, and the personal technical requirements of the sonographer are very high, so the diagnosis of complex CHD is limited (Lai et al., 2006).

Nowadays, low-dose CT has become a common focus in the medical imaging field. The management and supervision institutions of the medical and health industry constantly introduce relevant laws and regulations to guide manufacturers and scientific research

Table 2
Guidelines for CT scan of CHD in children

Radiologists Preparation
<ol style="list-style-type: none"> 1. The examination is conducted under the guidance of an experienced radiologist. 2. Ensure the quality of diagnostic images.
Patient Preparation
<ol style="list-style-type: none"> 1. Check the patient's name and inpatient/outpatient number. Record patient age, height, weight, body mass index (BMI), cooperation ability, and vital signs. 2. Using minimal or no sedatives in infants and children is recommended to avoid neurodevelopmental disorders. If sedatives were used, pediatric patients were monitored and observed until they were awake after a CT scan.
Scanning Parameters
<ol style="list-style-type: none"> 1. Recommend holding breath during the scan to avoid breathing movement artifacts; for young children and infants, use free breathing while scanning. 2. Scanning range: pediatric cardiothoracic CT scans typically include the upper aortic arch and the diaphragm. The scanning range to assess the pulmonary veins is from supraclavicular to subphrenic.
Scanning Techniques
<ol style="list-style-type: none"> 1. Set according to the patient's medical history and clinical requirements. 2. Non-ECG-gated acquisition <ol style="list-style-type: none"> (1) Using high-pitch dual-source scanning to observe the extracardiac and cardiac structures. (2) This pattern can be used in free-breathing children with CHD. (3) High-pitch dual-source non-ECG spiral scan reduces cardiac and respiratory motion artifacts in thoracic CT. 3. Retrospective ECG-gated spiral acquisition <ol style="list-style-type: none"> (1) Observe intracardiac structures and coronary arteries and perform a functional evaluation. (2) The radiation dose of this technique is higher than prospective ECG-triggered sequential acquisition. (3) This pattern provides excellent image quality (4) Breath should be held during scanning to avoid breathing movement artifacts. 4. Prospective ECG-triggered sequential acquisition. <ol style="list-style-type: none"> (1) Observe intracardiac structures and coronary arteries and perform a functional evaluation. (2) The radiation dose of this technique is lower than retrospective ECG-gated spiral acquisition.
Radiation Dose Reduction and Optimization
<ol style="list-style-type: none"> 1. Follow the ALARA principle.
The Iodinated Contrast Medium
<ol style="list-style-type: none"> 1. Recommend the tri-phasic injection protocol. 2. The appropriate injection method is selected according to the patient's condition and medical history. 3. The injection rate depends on the size of the vascular catheter and CT indications.

BMI = body mass index, ALARA = As Low As Reasonably Achievable

institutions to develop more advanced and effective innovative designs. At the same time, medical practitioners should be regulated to rationally use low doses in practical work, abide by the ALARA principle, and try to reduce patient doses to meet the requirements of clinical diagnosis.

MSCT, especially 64-slice CT, has been used to diagnose CHD in infants and young children, but its temporal resolution is low. The patient's heart rate must be less than 65 bpm

to ensure the image quality of coronary artery imaging. The heart rate of newborns and infants is often 150 bpm, so it is difficult to show the origin and course of the coronary arteries clearly.

Compared with ordinary 64-slice CT scanning technology, DSCT has higher temporal resolution and relatively low heart rate and breath-holding requirements. In addition, there are more scanning modes to choose from: personalization of scanning. As the pitch increases, the scanning speeds up, and the radiation dose can be reduced.

320-slice CT has extremely high temporal and spatial resolution and powerful post-processing functions. It can clearly and intuitively display most intra- and extracardiac malformations of complex CHD. It is safe and non-invasive, and the scanning dose is low, so it is very suitable for the examination of newborns and infants. Compared with other imaging methods, its clinical application value is more reflected in the accurate display of the anatomical morphology and collateral circulation of the connection between the heart and the great blood vessels, the part of the extracardiac great blood vessels. The severity of the collateral circulation will determine whether the child can undergo surgery, the formulation of the surgical plan, and the prognosis after surgery. Showing the origin and course of the coronary arteries is of great significance for the preparation of a detailed surgical plan before surgery, thereby minimizing intraoperative risks (Sreedher et al., 2021).

MRI has no ionizing radiation, does not necessarily require intravenous contrast and can acquire images at multiple angles. However, the temporal and spatial resolution is relatively low, the scanning time is long, and the large noise during the scanning process makes it difficult to obtain better imaging results in newborns and infants.

In addition, for patients, the cost and cost-effectiveness of the test also need to be considered by the patient's family and the medical field.

CONCLUSION

In recent years, medical imaging equipment and technology have been developed and promoted, and various new equipment and technology that can be used in the pediatric cardiac examination and imaging diagnosis of CHD have entered a new era. While each method has its inherent advantages and disadvantages, radiologists, physicists, manufacturers and national supervisory institutions must work together to fully recognize the strong sensitivity of infants and children to radiation in order to ensure access to the lowest dose of diagnostic imaging and the rational use of low dose, achieve individualized exposure doses, and effectively coordinate multiple imaging methods to improve the diagnosis and treatment of heart disease in children and infants. For patients, avoid unnecessary inspection as much as possible and cooperate with doctors during the scan to reduce unnecessarily repeated scans.

In this review, we introduced several current cardiovascular imaging methods for CHD. Nowadays, the integration of various imaging technologies has been widely recognized and developed for the accuracy and clinical practicability of the diagnosis of CHD in children. Therefore, radiologists and sonographers perform MPR through workstations to obtain more and clearer images, broaden diagnostic thinking and reflect the complementary advantages of imaging technology. Therefore, it is conceivable that the combination of multiple imaging technologies will enter a new, exciting, and high potential stage in diagnosing and treating CHD.

ACKNOWLEDGEMENT

The authors gratefully acknowledge all journals that permitted them to use images in papers published in their journals. The authors declare no potential conflicts of interest.

REFERENCES

- Acar, P., Dulac, Y., Taktak, A., & Villacèque, M. (2004). Echocardiographie 3D temps réel des cardiopathies congénitales [Real time 3D echocardiography in congenital heart disease]. *Archives des maladies du coeur et des vaisseaux*, 97(5), 472-478.
- Arlachov, Y., & Ganatra, R. H. (2012). Sedation/anaesthesia in paediatric radiology. *The British Journal of Radiology*, 85(1019), e1018-e1031. <https://doi.org/10.1259/bjr/28871143>
- Barkovich, M. J., Xu, D., Desikan, R. S., Williams, C., & Barkovich, A. J. (2018). Pediatric neuro MRI: Tricks to minimize sedation. *Pediatric Radiology*, 48(1), 50-55. <https://doi.org/10.1007/s00247-017-3785-1>
- Booij, R., Dijkshoorn, M. L., van Straten, M., du Plessis, F. A., Budde, R. P. J., Moelker, A., Krestin, G. P., & Ouhlous, M. (2016). Cardiovascular imaging in pediatric patients using dual source CT. *Journal of Cardiovascular Computed Tomography*, 10(1), 13-21. <https://doi.org/10.1016/j.jcct.2015.10.003>
- Brenner, D. J., & Hall, E. J. (2007). Computed tomography - An increasing source of radiation exposure. *The New England Journal of Medicine*, 357(22), 2277-2284. <https://doi.org/10.1056/NEJMra072149>
- Char, D., Ramamoorthy, C., & Wise-Faberowski, L. (2016). Cognitive dysfunction in children with heart disease: The role of anesthesia and sedation. *Congenital Heart Disease*, 11(3), 221-229. <https://doi.org/10.1111/chd.12352>
- Chen, D., Zhong, H. B., Lin, D. X., Chen, W. R., & Xie, L. D. (2017). The predictive value of echocardiography combined with electrocardiogram in the diagnosis of pulmonary arterial hypertension associated with congenital heart disease. *Chinese Journal of Hypertension*, 25(8), 767-772. <https://doi.org/10.16439/j.cnki.1673-7245.2017.08.024>
- Chen, X. F., Jiang, F., Li, L., Chen, Y., Chen, X., Jiang, Y. Y., Xiang, L., & Ma, X. J. (2017). Application of low-dose dual-source computed tomography angiography in children with complex congenital heart disease. *Experimental and Therapeutic Medicine*, 14(2), 1177-1183. <https://doi.org/10.3892/etm.2017.4591>
- Crean, A. (2007). Cardiovascular MR and CT in congenital heart. disease. *Heart (British Cardiac Society)*, 93(12), 1637-1647. <https://doi.org/10.1136/hrt.2006.104729>

- Davis, C. P., & Cunha, J. P. (2019). *CT scan vs. MRI differences between machines, costs, uses*. MedicineNet. https://www.medicinenet.com/ct_scan_vs_mri/article.htm
- Davison, R., & Mead, P. (2010). Nephrogenic systemic fibrosis (NSF): The role of tamoxifen. *NDT Plus*, 3(5), 505-512. <https://doi.org/10.1093/ndtplus/sfq111>
- Earls, J. P., Berman, E. L., Urban, B. A., Curry, C. A., Lane, J. L., Jennings, R. S., McCulloch, C. C., Hsieh, J., & Londt, J. H. (2008). Prospectively gated transverse coronary CT angiography versus retrospectively gated helical technique: Improved image quality and reduced radiation dose. *Radiology*, 246(3), 742-753. <https://doi.org/10.1148/radiol.2463070989>
- Elmholdt, T. R., Pedersen, M., Jørgensen, B., Søndergaard, K., Jensen, J. D., Ramsing, M., & Olesen, A. B. (2011). Nephrogenic systemic fibrosis is found only among gadolinium-exposed patients with renal insufficiency: A case-control study from Denmark. *The British Journal of Dermatology*, 165(4), 828-836. <https://doi.org/10.1111/j.1365-2133.2011.10465.x>
- Enaba, M. M., Hasan, D. I., Alsowey, A. M., & Elsayed, H. (2017). Multidetector computed tomography (CT) in evaluation of congenital cyanotic heart diseases. *Polish Journal of Radiology*, 82, 645-659. <https://doi.org/10.12659/PJR.903222>
- Feng, C. Z., & Xing, F. Q. (2012). Diagnose of neonate with complex congenital heart disease on 64-detector CT. *Journal of Chengdu College*, 7(2), 7-14.
- Fisher, J. K., & Sawyers, T. (2020). *CT scan vs. MRI*. Healthline. <https://www.healthline.com/health/ct-scan-vs-mri>
- Fratz, S., Chung, T., Greil, G. F., Samyn, M. M., Taylor, A. M., Buechel, E. R. V., Yoo, S. J., & Powell, A. J. (2013). Guidelines and protocols for cardiovascular magnetic resonance in children and adults with congenital heart disease: SCMR expert consensus group on congenital heart disease. *Journal of Cardiovascular Magnetic Resonance*, 15(1), Article 51. <https://doi.org/10.1186/1532-429X-15-51>
- Gilboa, S. M., Devine, O. J., Kucik, J. E., Oster, M. E., Riehle-Colarusso, T., Nembhard, W. N., Xu, P., Correa, A., Jenkins, K., & Marelli, A. J. (2016). Congenital heart defects in the United States: Estimating the magnitude of the affected population in 2010. *Circulation*, 134(2), 101-109. <https://doi.org/10.1161/CIRCULATIONAHA.115.019307>
- Ginat, D. T., Fong, M. W., Tuttle, D. J., Hobbs, S. K., & Vyas, R. C. (2011). Cardiac imaging: Part 1, MR pulse sequences, imaging planes, and basic anatomy. *American Journal of Roentgenology*, 197(4), 808-815. <https://doi.org/10.2214/AJR.10.7231>
- Glockner, J. F., Johnston, D. L., & McGee, K. P. (2003). Evaluation of cardiac valvular disease with MR imaging: Qualitative and quantitative techniques. *Radiographics*, 23(1), Article e9. <https://doi.org/10.1148/rg.e9>
- Goo, H. W. (2010). State-of-the-art CT imaging techniques for congenital heart disease. *Korean Journal of Radiology*, 11(1), 4-18. <https://doi.org/10.3348/kjr.2010.11.1.4>
- Goo, H. W., Park, I. S., Ko, J. K., Kim, Y. H., Seo, D. M., Yun, T. J., & Park, J. J. (2005). Visibility of the origin and proximal course of coronary arteries on non-ECG-gated heart CT in patients with congenital heart disease. *Pediatric Radiology*, 35(8), 792-798. <https://doi.org/10.1007/s00247-005-1482-y>

- Gopal, A., Mao, S. S., Karlsberg, D., Young, E., Waggoner, J., Ahmadi, N., Pal, R. S., Leal, J., Karlsberg, R. P., & Budoff, M. J. (2009). Radiation reduction with prospective ECG-triggering acquisition using 64-multidetector computed tomographic angiography. *The International Journal of Cardiovascular Imaging*, 25(4), 405-416. <https://doi.org/10.1007/s10554-008-9396-z>
- Gutgesell, H. P., & Rembold, C. M. (1990). Growth of the human heart relative to body surface area. *The American Journal of Cardiology*, 65(9), 662-668. [https://doi.org/10.1016/0002-9149\(90\)91048-b](https://doi.org/10.1016/0002-9149(90)91048-b)
- Hausleiter, J., Meyer, T. S., Martuscelli, E., Spagnolo, P., Yamamoto, H., Carrascosa, P., Anger, T., Lehmkuhl, L., Alkadhi, H., Martinoff, S., Hadamitzky, M., Hein, F., Bischoff, B., Kuse, M., Schömig, A., & Achenbach, S. (2012). Image quality and radiation exposure with prospectively ECG-triggered axial scanning for coronary CT angiography: The multicenter, multivendor, randomized Protection-III study. *JACC: Cardiovascular Imaging*, 5(5), 484-493. <https://doi.org/10.1016/j.jcmg.2011.12.017>
- Hoffman, J., & Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, 39(12), 1890-1900. [https://doi.org/10.1016/s0735-1097\(02\)01886-7](https://doi.org/10.1016/s0735-1097(02)01886-7)
- Horehledova, B., Muhl, C., Boswijk, E., Crombag, G., Nijssen, E. C., Nelemans, P. J., Veenstra, L. F., Wildberger, J. E., & Das, M. (2020). Retrospectively ECG-gated helical vs. non-ECG-synchronized high-pitch CTA of the aortic root for TAVI planning. *PLoS One*, 15(5), Article e0232673. <https://doi.org/10.1371/journal.pone.0232673>
- Hu, S. (2020). Report on cardiovascular health and diseases in China 2019: Updated summary. *Chinese Circulation Journal*, 35(9), 833-854.
- Husmann, L., Valenta, I., Gaemperli, O., Adda, O., Treyer, V., Wyss, C. A., Veit-Haibach, P., Fuminari, T., Schulthess, G. K., & Kaufmann, P. A. (2008). Feasibility of low-dose coronary CT angiography: First experience with prospective ECG-gating. *European Heart Journal*, 29(2), 191-197. <https://doi.org/10.1093/eurheartj/ehm614>
- ICRP. (2007). 2007 Recommendations of the international commission on radiological protection. *Annals of the ICRP*, 37(2-4), 1-94. <https://doi.org/10.1016/j.icrp.2007.10.003>
- Jin, K. N., Park, E. A., Shin, C. I., Lee, W., Chung, J. W., & Park, J. H. (2010). Retrospective versus prospective ECG-gated dual-source CT in pediatric patients with congenital heart diseases: Comparison of image quality and radiation dose. *The International Journal of Cardiovascular Imaging*, 26(1), 63-73. <https://doi.org/10.1007/s10554-009-9579-2>
- Jing, L., Pulenthiran, A., Nevius, C. D., Mejia-Spiegeler, A., Suever, J. D., Wehner, G. J., Kirchner, H. L., Haggerty, C. M., & Fornwalt, B. K. (2017). Impaired right ventricular contractile function in childhood obesity and its association with right and left ventricular changes: A cine DENSE cardiac magnetic resonance study. *Journal of Cardiovascular Magnetic Resonance*, 19(1), Article 49. <https://doi.org/10.1186/s12968-017-0363-5>
- Kleijn, S. A., & Kamp, O. (2009). Clinical application of three-dimensional echocardiography: Past, present and future. *Netherlands Heart Journal*, 17(1), 18-24. <https://doi.org/10.1007/BF03086210>
- Koestenberger, M., Friedberg, M. K., Ravekes, W., Nestaas, E., & Hansmann, G. (2012). Non-invasive imaging for congenital heart disease: Recent innovations in transthoracic echocardiography. *Journal of Clinical & Experimental Cardiology*, S8(002), 1-5. <https://doi.org/10.4172/2155-9880.S8-002>

- Kutty, S., Xiao, Y., Olson, J., Xie, F., Danford, D. A., Erickson, C. C., & Porter, T. R. (2016). Safety and efficacy of cardiac ultrasound contrast in children and adolescents for resting and stress echocardiography. *Journal of the American Society of Echocardiography*, *29*(7), 655-662. <https://doi.org/10.1016/j.echo.2016.02.019>
- Lai, W. W., Geva, T., Shirali, G. S., Frommelt, P. C., Humes, R. A., Brook, M. M., Pignatelli, R. H., & Rychik, J. (2006). Guidelines and standards for performance of a pediatric echocardiogram: A report from the Task Force of the Pediatric Council of the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, *19*(12), 1413-1430. <https://doi.org/10.1016/j.echo.2006.09.001>
- Lang, R. M., Badano, L. P., Tsang, W., Adams, D. H., Agricola, E., Buck, T., Faletra, F. F., Franke, A., Hung, J., de Isla, L. P., Kamp, O., Kasprzak, J. D., Lancellotti, P., Marwick, T. H., McCulloch, M. L., Monaghan, M. J., Nihoyannopoulos, P., Pandian, N. G., Pellikka, P. A., & Pepi, M. (2012). EAE/ASE recommendations for image acquisition and display using three-dimensional echocardiography. *Journal of the American Society of Echocardiography*, *25*(1), 3-46. <https://doi.org/10.1016/j.echo.2011.11.010>
- Larson, D. B., Johnson, L. W., Schnell, B. M., Goske, M. J., Salisbury, S. R., & Forman, H. P. (2011). Rising use of CT in child visits to the emergency department in the United States, 1995-2008. *Radiology*, *259*(3), 793-801. <https://doi.org/10.1148/radiol.11101939>
- Li, F., Yang, G., & Zhu, Z.M. (2009). Investigation on technical principle and clinical application of dual-source CT. *Chinese Medical Equipment Journal*, *30*(2), 112-113. <https://doi.org/10.3969/j.issn.1003-8868.2009.02.050>
- Li, M., Gan, T., Shun, K., Song, X. R., & Wu, R. N. (2016). Diagnostic value for complex congenital heart disease: The comparative study of low doses of large pitch dual-source CT imaging and echocardiography. *Progress in Modern Biomedicine*, *16*(2), 229-308.
- Lian, J., Wang, X. Q., & Wang, P. F. (2018). Application value of transthoracic echocardiography and DSCT in preoperation of children with congenital heart disease. *Chinese Journal of CT and MRI*, *16*(1), 72-74. <https://doi.org/10.3969/j.issn.1672-5131.2018.01.023>
- Lin, E., & Alessio, A. (2009). What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT? *Journal of Cardiovascular Computed Tomography*, *3*(6), 403-408. <https://doi.org/10.1016/j.jcct.2009.07.003>
- Liu, N., & Wang, Q. Z. (2017). The diagnosis value of MSCT for complex congenital heart disease in children. *Chinese Journal of CT and MRI*, *15*(9), 47-50. <https://doi.org/10.3969/j.issn.1672-5131.2017.09.015>
- Lu, M. J., & Zhao, S. H. (2010). Magnetic resonance imaging in complex congenital heart disease. *Advances in Cardiovascular Disease*, *5*(3), 652-655.
- Lv, J. H., Lu, M. J., Zhao, S. H., Jiang, S. L., Zhang, G. J., Zhang, Y., & Jin, L. X. (2010). "One-stop shop" examination of complicated and complex congenital heart diseases with MRI. *Chinese Journal of Medical Imaging Technology*, *26*(10), 1864-1868.
- Marin, J. R., Zuckerbraun, N. S., & Kahn, J. M. (2012). Use of emergency ultrasound in United States pediatric emergency medicine fellowship programs in 2011. *Journal of Ultrasound in Medicine*, *31*(9), 1357-1363. <https://doi.org/10.7863/jum.2012.31.9.1357>
- Mazrani, W., McHugh, K., & Marsden, P. J. (2007). The radiation burden of radiological investigations. *Archives of Disease in Childhood*, *92*(12), 1127-1131. <https://doi.org/10.1136/adc.2006.101782>

- Miao, Y., Bing, J. W., & Bu, G. L. (2016). Application of prospectively electrocardiograph-gated 128-slice spiral CT angiography in children with complex congenital heart disease. *Chinese Journal of CT and MRI*, 14(10), 51-53.
- MOH. (2007). *Health indicators: Indicators for monitoring and evaluation of strategy health for all*. Ministry of Health. https://www.moh.gov.my/moh/images/gallery/publications/md/hi/hi_2007.pdf
- MOH. (2008a). *Health facts 2007*. Ministry of Health. http://vlib.moh.gov.my/cms/documentstorage/com.tms.cms.document.Document_2e89ce05-a0188549-d5315d00-5bfa49a4/2007.pdf
- MOH. (2008b). *Annual report 2007* (pp.75-76). Ministry of Health. <https://www.moh.gov.my/moh/images/gallery/publications/md/ar/2007-2.pdf>
- Nakagawa, M., Ozawa, Y., Nomura, N., Inukai, S., Tsubokura, S., Sakurai, K., Shimohira, M., Ogawa, M., & Shibamoto, Y. (2016). Utility of dual source CT with ECG-triggered high-pitch spiral acquisition (Flash Spiral Cardio mode) to evaluate morphological features of ventricles in children with complex congenital heart defects. *Japanese Journal of Radiology*, 34(4), 284-291. <https://doi.org/10.1007/s11604-016-0522-x>
- Nie, P., Yang, G., Wang, X., Duan, Y., Xu, W., Li, H., Cao, T., Liu, X., Ji, X., Cheng, Z., & Wang, A. (2014). Application of prospective ECG-gated high-pitch 128-slice dual-source CT angiography in the diagnosis of congenital extracardiac vascular anomalies in infants and children. *PLoS One*, 9(12), Article e115793. <https://doi.org/10.1371/journal.pone.0115793>
- Pache, G., Grohmann, J., Bulla, S., Arnold, R., Stiller, B., Schlensak, C., Langer, M., & Blanke, P. (2011). Prospective electrocardiography-triggered CT angiography of the great thoracic vessels in infants and toddlers with congenital heart disease: Feasibility and image quality. *European Journal of Radiology*, 80(3), e440-e445. <https://doi.org/10.1016/j.ejrad.2011.01.032>
- Pavithran, S., Natarajan, K., Vishwambaran, B., Arke, A. D., & Sivakumar, K. (2014). Preliminary evaluation of a microtransesophageal probe in neonates and young infants undergoing surgery for congenital heart disease. *Annals of Pediatric Cardiology*, 7(3), 173-179. <https://doi.org/10.4103/0974-2069.140829>
- Pearce, M. S., Salotti, J. A., Little, M. P., McHugh, K., Lee, C., Kim, K. P., Howe, N. L., Ronckers, C. M., Rajaraman, P., Craft, A. W., Parker, L., & de González DPhil, A. B. (2012). Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: A retrospective cohort study. *The Lancet*, 380(9840), 499-505. [https://doi.org/10.1016/S0140-6736\(12\)60815-0](https://doi.org/10.1016/S0140-6736(12)60815-0)
- Porter, T. R., Mulvagh, S. L., Abdelmoneim, S. S., Becher, H., Belcik, J. T., Bierig, M., Choy, J., Gaibazzi, N., Gillam, L. D., Janardhanan, R., Kutty, S., Leong-Poi, H., Linder, J. R., Main, M. L., Mathias Jr, W., Park, M. M., Senior, R., & Villanueva, F. (2018). Clinical applications of ultrasonic enhancing agents in echocardiography: 2018 American Society of echocardiography guidelines update. *Journal of the American Society of Echocardiography*, 31(3), 241-274. <https://doi.org/10.1016/j.echo.2017.11.013>
- Rehman, R., Yelamanchili, V. S., & Makaryus, A. N. (2021). *Cardiac imaging*. StatPearls Publishing LLC.
- Roberts, W. T., Bax, J. J., & Davies, L. C. (2008). Cardiac CT and CT coronary angiography: Technology and application. *Heart (British Cardiac Society)*, 94(6), 781-792. <https://doi.org/10.1136/hrt.2007.116392>
- Ropers, D., Rixe, J., Anders, K., Küttner, A., Baum, U., Bautz, W., Daniel, W. G., & Achenbach, S. (2006). Usefulness of multidetector row spiral computed tomography with 64- x 0.6-mm collimation and 330-ms

- rotation for the noninvasive detection of significant coronary artery stenoses. *The American Journal of Cardiology*, 97(3), 343-348. <https://doi.org/10.1016/j.amjcard.2005.08.050>
- Seibert, J. A. (2004). Tradeoffs between image quality and dose. *Pediatric Radiology*, 34(3), S183-S195. <https://doi.org/10.1007/s00247-004-1268-7>
- Shrimpton, P. C., Miller, H. C., Lewis, M. A., & Dunn, M. (2005). *Doses from computed tomography (CT) examinations in the UK-2003 review*. National Radiological Protection Board.
- Siegel, M. J. (2003). Multiplanar and three-dimensional multi-detector row CT of thoracic vessels and airways in the pediatric population. *Radiology*, 229(3), 641-650. <https://doi.org/10.1148/radiol.2293020999>
- Slovis, T. L. (2003). Children, computed tomography radiation dose, and the as low as reasonably achievable (ALARA) concept. *Pediatrics*, 112(4), 971-972. <https://doi.org/10.1542/peds.112.4.971>
- Sluysmans, T., & Colan, S. D. (2005). Theoretical and empirical derivation of cardiovascular allometric relationships in children. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, 99(2), 445-457. <https://doi.org/10.1152/jappphysiol.01144.2004>
- Somerville, J., & Grech, V. (2009). The chest X-ray in congenital heart disease 1. Total anomalous pulmonary venous drainage and coarctation of the aorta. *Images in Paediatric Cardiology*, 11(1), 7-9.
- Sreedhar, C. M., Ram, M. S., Alam, A., & Indrajit, I. K. (2005). Cardiac MRI in congenital heart disease - Our experience. *Medical Journal, Armed Forces India*, 61(1), 57-62. [https://doi.org/10.1016/S0377-1237\(05\)80122-4](https://doi.org/10.1016/S0377-1237(05)80122-4)
- Sreedher, G., Bruckman, D., & Ganapathy, S. S. (2021). 320 slice CT in imaging of congenital heart diseases in infants: A single-center experience. *Cureus*, 13(2), Article e13348. <https://doi.org/10.7759/cureus.13348>
- Sriharan, M., Lazoura, O., Pavitt, C. W., Castellano, I., Owens, C. M., Rubens, M. B., Padley, S. P., & Nicol, E. D. (2016). Evaluation of high-pitch ungated pediatric cardiovascular computed tomography for the assessment of cardiac structures in neonates. *Journal of Thoracic Imaging*, 31(3), 177-182. <https://doi.org/10.1097/RTI.0000000000000201>
- Stolzmann, P., Goetti, R., Baumueller, S., Plass, A., Falk, V., Scheffel, H., Feuchtner, G., Marincek, B., Alkadhi, H., & Leschka, S. (2011). Prospective and retrospective ECG-gating for CT coronary angiography perform similarly accurate at low heart rates. *European Journal of Radiology*, 79(1), 85-91. <https://doi.org/10.1016/j.ejrad.2009.12.016>
- Stolzmann, P., Leschka, S., Scheffel, H., Krauss, T., Desbiolles, L., Plass, A., Genoni, M., Flohr, T. G., Wildermuth, S., Marincek, B., & Alkadhi, H. (2008). Dual-source CT in step-and-shoot mode: Noninvasive coronary angiography with low radiation dose. *Radiology*, 249(1), 71-80. <https://doi.org/10.1148/radiol.2483072032>
- Tsai, I. C., & Goo, H. W. (2013). Cardiac CT and MRI for congenital heart disease in Asian countries: Recent trends in publication based on a scientific database. *The International Journal of Cardiovascular Imaging*, 29(1), 1-5. <https://doi.org/10.1007/s10554-013-0184-z>
- Ünal, E., Yıldız, A. E., Güler, E., Karcaaltıncaba, M., Akata, D., Kılınçer, A., Atlı, E., Topçuoğlu, M., & Hazirolan, T. (2015). Comparison of image quality and radiation dose between prospectively ECG-triggered and retrospectively ECG-gated CT angiography: Establishing heart rate cut-off values in first-

- generation dual-source CT. *Anatolian Journal of Cardiology*, 15(9), 759-764. <https://doi.org/10.5152/akd.2014.5720>
- UNSCEAR. (2006). *Effects on ionizing radiation: United Nations scientific committee on the effects of atomic radiation* (Vol. 1). United Nations Scientific Committee on the Effects of Atomic Radiation. https://www.unscear.org/docs/publications/2006/UNSCEAR_2006_Report_Vol.I.pdf
- UNSCEAR. (2013). *Effects of radiation exposure of children. Report to the general assembly with scientific annexes* (Vol. 1). United Nations Scientific Committee on the Effects of Atomic Radiation. https://www.unscear.org/unscear/uploads/documents/publications/UNSCEAR_2013_Annex-A-CORR.pdf
- Wang, R., Xu, X. J., Huang, G., Zhou, X., Zhang, W. W., Ma, Y. Q., & Zuo, X. N. (2017). Comparison of image quality, diagnostic accuracy and radiation dose between flash model and retrospective ECG-triggered protocols in dual source computed tomography (DSCT) in congenital heart diseases. *Polish Journal of Radiology*, 82, 114-119. <https://doi.org/10.12659/PJR.899876>
- Xu, L., Yang, L., Zhang, Z., Li, Y., Fan, Z., Ma, X., Lv, B., & Yu, W. (2010). Low-dose adaptive sequential scan for dual-source CT coronary angiography in patients with high heart rate: Comparison with retrospective ECG gating. *European Journal of Radiology*, 76(2), 183-187. <https://doi.org/10.1016/j.ejrad.2009.06.003>
- Yao, Q., Hu, X. H., Shen, Q. L., Qiao, Z. W., Pa, M. E., Qian, B., Yan, W. L., & Huang, G. Y. (2016). Differential effect of the ratio of right ventricular volume to left ventricular volume in children with repaired tetralogy of Fallot. *Cardiology*, 133(3), 135-140. <https://doi.org/10.1159/000441291>
- Zeng, Q. B., Guo, Q. N., Luo, Q., Wu, G. Y., & Zhou, X. (2014). Manganese-based contrast agents for MRI. *Chinese Journal of Magnetic Resonance Imaging*, 5(4), 315-320.
- Zhang, Y., Zou, C. J., & Xu, Y. Q. (2012). The clinical value of MSCT in anomalous pulmonary venous connection of neonate. *Journal of Medical Imaging*, 22(11), 1876-1878.
- Zimmerman, M. S., Smith, A. G. C., Sable, C. A., Echko, M. M., Wilner, L. B., Olsen, H. E., Atalay, H. T., Awasthi, A., Bhutta, Z. A., Boucher, J. L., Castro, F., Cortesi, P. A., Dubey, M., Fischer, F., Hamidi, S., Hay, S. I., Hoang, C. L., Hugo-Hamman, C., Jenkins, K. J., ... & Kassebaum, N. J. (2020). Global, regional, and national burden of congenital heart disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Child & Adolescent Health*, 4(3), 185-200. [https://doi.org/10.1016/s2352-4642\(19\)30402-X](https://doi.org/10.1016/s2352-4642(19)30402-X)

Evaluation of Phytochemical and Mineral Composition of Malaysia's Purple-Flesh Sweet Potato

Rosnah Shamsudin^{1,2*}, Nurfarhana Shaari¹, Mohd Zuhair Mohd Noor¹,
Nazatul Shima Azmi^{1,3} and Norhashila Hashim⁴

¹Department of Process and Food Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Institute of Advanced Technology, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

³Department of Mechanical Engineering Technology, Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, 86400 UTHM, Pagoh, Johor, Malaysia

⁴Department of Agricultural and Biology, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Sweet potato (*Ipomoea batatas* (L.) Lam) is one of Malaysia's underutilized main crops. However, systematic characterization of phytochemical and mineral contents of the sweet potato tuber and flour is still required for further specific food product development. Therefore, this study aims to evaluate the effects of the peeling conditions on the nutritional composition of sweet potato tuber and flour of the *Anggun 1* cultivar. The following properties were investigated in the different peeling conditions of the sweet potato tuber (unpeeled, peeled, and skin only), i.e., the phytochemical properties (total phenolic, total flavonoid, anthocyanin content) and mineral elements (calcium and iron). The results indicated a significant interaction between the peeling conditions of the sweet potato tuber and flour on all the examined properties ($p < 0.05$), total flavonoid content (2615.05 mg quercetin/100g and 3362.96 mg quercetinrties ($p < 0.05$). The unpeeled sweet potato tuber and flour had the highest value of total phenolic content (3142.98 mg GAE/100g and 4303.80 mg GAE/10/100g), anthocyanin (628.35 mg/100g and 960.30 mg/100g) and iron (5.45 mg/100g for flour only). Moreover, the result had indicated a higher calcium content (701.87 mg/100g) in the skin of the sweet potato. Therefore, this study provides a broad insight into the

ARTICLE INFO

Article history:

Received: 29 July 2021

Accepted: 16 February 2022

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.09>

E-mail addresses:

rosnahs@upm.edu.my (Rosnah Shamsudin)

nurfashaari2008@gmail.com (Nurfarhana Shaari)

zuhair@upm.edu.my (Mohd Zuhair Mohd Noor)

nazatul_shima89@yahoo.com (Nazatul Shima Azmi)

norhashila@upm.edu.my (Norhashila Hashim)

* Corresponding author

sweet potato's phytochemical properties and the mineral composition, which can be used for further specific food product development and processing.

Keywords: Anthocyanin, calcium, iron, purple sweet potato, total flavonoid, total phenolic

INTRODUCTION

The sweet potato is a tuber crop widely consumed in the human diet, eaten either fresh or processed in meals and beverages. Sweet potato has been ranked 7th among staple foods globally for its multiple nutritional benefits (Tan, 2015). Malaysia accounts for an increment from 26,582 in 2011 to 41,245 in 2017 tons of production, mainly from the states of Perak, Kelantan, and Terengganu (Yusoff et al., 2018). The Malaysian Agriculture and Development Institute (MARDI) has developed many famous varieties of sweet potatoes, such as Jalomas, Gendut, and VitAto. In 2017, MARDI had released three types of *Anggun* viz., *Anggun 1*, *Anggun 2* and *Anggun 3*. They have similar purple skin and flesh but differ in morphological characteristics (Shafienaz et al., 2019). However, *Anggun 1* (Figure 1) was recognized as the ideal one as it has the ideal shape that meets the customer demand. Moreover, it also has a high anthocyanin content, is capable of survival in low-fertility soil, and is tolerant to drought and disease, making this variety one step further than the other varieties of sweet potatoes (Shaari et al., 2021).

The sweet potato storage roots make important dietary contributions of dietary fibers, carbohydrates, minerals (calcium, magnesium, potassium, and zinc), and vitamins (C, B1, B2, E, and A). Often, minerals such as calcium (Ca) and iron (Fe) are deficient in human diets. In addition, in human diets, the recommended daily intake (RDI) of iron for males and females is 8 mg/day and 18 mg/day (Trumbo et al., 2001). The deficiency of iron is the sixth leading cause of illness and disease. Other than the mineral compounds, sweet potato also contains phytochemical properties (Alam et al., 2016), which commonly exist in the purple-flesh sweet potato phenolic and anthocyanin content, which has better radical scavenging activity than the red cabbage, grape skin, and purple corn (Rumbaoa et al., 2009). Phenolics are molecules that consist of at least one aromatic ring and one or more hydroxyl groups (Shen et al., 2009). These phenolic compounds are responsible for the antioxidant activities that scavenge the free radicals. Free radicals are unstable and highly reactive atoms or molecules with at least one or more unpaired electrons. Over-



Figure 1. The local sweet potato *Anggun 1* variety

production of free radicals can damage biomolecules (lipids, proteins, DNA), which can eventually cause many chronic diseases in humans, such as cancer, cardiovascular disease, diabetes, and atherosclerosis (Fang et al., 2002; Uttara et al., 2009). An antioxidant is a stable molecule that helps to reduce the damage by donating an electron to a charged free radical and neutralizing it (Lobo et al., 2010).

Various products can be obtained from sweet potatoes, such as cookies, biscuits, muffins, noodles, and pies, with a longer shelf-life and improved characteristics. The sweet potato tuber is transformed into good processed goods, for example, flour, which is less bulky and more stable. Flour can be utilized as a thickener in soup, gravy, fabricated snacks, and bakery products (Ahmed et al., 2010). A study by Van Hal (2000) has reported that the unpeeled flour is darker than the peeled flour due to the effect on the phenolic content. It was found that the phenolic content was high on the outer surface of the sweet potato and thus increased the action of the oxidase enzyme, which led to the browning effect. Based on the common method of processing through direct grinding without peeling off its surface will shorten the processing procedure, and more nutrients are retained. In the previous study by Shaari et al. (2021), it was observed that flour processed from unpeeled sweet potatoes reduced the loss of physical and chemical properties and enhanced the stability of the sweet potato with better nutritional value, such as carbohydrate and fiber.

In Malaysia, while other crops such as rice, cassava, and potato have obtained more attention among researchers, the sweet potato has very limited available literature resources. It includes the phytochemical and nutrients contents of the different parts of the local sweet potato (Shaari et al., 2020; Zulkifli et al., 2021). Some studies have been conducted on the nutrient and phytochemical composition of the new sweet potato VitAto and the different *Anggun* varieties (Yusoff et al., 2018; Shaari et al., 2020; Zulkifli et al., 2021), but the composition of the cultivated sweet potatoes on the different peeling conditions is yet to be explored. In this current study, the different peeling conditions (unpeeled, peeled, and skin only) of the *Anggun 1* sweet potato cultivar that is grown in Malaysia are analyzed in terms of its nutritional composition through the assessments of phytochemical properties (total phenolic, total flavonoid, and anthocyanin content) and minerals (Calcium and Iron). Therefore, this study will provide knowledge and guidance on the effects of the different peeling conditions on the phytochemical properties of the *Anggun 1* sweet potato, which will be a benefit in developing comprehensive processing of sweet potato in the future.

MATERIALS AND METHODS

Plant Material

The sample preparation was adapted from the method described by Shaari et al. (2020). The sweet potato cultivar (*Ipomoea batatas* Lam cv. *Anggun 1*) was grown and harvested on a local farm located at Semenyih, Selangor (Malaysia). The farmers harvested the sweet

potato after 3 to 4 months of plantation according to the maturation stage (Van de Fliert & Braun, 1999). About three hundred and fifty tubers (~20 kg) were selected and transported to the laboratory. The tubers were washed with tap water to remove the dirt and soil. After being air-dried at a room temperature of 27°C, the washed tubers were stored at 4°C for further analysis.

Preparation of Sweet Potato Raw Tubers

The sweet potato tubers were divided into three groups. In the first group (C1), unpeeled sweet potatoes were cut into slices of 5 mm in thickness (Amir et al., 2017) using a sharp stainless-steel knife (86902, John Craft, Malaysia). For the second group (C2), the sweet potatoes were peeled with a hand peeler (Peeler Stainless Steel 22, Barbarian, Thailand). For the third group (C3), the skins of the sweet potatoes were peeled about 1 mm in thickness from the tuber. All the prepared samples for phytochemical and nutritional factors were placed in an auto seal bag and kept in a desiccator to inhibit the addition or reduction of moisture.

Preparation of Sweet Potato Flour Sample

The sweet potato tubers were divided into three groups, as mentioned in Section 2.2. Then, all the samples were oven-dried at 60°C for 7–8 hours using a DO6836 oven (Mettler GmbH, Schwabach, Germany) until they reached a 6–7% moisture content (Ahmed et al., 2010). Next, the flour was obtained by grinding the dried samples using a laboratory-scale grinder (HR-20B-AEC, AEC Machinery, Malaysia) and sieved through a 300 µm sieve (AS200, Retsch, Germany). Finally, the flour samples were packaged in plastic bags, sealed, and stored in a chiller (LF817LD, ASec0®, Japan) at 4°C for further analysis.

Determination of Phytochemical Properties

Extraction of Sample. The extraction of the sample has followed the method of (Huang et al., 2006). Approximately 1 g of sweet potato tuber and flour samples (C1, C2, and C3) were treated with 15 ml of 80% methanol and centrifuged at 1600 × g for 15 min. The supernatant was collected while the remaining suspension was re-extracted using another 10 ml of 80% methanol, and the process was repeated. The collected supernatant was combined and filtered through Whatman No. 4 filter paper into a 25 ml volumetric flask. The extracts were diluted to the volume and were stored at 4°C until further analysis.

Total Phenolic Content (TPC). Determination of total phenolic content (TPC) was used to determine the amount of phenolic content in the samples. According to the Folin-Ciocalteu (FC) assay, the TPC was measured by Huang et al. (2006). First, 0.2 ml of an aliquot from the prior extract was mixed with 1.0 ml of Folin-Ciocalteu's reagent and 0.8

ml of 7.5% (w/v) sodium carbonate solution. Next, the mixture was homogenized using a vortex (EE0475051, Cole Palmer, United States) and allowed to stand for 30 min at room temperature. Then, the absorbance of the sample was measured versus a blank at 765 nm using a UV-Vis spectrophotometer (LT 291, Uvsar, India). Finally, Gallic acid was used as a standard solution, and the results were expressed as gallic acid equivalent (mg GAE g⁻¹).

Total Flavonoid Content. The total flavonoid content (TFC) was determined by following the spectrophotometric method based on the aluminum chloride (AlCl₃) complexation (Huang et al., 2006). Approximately 0.5 ml of an aliquot from the previous extract was reacted with 1.0 ml of 2% (w/v) methanolic aluminum chloride (AlCl₃·6H₂O) and was mixed using a vortex (EE0475051, Cole Palmer, United States). The mixture was kept for 10 min at a room temperature of 27°C. Then, the absorbance was taken using a spectrophotometer at 430 nm versus a blank. The TFC of the sample was calculated using a calibration curve of quercetin and expressed as mg quercetin g⁻¹.

Anthocyanin. The anthocyanin content was determined according to Huang et al. (2006). Approximately 0.5 g of sample was extracted and centrifuged at 1600 × g for 15 min with 10 ml of acidified methanol (1% hydrochloric acid (HCl)). Then, the supernatant was collected, and the suspension was extracted again using another 10 ml of acidified methanol. The process was repeated twice. Finally, the supernatant collected was combined in a 25 ml volumetric flask and diluted to volume. The absorbance was read at 530 nm. The anthocyanin content for the samples was determined according to Equation 1:

$$\text{Anthocyanin content (mg 100g}^{-1}\text{)} = A \times MW \times DF \times 100 / (\epsilon \times W) \quad (1)$$

Where;

A = absorbance

MW = molecular weight of cyanidin-3-glucoside chloride (C₂₁H₂₁ClO₁₁, 484.84 Da)

DF = Dilution factor

ε = Molar absorptivity (34,300)

W = sample weight (g)

Determination of Mineral Composition

Mineral (Calcium and Iron). Calcium (Ca) and Iron (Fe) were determined using inductively coupled plasma-atomic emission spectrometry (ICP-AES) (Sun et al., 2011). First, 0.25 g of flour samples (C1, C2, and C3) were weighed into a polytetrafluoroethylene digestion tube, and 8 mL 65% (v/v) HNO₃ was added for pre-digestion. After 1 h, 30% (v/v) H₂O₂ was added and digested using a microwave digestion system (MARS 5, CEM Co., Matthews, NC, USA). Then, the digested solution was diluted with 100 mL Milli-Q

water (Bedford, MA, USA) and stored at 4°C in plastic tubes for further analysis. The mineral elements were measured with the following operating conditions: radio frequency power, 1280 W; nebulization chamber temperature, 20°C; sampling depth, 8 mm; cooling rate, 1.47 L min⁻¹; carrier gas flow rate, 1 L min⁻¹ and auxiliary gas flow rate, 1 L min⁻¹.

Statistical Analysis

All the analyses were done in triplicate, and the data were analyzed as mean values and standard deviation (SD) using the one-way analysis of variance (ANOVA) of IBM Statistic 22.0. According to Tukey HSD (Honestly Significant Difference), significant differences were established and determined at $p \leq 0.05$. The correlation coefficient was performed using the Statistical Package for Social Sciences (SPSS Inc. Waker Drive, Chicago, IL, USA).

RESULTS AND DISCUSSION

Determination of Phytochemical Properties

Total Phenolic Content (TPC). Table 1 shows the experimental data of the total phenolic content of the sweet potato tuber and flour samples at three different conditions (C1, C2, and C3). For the tuber sample, TPC values of C1, C2, and C3 were 3142.98±27.39, 1916.37±61.46, and 1189.77±55.88 mg GAE 100g⁻¹, respectively. The highest TPC was found in C1 (3142.98±27.39 mg 100g⁻¹), and it was greater than the reported data by (Akyol et al., 2016), where they found the TPC of potato tubers ranged from 0.5 to 0.6 mg 100g⁻¹ fresh matter. By comparing the TFC value of C2 and C3, C1 is the highest due to the accumulation of phenolic compounds in the flesh of the sweet potato (Shaari et al., 2020). After peeling the skin, the sweet potato tuber showed a significant TPC reduction

Table 1
Effect of peeling conditions on the total phenolic, total flavonoid, and anthocyanin content of *Anggun 1*

Sample	Total phenolic content (TPC) (mg GAE 100g ⁻¹)		Total flavonoid content (TFC) (mg quercetin 100g ⁻¹)		Anthocyanin (mg 100 ⁻¹)	
	Tuber	Flour	Tuber	Flour	Tuber	Flour
C1	3142.98 ± 27.39 ^{ab}	4303.80 ± 35.45 ^{aA}	2615.05 ± 29.34 ^{ab}	3362.96 ± 34.07 ^{aA}	628.35 ± 46.38 ^{ab}	960.30 ± 6.75 ^{aA}
C2	1916.37 ± 61.46 ^{bb}	4166.37 ± 30.81 ^{bA}	2480.05 ± 39.48 ^{ab}	2957.95 ± 63.75 ^{bA}	541.46 ± 24.75 ^{ab}	942.88 ± 7.54 ^{bA}
C3	1189.77 ± 55.88 ^{cb}	2890.06 ± 74.22 ^{cA}	2290.08 ± 81.92 ^{bb}	2604.30 ± 66.96 ^{cA}	333.48 ± 42.38 ^{bb}	521.08 ± 2.26 ^{cA}

± SD – standard deviation; C1 – Whole tuber; C2 – Peeled tuber; C3 – Skin of tuber. Different small letters within the same column indicate statistical difference ($p < 0.05$) among the processing conditions (C1, C2, C3), while different capital letters (A, B) within the same row indicate statistical difference ($p < 0.05$) among the different sweet potato form (raw, flour) on the Tukey's HSD comparison test.

($p < 0.05$) by 39.03% in C2. It might be due to removing the skin, where the skin (C3) had contributed a huge portion of TPC in C1 by 37.85%. It agreed with Mondy and Gosselin (1998), who stated that the skin of the sweet potato tuber is known to be high in phenolic content.

Similarly, the TPC of the sweet potato flour varied significantly ($p < 0.05$) under the different conditions. C1 had the highest TPC (4303.80 ± 35.45 mg GAE 100g^{-1}), whereas C3 (2890.06 ± 74.22 mg GAE 100g^{-1}) had the lowest. After peeling, the sweet potato flour showed a slight TPC change ($p < 0.05$) in C2, about 3.19%. This finding was contradicted by Yin et al. (2016), who had found a greater value of TPC in the skin (648.65 mg 100g^{-1}) rather than in the flesh (89.12 mg 100g^{-1}) of potato flour. In addition, the flour samples contained a higher TPC compared to the tuber samples. This finding agreed with the study by Ruttarattanamongkol et al. (2016), who found that the TPC flour sample of the purple-fleshed sweet potato (Phicit 65-3) was higher than the raw potato sample. The heat treatment during the drying process could cause damage to the cell structures of the sweet potato tissues and result in more easy extraction of antioxidant properties such as phenolic, flavonoid, anthocyanin, and carotenoid (Huang et al., 2006; Tokusoglu & Yildirim, 2012).

A significant difference ($p < 0.05$) was found between the tuber and flour samples in each condition. The sweet potato flour's TPC of C1, C2, and C3 increased to 4303.80 ± 35.45 , 4166.37 ± 30.81 , and 2890.06 mg GAE 100g^{-1} were 26.97, 54.00, and 58.83% higher than those that were in tuber conditions, respectively. These indicated that the sweet potato flour had retained more TPC than the tuber. It might be due to the drying process experienced by the sweet potato flour, where the thermal treatment had released more bound phenolic from the breakdown of cellular structure (Yang et al., 2010). Furthermore, this study has provided further support for the hypothesis that the skin contributes a high TPC to the sweet potato. According to Friedman (1997), phenolic compounds usually accumulate in the peel of potato tubers than in the flesh. Almost 50% of phenolic were found in the peel and adjoining tissues of the tubers, and their concentration decreased towards the center of the tuber (Friedman, 1997; Yang et al., 2010). Hence, to acquire more TPC from the sweet potato, the suggestion is to maximize the usage of the skin rather than discard it.

Total Flavonoid Content (TFC). The experimental data of total flavonoid content (TFC) for the different conditions (C1, C2, and C3) of the sweet potato tuber and flour samples are summarized in Table 1. The TFC of the sweet potato tuber sample in the different conditions of C1, C2, and C3 were 2615.05 ± 29.43 , 2480.05 ± 39.48 , and 2290.08 ± 81.92 mg quercetin 100g^{-1} , respectively. The TFC found in this current *Anggun* study variety with purple-fleshed was higher than the fresh white-fleshed potato (30 mg 100g^{-1}) (Perla, Holm, & Jayanty, 2012). The result showed that C1 had the highest TFC while C3 had the lowest TFC ($p < 0.05$). There was no significant difference observed between C1 and C2. It means that peeling does not affect the reduction of TFC in sweet potato tuber.

For sweet potato flour, the measured values of TFC in C1, C2, and C3 were 3362.96 ± 34.07 , 2957.95 ± 63.75 , and 2604.30 ± 66.96 mg quercetin 100g^{-1} , respectively. The TFC of sweet potato flour varied significantly ($p < 0.05$) under different conditions. Similarly, it is shown that the highest TFC value was reported in C1 while the lowest value resulted in C3. After the peeling process, the TFC in C2 reduced significantly ($p < 0.05$) from C1 to around 12.01%. This result is in agreement with the finding by Khajehei, Merkt, Claupein, and Graeff-Hoenninger (2018), where they have found that the TFC in New Zealand yacon tubers has decreased significantly ($p < 0.05$) from 3221.47 to 1041.689 mg RE 100g^{-1} dry weight after the skin removal.

After the sweet potato tuber was processed into flour, a significant difference ($p < 0.05$) was obtained between the tuber and flour samples. The TFC of C1, C2, and C3 of the flour sample was significantly higher ($p < 0.05$) than those tuber samples by 22.24, 16.16, and 12.07%, respectively. In addition, this study has shown that the skin of the sweet potato tuber and flour samples have a large contribution of TFC to the unpeeled sweet potato. It agrees with the result of the yacon tuber skin that contains high amounts of TFC (Khajehei et al., 2018). TFC represents the most common group of plant phenolic compounds, and their existence will influence the flavor and color of fruits and vegetables (Akyol et al., 2016). Furthermore, major bioactive compounds promote many potential health benefits, which have been used against chronic diseases, such as antiviral, cancer, inflammation, cardiovascular and neurodegenerative disorders. Therefore, from a quality and nutritional perspective, the suggestion for sweet potato consumption, including its skin, could maximize the intake of TFC in the human body.

Anthocyanin. As illustrated in Table 1, the anthocyanin content for the different conditions (C1, C2, and C3) of the sweet potato tuber and flour samples were observed. Based on the results obtained from the tuber sample, the anthocyanin content of C1, C2, and C3 were 628.35 ± 46.38 , 541.46 ± 24.75 , and 333.48 ± 42.38 mg 100g^{-1} , respectively. All the conditions in this current study showed a higher anthocyanin content than the potato studied by Yin et al. (2016). A significant ($p < 0.05$) highest value of anthocyanin content was exhibited in C1 and C2, while C3 showed the lowest value. There was no significant difference found between C1 and C2. This result was supported by the finding of Steed and Truong (2008), who had also found the insignificant value of anthocyanin content between unpeeled and peeled sweet potatoes. Therefore, it was suggested that peeling did not affect the reduction of the anthocyanin content. Furthermore, this current study found that the average anthocyanin content in the peeled sweet potato (C2) is higher than the skin (C3) by 1.6 times. In contrast to this finding, Steed and Truong (2008) and Albishi et al. (2013) have found more anthocyanin content in the skin than in peeled potatoes. This difference may result from multiple factors such as variety, several cultivars investigated, the thickness of skin peeled, and the methods of anthocyanin extraction.

Meanwhile, the anthocyanin content for C1, C2, and C3 in the sweet potato flour ranged from 521.08 to 960.30 mg/100g. On average, C1 had the highest anthocyanin (960.30 ± 6.75 mg/100g), whereas C3 had the lowest anthocyanin (521.08 ± 2.26 mg 100g⁻¹). The results showed a significant difference ($p < 0.05$) between each condition. After peeling, there was a slight change in the anthocyanin content of C2, about 1.81% ($p < 0.05$). Contrary to the result of the sweet potato tuber, peeling affects the anthocyanin content of the sweet potato flour. In addition, the mean anthocyanin content in the C2 was higher than in C3 by 44.74% ($p < 0.05$).

The anthocyanin content of the sweet potato tuber was significantly increased ($p < 0.05$) after being processed into flour. Higher anthocyanin content in the sweet potato flour rather than in the tuber was expected due to the inactivated enzymes that degraded the anthocyanin pigments and other antioxidant compounds (Aziz et al., 2018). Sweet potato is a good source of anthocyanin, and the level of this compound is increased when it is processed in a different form. Moreover, the evidence that has been presented in this section provides that the unpeeled sweet potato has reported higher rates of anthocyanin content compared to other conditions. This high value is contributed by the skin of the sweet potato, which is rich in anthocyanin content. Anthocyanin is often associated with preventive health effects and reduces risks of age-related macular degeneration (Jang et al., 2005), anti-carcinogenic activity, and cardiovascular disorders (Mazza, 2007). Therefore, concerning this aspect, it is suggested that the sweet potato is consumed unpeeled rather than peeled to maximize the anthocyanin intake.

Determination of Mineral Composition

Calcium. The variation of calcium for different conditions in the sweet potato flour is shown in Table 2. The present study's finding indicates that C1, C2, and C3 have exhibited a calcium content of 242.33 ± 6.20 , 216.98 ± 10.93 , and 701.87 ± 2.61 mg 100g⁻¹, respectively. The calcium content of C2 showed a higher value than Boni et al. (2018), which reported the calcium content ranged from 30 to 47 mg 100g⁻¹ for peeled sweet potato cultivars of *Kabode*, *Irenei*, *Fatoni*, *Tib*, and *BelaBela*. The change in calcium content in the sweet potatoes may be attributed to the differences in cultivars, climate, soil type, geographical location, and several other factors (Boni et al., 2018). Furthermore, C3 has significantly shown the highest ($p < 0.05$) calcium content of other conditions. The greatest difference was noted, in which the calcium content was as much as three times higher in the skin (C3) than in the flesh (C2). Similar observations have been made by Czech et al. (2020), who have also found calcium content in all citrus fruits to be more than 50% higher ($p < 0.05$) in the peel than in the pulp. Therefore, removing the sweet potato skin greatly reduces its nutritional value and allows this element to enter the environment. A finding has shown that the human digestive tract absorbs calcium of vegetable origin; hence, its loss should be limited by including the sweet potato skin during consumption (Yang et al., 2012).

Table 2
Effect of peeling condition on the calcium and iron content of *Anggun 1*

Sample	Mineral (Calcium) (mg 100g ⁻¹)	Mineral (Iron) (mg 100g ⁻¹)
	Flour	Flour
C1	242.33 ± 6.20 ^b	5.45 ± 3.04 ^b
C2	216.98 ± 10.93 ^a	5.26 ± 1.30 ^{ab}
C3	701.87 ± 2.61 ^c	4.84 ± 0.93 ^{ab}

± SD – standard deviation; C1 – Whole tuber; C2 – Peeled tuber; C3 – Skin of tuber. Results involving flour only. Different small letters within the same column indicate statistical difference ($p < 0.05$) among the processing conditions (C1, C2, C3), while different capital letters (a,b) within the same row indicate statistical difference ($p < 0.05$) among the different sweet potato form (raw, flour) on the Tukey's HSD comparison test.

Iron. Table 2 presents the effects of the conditions (C1, C2, C3) on the iron content of sweet potato flour. The means of iron contents for C1, C2, and C3 were obtained as follows, 5.45±3.04, 5.26±1.30, and 4.84±0.93 mg 100g⁻¹, respectively. However, no significant difference ($p > 0.05$) was found between the conditions. Ju et al. (2017) found a lower iron content in the sweet potato, ranging from 0.782 to 1.818 mg 100g⁻¹. On the other hand, Dako et al. (2016) and Senanayake et al. (2013) have found a higher iron content in the sweet potato varieties from Sri Lanka (4.2 to 6.3 mg 100g⁻¹) and Ethiopia (8.7 to 11.45 mg 100g⁻¹). Iron is responsible for many metabolic reactions. It involves a component of heme proteins, such as hemoglobin and myoglobin, in the human body. In addition, it is also a part of enzymes, such as cytochromes, catalases, and peroxidases (Antoine et al., 2012). Hence, C1, C2, and C3 cannot be classified as 'rich in' or a 'source of' iron due to the lower iron content. As suggested, the intake of sweet potatoes should be complemented with other food components to achieve the desired RDI.

CONCLUSION

This study determines the effects of peeling on the phytochemical properties and mineral elements of the most important sweet potato cultivar (cv. *Anggun 1*) harvested in the western region of Malaysia. The result showed that the peeling conditions had significant effects ($p < 0.05$) on some quality attributes, including the phytochemical and mineral content. On the other hand, the unpeeled sweet potato tuber and flour had a better nutritional composition by assessing phytochemical properties and mineral elements compared to other peeling conditions. Therefore, it is proven that the unpeeled sweet potato has a more significant condition in improving the quality of the sweet potato and could contribute to more potential applications in the food industry. Therefore, to maximize sweet potato usage and minimize its waste, it is suggested that the food processing industry uses unpeeled sweet potatoes.

ACKNOWLEDGMENTS

The authors express their gratitude to the Universiti Putra Malaysia for providing financial and technical support under grant GP-IPB/2018/9660301 to conduct this research work.

REFERENCES

- Ahmed, M., Akter, M. S., & Eun, J. B. (2010). Peeling, drying temperatures, and sulphite - Treatment affect physicochemical properties and nutritional quality of sweet potato flour. *Food Chemistry*, *121*(1), 112-118. <https://doi.org/10.1016/j.foodchem.2009.12.015>
- Akyol, H., Riciputi, Y., Capanoglu, E., Caboni, M. F., & Verardo, V. (2016). Phenolic compounds in the potato and its byproducts: An overview. *International Journal of Molecular Sciences*, *17*(6), Article 835. <https://doi.org/10.3390/ijms17060835>
- Alam, M. K, Rana, Z. H., & Islam, S. N. (2016). Comparison of the proximate composition, total carotenoids and total polyphenol content of nine orange-fleshed sweet potato varieties grown in Bangladesh. *Foods*, *5*(4), Article 64. <https://doi.org/10.3390/foods5030064>
- Albishi, T., John, J. A., Al-Khalifa, A. S., & Shahidi, F. (2013). Phenolic content and antioxidant activities of selected potato varieties and their processing by-products. *Journal of Functional Foods*, *5*(2), 590-600. <https://doi.org/10.1016/j.jff.2012.11.019>
- Amir, H. M. S., Wong, M. M., Adibah, Y., Madihah, M. S., Nida, I., Rashidah, W. A. K. W., Rozira, A., Sharifudin, M. S., Mariam, F. M. N., & Mona, Z. (2017). Biochemical content of two sweet potatoes (*Ipomoea batatas*) cultivars under dry heat and moist heat treatment. *Transactions on Science and Technology*, *4*(3), 224-231.
- Antoine, J. M. R., Fung, L. A. H., Grant, C. N., Dennis, H. T., & Lalor, G. C. (2012). Dietary intake of minerals and trace elements in rice on the Jamaican market. *Journal of Food Composition and Analysis*, *26*(1-2), 111-121. <https://doi.org/10.1016/j.jfca.2012.01.003>
- Aziz, A. A., Padzil, A. M., & Muhamad, I. I. (2018). Effect of incorporating purple-fleshed sweet potato in biscuit on antioxidant content, antioxidant capacity and colour characteristics. *Malaysian Journal of Analytical Sciences*, *22*(4), 667-675. <https://doi.org/10.17576/mjas-2018-2204-13>
- Boni, V. W. T., Digbeu, Y. D., Dje, M. K., Kouamé, L. P., & Dué, E. A. (2018). Proximate and mineral composition and some anti-nutritional factors of flour from five improved sweet potato (*Ipomoea batatas* (L.) Lam) varieties roots. *European Journal of Pharmaceutical and Medical Research*, *5*(11), 126-137.
- Czech, A., Zarycka, E., Yanovych, D., Zasadna, Z., Grzegorzczuk, I., & Klys, S. (2020). Mineral content of the pulp and peel of various citrus fruit cultivars. *Biological Trace Element Research*, *193*, 555-563. <https://doi.org/10.1007/S12011-019-01727-1>
- Dako, E., Retta, N., & Desse, G. (2016). Comparison of three sweet potato (*Ipomoea batatas* (L.) Lam) varieties on nutritional and anti-nutritional factors. *Global Journal of Science Frontier Research*, *16*(4), 63-72.
- Fang, Y. Z., Yang, S. & Wu, G. (2002). Free radicals, antioxidants, and nutrition. *Nutrition*, *18*(10), 872-879. [https://doi.org/10.1016/S0899-9007\(02\)00916-4](https://doi.org/10.1016/S0899-9007(02)00916-4)

- Friedman, M. (1997). Chemistry, biochemistry, and dietary role of potato polyphenols. A review. *Journal of Agricultural and Food Chemistry*, 45(5), 1523-1540. <https://doi.org/10.1021/Jf960900s>
- Huang, Y. C., Chang, Y. H., & Shao, Y. Y. (2006). Effects of genotype and treatment on the antioxidant activity of sweet potato in Taiwan. *Food Chemistry*, 98(3), 529-538. <https://doi.org/10.1016/j.foodchem.2005.05.083>
- Jang, Y. P., Zhou, J., Nakanishi, K., & Sparrow, J. R. (2005). Anthocyanins protect against A2e photooxidation and membrane permeabilization in retinal pigment epithelial cells. *Photochemistry and Photobiology*, 81(3), 529-536. <https://doi.org/10.1111/j.1751-1097.2005.tb00221.x>
- Ju, D., Mu, T. H., & Sun, H. N. (2017). Sweet potato and potato residual flours as potential nutritional and healthy food material. *Journal of Integrative Agriculture*, 16(11), 2632-2645. [https://doi.org/10.1016/S2095-3119\(16\)61601-5](https://doi.org/10.1016/S2095-3119(16)61601-5)
- Khajehci, F., Merkt, N., Claupein, W., & Graeff-Hoeningner, S. (2018). Yacon (*Smallanthus sonchifolius* Poepp. & Endl.) as a novel source of health promoting compounds: Antioxidant activity, phytochemicals and sugar content in flesh, peel, and whole tubers of seven cultivars. *Molecules*, 23(2), Article 278. <https://doi.org/10.3390/Molecules23020278>
- Lobo, V., Patil, A., Phatak, A., & Chandra, N. (2010) Free radicals, antioxidants and functional foods: Impact on human health. *Pharmacognosy Reviews*, 4(8), 118-126. <https://doi.org/10.4103/0973-7847.70902>
- Mazza, G. (2007). Anthocyanins and heart health. *Annali-Istituto Superiore Di Sanita*, 43(4), 369-374.
- Mondy, N. I., & Gosselin, B. (1988). Effect of peeling on total phenols, total glycoalkaloids, discoloration and flavor of cooked potatoes. *Journal of Food Science*, 53(3), 756-759. <https://doi.org/10.1111/J.1365-2621.1988.Tb08949.x>
- Perla, V., Holm, D. G., & Jayanty, S. S. (2012). Effects of cooking methods on polyphenols, pigments and antioxidant activity in potato tubers. *LWT - Food Science and Technology*, 45(2), 161-171. <https://doi.org/10.1016/J.Lwt.2011.08.005>
- Rumbaoa, R. G. O., Cornago, D. F., & Geronimo, I. M. (2009). Phenolic content and antioxidant capacity of Philippine sweet potato (*Ipomoea batatas*) varieties, 113(4), 1133-1138. <https://doi.org/10.1016/j.foodchem.2008.08.088>
- Ruttarattanamongkol, K., Chittrakorn, S., Weerawatanakorn, M., & Dangpium, N. (2016). Effect of drying conditions on properties, pigments and antioxidant activity retentions of pretreated orange and purple-fleshed sweet potato flours. *Journal of Food Science and Technology*, 53(4), 1811-1822. <https://doi.org/10.1007/S13197-015-2086-7>
- Senanayake, S. A., Ranaweera, K. K. D. S., Gunaratne, A., & Bamunuarachchi, A. (2013). Comparative analysis of nutritional quality of five different cultivars of sweet potatoes (*Ipomoea batatas* (L) Lam) in Sri Lanka. *Food Science & Nutrition*, 1(4), 284-291. <https://doi.org/10.1002/fsn3.38>
- Shaari, N., Shamsudin, R., Mohd Nor, M. Z., & Hashim, N. (2020). Phenolic, flavonoid and anthocyanin contents of local sweet potato (*Ipomoea batatas*). *Food Research*, 4, 74-77. [https://doi.org/10.26656/fr.2017.4\(s1\).s01](https://doi.org/10.26656/fr.2017.4(s1).s01)
- Shaari, N., Shamsudin, R., Nor, M. Z. M., & Hashim, N. (2021). Quality attributes of Malaysia purple-fleshed sweet potato at different condition. *Agronomy*, 11(5), 872-886. <https://doi.org/10.3390/agronomy11050872>

- Shafienaz, N. A., Najib, M., Omar, H., & Jupri, A. (2019). Tuber yields of three MARDI purple flesh sweet potatoes (*Ipomoea batatas* [L.] Lam) varieties as affected by different age of planting material. In *28th Malaysian Society of Plant Physiology Conference (MSPPC 2018), Challenges and strategies for plant productivity and resilience* (pp. 255-257). Malaysian Society of Plant Physiology (MSPP)
- Shen, Y., Jin, L., Xiao, P., Lu, Y., & Bao, J. (2009). Total phenolics, flavonoids, antioxidant capacity in rice grain and their relations to grain color, size and weight. *Journal of Cereal Science*, *49*(1), 106-111. <https://doi.org/10.1016/j.jcs.2008.07.010>
- Steed, L. E., & Truong, V. D. (2008). Anthocyanin content, antioxidant activity, and selected physical properties of flowable purple-fleshed sweet potato purees. *Journal of Food Science*, *73*(5), S215-S221. <https://doi.org/10.1111/j.1750-3841.2008.00774.x>
- Sun, S., Guo, B., Wei, Y., & Fan, M. (2011). Multi-Element analysis for determining the geographical origin of mutton from different regions of China. *Food Chemistry*, *124*(3), 1151-1156. <https://doi.org/10.1016/j.foodchem.2010.07.027>
- Tan, S. L. (2015). Sweetpotato - *Ipomoea batatas* - A great health food. *UTAR Agriculture Science Journal*, *1*(3), 15-28.
- Tokusoglu, O., & Yildirim, Z. (2012). Effects of cooking methods on the anthocyanin levels and antioxidant activity of a local Turkish sweetpotato [*Ipomoea batatas* (L.) Lam] cultivar Hatay Kirmizi: Boiling. *Turkish Journal of Field Crops*, *17*(1), 87-90.
- Trumbo, P., Yates, A. A., Schlicker, S., & Poos, M. (2001). Dietary reference intakes: Vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. *Journal of the American Dietetic Association*, *101*(3), 294-301. [https://doi.org/10.1016/S0002-8223\(01\)00078-5](https://doi.org/10.1016/S0002-8223(01)00078-5)
- Uttara, B., Singh, A. V., Zamboni, P., & Mahajan, R. T. (2009). Oxidative stress and neurodegenerative diseases: A review of upstream and downstream antioxidant therapeutic options. *Current Neuropharmacology*, *7*(1), 65-74.
- Van de Fliert, E., & Braun, A. R. (1999). *Farmer field school for integrated crop management of sweet potato. Field Guides Technical Manual*. International Potato Center.
- Van Hal, M. (2000). Quality of sweetpotato flour during processing and storage. *Food Reviews International*, *16*(1), 1-37. <https://doi.org/10.1081/fri-100100280>
- Yang, J., Chen, J. F., Zhao, Y. Y., & Mao, L. C. (2010). Effects of drying processes on the antioxidant properties in sweet potatoes. *Agricultural Sciences in China*, *9*(10), 1522-1529. [https://doi.org/10.1016/S1671-2927\(09\)60246-7](https://doi.org/10.1016/S1671-2927(09)60246-7)
- Yang, J., Punshon, T., Guerinot, M. L., & Hirschi, K. D. (2012). Plant calcium content: Ready to remodel. *Nutrients*, *4*(8), 1120-1136. <https://doi.org/10.3390/nu4081120>
- Yin, L., Chen, T., Li, Y., Fu, S., Li, L., Xu, M., & Niu, Y. (2016). A comparative study on total anthocyanin content, composition of anthocyanidin, total phenolic content and antioxidant activity of pigmented potato peel and flesh. *Food Science and Technology Research*, *22*(2), 219-226. <https://doi.org/10.3136/fstr.22.219>

- Yusoff, M. M., Abdullah, S. N., Halim, M. R. A., Shari, E. S., Ismail, N. A., & Yusoff, M. M. (2018). Growth and yield performance of five purple sweet potato (*Ipomoea batatas*) accessions on colluvium soil. *Pertanika Journal of Tropical Agricultural Science*, 41(3), 975-986.
- Zulkifli, N. A., Nor, M. Z. M., Omar, F. N., Sulaiman, A., & Mokhtar, M. N. (2021). Proximate composition of Malaysian local sweet potatoes. *Food Research*. 5(Suppl 1), 73-79. [https://doi/10.26656/fr.2017.5\(S1\).045](https://doi/10.26656/fr.2017.5(S1).045)

Cultivation of *Rhodotorula Toruloides* Using Palm Oil Mill Effluent: Effect on the Growth, Lipid Production, and Waste Removal

Ignatia Justine, Grace Joy Wei Lie Chin, Wilson Thau Lym Yong and Mailin Misson*

Biotechnology Research Institute, University Malaysia Sabah, 88400 UMS, Kota Kinabalu, Sabah, Malaysia

ABSTRACT

Using Palm Oil Mill Effluent (POME) as a medium for culturing oleaginous yeast is advantageous for simultaneous lipid production and waste removal. The organic compounds in POME can be utilised as a nutrient source for yeast growth. *Rhodotorula toruloides* yeast was cultivated in filtered and unfiltered raw POME as growth media in this study. The yeast growth, pH changes in media, lipid production and removal of chemical oxygen demand (COD) of *Rhodotorula toruloides* cultivated in POME were examined and compared to *Rhodotorula toruloides* grown in yeast peptone dextrose (YPD) control media. The COD level of filtered POME was reduced by nearly 50% after filtration. The biomass concentration of *Rhodotorula toruloides* in filtered POME surpassed the other media in the following order: filtered POME > YPD > unfiltered POME (152 mg/ml > 121 mg/ml > 37 mg/ml). The filtered POME was found favourable for yeast growth due to the minimal amount of colloidal particles and suspended solids. Meanwhile, the lipid production (4.51 %) in filtered POME was 4.8-fold higher than in control media. The water analysis indicated about 43% of COD reduction, signifying the ability of *Rhodotorula toruloides* to utilise nutrient components present in POME for growth. This study provides insightful knowledge on the utilisation of oleaginous yeast for simultaneous green waste disposal and sustainable microbial oil production.

Keywords: Chemical oxygen demand, lipid production, palm oil mill effluent, *Rhodotorula toruloides*, waste treatment

ARTICLE INFO

Article history:

Received: 29 August 2021

Accepted: 28 December 2021

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.10>

E-mail addresses:

ignatia94.j@gmail.com (Ignatia Justine)

mailin@ums.edu.my (Mailin Misson)

gracejoy@ums.edu.my (Grace Joy Wei Lie Chin)

wilsonyg@ums.edu.my (Wilson Thau Lym Yong)

* Corresponding author

INTRODUCTION

The palm oil industry is an important commodity and the largest agriculture-based industry in Malaysia (Chin et al., 2013). Compared to other countries, Malaysia is

rich in tropical forests and agricultural vegetation, which has led to the industry's success, which sought the growth of palm trees (Foo & Hameed, 2010). The increased demand for palm oil has contributed to palm oil production worldwide. Palm oil production is anticipated to increase from 72.27 million metric tonnes in 2019/2020 to 73.23 million metric tonnes in 2020/2021 (Shahbandeh, 2021). Palm oil industries contributed about 5% to 7% of Malaysia's gross domestic product (GDP) (Lee et al., 2020). The rapid production of palm oil, however, generates an abundance of wastes such as empty fruit bunches, mesocarp fibre, palm kernel shells, oil palm frond, oil palm trunk, and palm oil mill effluents (POME) (Lee et al., 2020).

Concerns about the detrimental environmental effect of the massive amount of palm oil-based wastes have been raised. Raw POME contains high chemical oxygen demand (16,000–100,000 mg/L), biochemical oxygen demand (10,250–43,750 mg/L), total suspended solids (5,000–54,000 mg/L), oils and greases, and ammoniacal nitrogen with highly acidic pH determined around 3 to 4 (Abdulsalam et al., 2018). The presence of these contaminants makes the disposal of POME a pollution issue. Therefore, further research into developing alternative approaches for POME treatment becomes crucial in combating the environmental impact.

According to Amat et al. (2015), there are few known conventional methods for treating POME, including evaporation, adsorption, and advanced oxidation (ozonation technique), which primarily involves decolourising the POME from its thick brownish colour into the colour of pure, clean water. Out of the many alternative methods explored for POME treatment, previous findings showed that aerobic and anaerobic biological processes have significantly catered to energy efficiency, low nutrient intake, and biomass yield (Cheng et al., 2010).

More than 85% of POME producers in Malaysia employ ponding systems with anaerobic technology for POME treatment due to the low cost of operation and capital (Abdurahman et al., 2011). Nonetheless, this approach has surged issues threatening the environment due to extensive land usage and long retention time (Abdurahman et al., 2011). Therefore, it indicates that further approaches for a green and cost-effective POME treatment should be explored.

The employment of microbial oils on POME as cheap media is beneficial for green waste management and essential for the cost-effective production of valuable products (Marjakangas et al., 2015). Oleaginous microorganisms, such as yeast, microalgae, fungi, and bacteria, can produce more than 20–25% microbial lipids or single-cell oils like vegetable oils composition (Liang & Jiang, 2013). The application of oleaginous microorganisms in treating POME has gained widespread acceptance as a potential replacement for vegetable oils and future biodiesel. Culturing oleaginous microorganisms is more practical since it is unaffected by seasons or climate change (Cheirsilp & Louhasakul, 2013).

Oleaginous yeasts are attractive due to their ability to use low-cost fermentation media, rapid growth, and feasibility for large-scale cultivation (Cheirsilp & Louhasakul, 2013). *Yarrowia sp.*, *Candida sp.*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus sp.*, *Trichosporon sp.*, and *Lipomyces sp.* are lipid-producing yeasts (Marjakangas et al., 2015). *Rhodospiridium sp.*, *Rhodotorula sp.*, and *Lipomyces sp.* can accumulate intracellular lipids up to 70% of their dry cell mass (Li et al., 2007). Marjakangas et al. (2015) proposed utilising oleaginous yeast, which is naturally present in POME for lipid production and COD removal, as wastewater sterilisation in large-scale operation is not economically practicable. Their study achieved 94.5% COD removal using oleaginous yeast *Graphium penicillioides* grown in POME, with 29.1% (w/w) lipid production.

According to Islam et al. (2018), yeast could take up minerals and vitamins from POME to stimulate cell growth and product formation. *Rhodotorula toruloides* of the *Pucciniomycotina* subphylum exhibited an unprecedented rate of high cell density fermentation, high carotenoid, and lipid productivity using cheap substrates as growth medium (Liu et al., 2018). Furthermore, *Rhodotorula toruloides* showed COD reduction in wastewater by 72.3% using the feeding mode strategy (Qin et al., 2017). Arous et al. (2019) demonstrated that POME bioremediation by yeast has contributed to a COD removal of about 90%. Meanwhile, Karim et al. (2021) demonstrated an enhanced lipid accumulation and COD removal of approximately 80% using a mixed culture of yeast and bacteria. Nevertheless, there has been limited study on the growth behaviour of *Rhodotorula toruloides* on raw POME media for lipid and biomass production.

A local oleaginous yeast *Rhodotorula toruloides* has been successfully isolated from an acidic environment from a copper mining site. Its growth behaviour on POME was examined in this study. The aim was to evaluate the potential of the yeast to reduce the COD levels of POME while simultaneously producing lipid as a valuable product. The results were compared with the yeast growth cultivated in yeast peptone dextrose (YPD) media as a controlled study. The resulting lipid production and COD removal of the yeast were further evaluated. The findings of this study will lead to the utilisation of oleaginous yeast for a cost-effective POME treatment as part of bioremediation and simultaneously provide sustainable microbial oil production.

MATERIALS AND METHODS

Microorganism

Oleaginous yeast *Rhodotorula toruloides* was locally isolated from runoff water of Ranau River, Sabah (6.030324°; 116.657833°).

Preparation of POME Media

Palm oil mill effluent (POME) was collected from Langkon Palm Oil Mill, Kota Marudu, Sabah (6.57098°; 116.70748°). The POME samples were placed into a 250-ml conical flask and then filtration through Whatman filter paper (0.2 nm) to remove mud and large particles before yeast cultivation. Unfiltered POME samples, which did not undergo any filtration process, were used as a comparison study for yeast cultivation. *Rhodotorula toruloides* were grown on a yeast peptone dextrose (YPD) agar plate and incubated for 24 hours at 30 °C. The cultures were then inoculated in a 250 ml Erlenmeyer flask containing 150 ml sterilised media (filtered POME, unfiltered POME, or YPD media) and incubated at 30 °C on an incubator shaker (New Brunswick Scientific) at 180 rpm for 24 hours for preparation of inoculum. Every new experiment was carried out using fresh culture from the inoculum.

Cultivation of Yeast in POME Media

Approximately 150 ml of filtered POME, unfiltered POME, and YPD media were prepared in 250 ml conical flasks. The initial pH of the media was adjusted to 6.0 and sterilised by autoclaving at 121 °C for 20 minutes before use. The media was inoculated with 10% (v/v) of a mid-log phase *Rhodotorula toruloides* inoculum (approximately $OD_{600} = 1.8-2.2$) and incubated at room 30 °C on an incubator shaker at 180 rpm for 72 hours. Samples were harvested every 2 hours interval for biomass determination to determine the growth behaviour of the yeast on POME. The pH changes in the cultivation media were also observed and determined using a pH meter (Professional Benchtop pH Meter BP3001, Trans Instruments, Singapore). Lipid production and COD content of the waste media were determined after 36 hours of the cultivation of *Rhodotorula toruloides*.

Characterisation of POME

The physicochemical properties of the filtered and unfiltered POME as a culture medium for *Rhodotorula toruloides* were analysed at the Water Research Unit, Faculty of Science and Natural Resources, Universiti Malaysia Sabah. The samples were analysed using the DOE method (Department of Environment, 2011) concerning Revised Standard Methods (1985). The tested water quality analyses include the biochemical oxygen demand (BOD), chemical oxygen demand (COD), total Kjeldahl nitrogen, total suspended solids, oil, grease, ammoniacal nitrogen, and orthophosphate.

Determination of Biomass Concentration

The harvested samples were placed in a 50-ml falcon tube, spun down at 3,400 x g for 5 minutes and washed with deionised distilled water. After that, the samples were dried at 60 °C temperature until constant weight (Kitcha & Cheirsilp, 2011). Biomass concentration was measured gravimetrically every two hours using Equation 1 (Nascimento et al., 2013).

$$\text{Biomass concentration (mg/ml)} = (x_2 - x_1) \quad (1)$$

Where x_2 was the weight of the centrifuge tube together with biomass after drying, and x_1 was the weight of the centrifuge tube.

Determination of Lipid Content

The harvested biomass samples were centrifuged, dried, and grounded into powder biomass. Lipid content in the sample was determined using Bligh and Dyer's (1959) method, whereby a 1:1:1 mixture of chloroform, methanol, and distilled water were added into the pre-weighed dried biomass samples inside a falcon tube. The mixtures were vortexed and left on the bench for 1 hour. Next, the samples were centrifuged at 5,500 x g for 5 min, resulting in the formation of two layers. The top layer (methanol and water) was discarded, while the bottom layer (chloroform and lipid) was transferred into a pre-weighed microcentrifuge tube. Speed-vacuum (V-AQ) was used for 1.5 hours to complete methanol removal. Finally, the extracted lipid was weighed, and the lipid content was calculated using Equation 2 (Cordeiro et al., 2017).

$$\text{Lipid content (\%)} = [\text{extracted lipid extract (g)} / \text{dry biomass (g)}] \times 100 \quad (2)$$

Determination of COD Removal

The COD content of POME samples was determined before and after treatment with *Rhodotorula toruloides*. The samples were analysed using the DOE method (Department of Environment, 2011) with reference to Revised Standard Methods (1985) at the Water Research Unit, Faculty of Science and Natural Resources, University Malaysia Sabah. The removal of COD was determined using Equation 3.

$$\text{COD removal (\%)} = \frac{\text{COD}_{\text{untreated POME}} - \text{COD}_{\text{treated POME}}}{\text{COD}_{\text{untreated POME}}} \quad (3)$$

Statistical Analysis

All experiments were carried out in triplicate. The data were analysed using the statistical tool in MS Excel 2010 and presented as a mean value with an average standard deviation (SD) of < 5%.

RESULT AND DISCUSSION

Characteristics of POME

The collected POME sample was pre-filtered to remove mud and solid wastes. Its physicochemical properties were determined and compared to the unfiltered POME sample

as tabulated in Table 1. The unfiltered POME was found very thick, dark brownish, and odorous. The filtered POME showed a cleaner look (Figure 1a) and was free from mud and solid wastes present in the unfiltered POME (Figure 1b). The unfiltered POME may contain colloidal suspensions, including water, oil, and suspended solids from palm fruit residues, as Kaman et al. (2016) described. The BOD and COD content of the unfiltered POME was determined to be about 50,750 mg/L of BOD and 68,031 mg/L. After the filtration process, the BOD and COD values of POME were largely reduced to 70% and 50%, respectively. This finding parallels the study Lokman et al. (2019) reported on the successful reduction of BOD concentration by approximately 50% to 65% after the filtration process. The total Kjeldahl nitrogen in the unfiltered POME exceeded 30,000 mg/L, while the ammoniacal nitrogen was about 11 365 mg/L. It is important to note that total nitrogen and ammoniacal nitrogen content dropped by about 99% upon filtration. A similar observation was reported by Sajjad et al. (2018) on the removal of ammoniacal nitrogen of about 99% using a combination of biofilm and nano-filtration membrane. Both filtered and unfiltered POME were further evaluated for yeast growth and lipid production.

Table 1
Characteristics of POME

Parameters	Unit	Unfiltered	Filtered
Biochemical oxygen demand	mg/L	50, 750	15, 275
Chemical oxygen demand	mg/L	68, 031	33, 923
Total Kjeldahl nitrogen	mg/L	30, 845	280
Ammoniacal nitrogen	mg/L	11, 365	74
Orthophosphate	mg/L	23.20	18.6
Total suspended solids	mg/L	23, 244	136
Oil and grease	mg/L	16, 188	8

Growth Profile of *Rhodotorula toruloides*

Rhodotorula toruloides yeast was grown in filtered (Figure 1a), and unfiltered POME (Figure 1b) medium, and their growth characteristic were benchmarked to yeast cultured on YPD defined media (Figure 1c). The POME samples were darker than the YPD media as

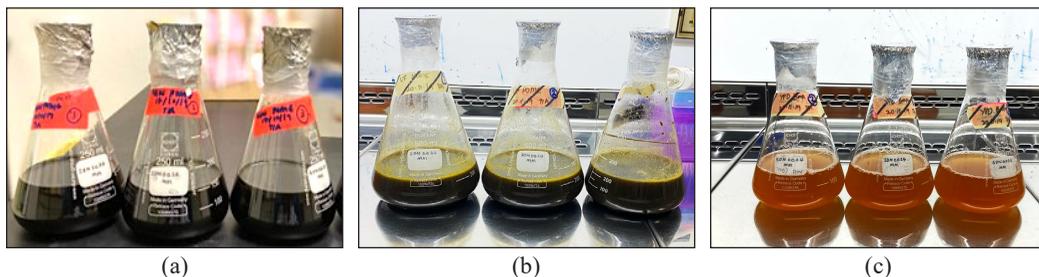


Figure 1. Cultivation of *Rhodotorula toruloides* in (a) filtered POME, (b) unfiltered POME, and (c) YPD broth

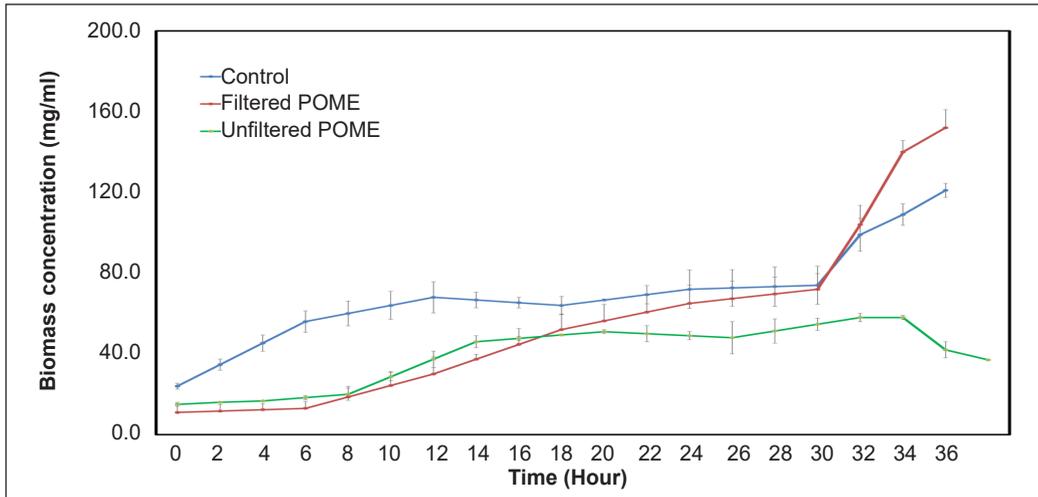


Figure 2. The growth profile of *Rhodotorula toruloides* in filtered POME, unfiltered POME, and YPD media

they contained high organic compounds. Figure 2 shows the growth profile of *Rhodotorula toruloides* in POME and YPD medium as a function of biomass concentration over time. As can be observed, the growth of *Rhodotorula toruloides* in filtered POME media resulted in a maximum biomass concentration of 152.0 mg/ml at 36 hours, as opposed to YPD media (121.0 mg/ml) and unfiltered POME media (37.0 mg/ml).

According to the findings, the yeast growth in filtered media achieved the highest growth compared to control and unfiltered POME media, 152 mg/ml > 121 mg/ml > 37 mg/ml, respectively. According to Hahn et al. (2005), microorganisms have grown more aggressively in rich, complex media than in defined media due to biosynthetic precursors, which can be transported directly into anabolic pathways and aid in reducing the production of biosynthetic precursors while saving metabolic energy. From 30 to 36 hours, both control and filtered POME medium showed a fast surge in growth, whereas the culture in unfiltered POME displayed an early death phase trend. The sudden rise in concentration from 30 to 36 hours, from 74 to 121 mg/ml, was most likely induced by the nutrient available in the media, particularly carbon source, which regulates the biosynthesis of internal lipids in yeast cells (Kot et al., 2019).

As for the cultivation of *Rhodotorula toruloides* in unfiltered POME, the biomass concentration did not exceed the biomass yield in the control media. The enormous quantity of colloidal particles, suspended solids, and bacteria inside the unfiltered POME, as indicated in Table 1, made such an event possible by impeding yeast development. The physical form of the unfiltered POME has most likely hindered cell mobility in search of food sources. The observation was seen here also corroborates the findings of earlier reported studies. Soleimaninanadegani et al. (2014) described the growth inhibition as due to the colloidal and suspended components floating on wastewater surface, obstructing the

respiration process of the microorganisms. Furthermore, Bala et al. (2018) showed that the existence of indigenous bacteria in POME could result in a food competition. The cells' growth dropped drastically after 32 hours of cultivation from 58.0 mg/ml to 37.0 mg/ml, implying the beginning of the death phase.

These growth trends of *Rhodotorula toruloides* demonstrated its ability to utilise carbon, nitrogen, and other nutrient sources in POME. It signifies the great potential of agricultural wastes as growth promoting material for microorganisms cultivation. Other studies have also described waste as growth media as a new technique for long-term agricultural development. By combining food waste with biochar obtained from oil palm trash, sugarcane, and sawdust, Ma et al. (2019) revealed the potential of reusing diaper waste as an orchid growing substrate. A unique combination of used diapers and other food wastes, such as banana peels and eggshells, has also been demonstrated to be a green method for mushroom cultivation with no accumulation of undesired toxic products (Ma et al., 2020, Nam et al., 2018).

pH Changes of *Rhodotorula toruloides*

The pH of the culture media was monitored. As shown in Figure 3, the pH of filtered POME has swiftly shifted from acidic to alkali compared to the control media. It is observed that during the early hours of yeast cultivation, from 0 to 6 hours, the pH value of filtered and unfiltered POME dropped from pH 6.0 to 5.6 and 5.7, respectively. Meanwhile, the pH value of control media remained constant throughout yeast cultivation. Nevertheless, filtered POME media showed a rapid increase in pH from 5.6 to 7.1 up to 34 hours of cultivation, then a slight decrease to 6.8 at hour 36. Unfiltered POME media followed the same pattern, although not as quickly as filtered POME media.

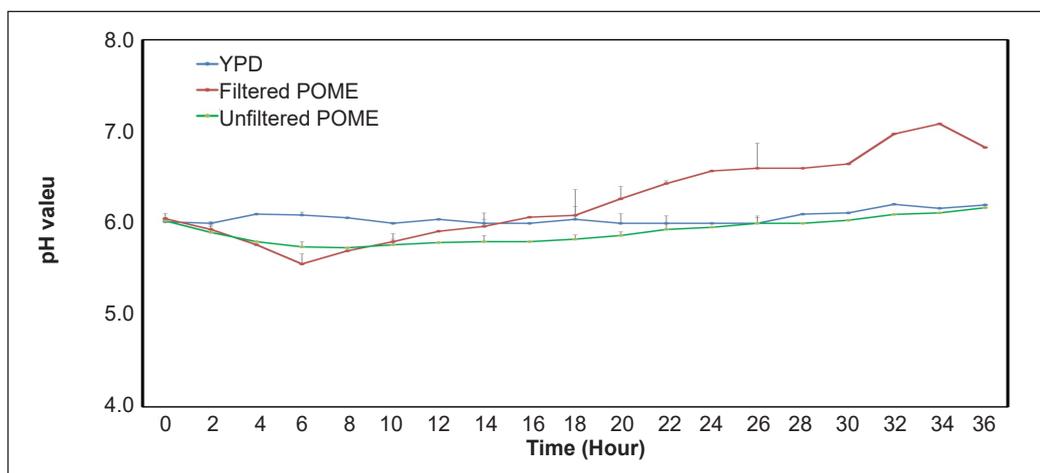


Figure 3. pH changes of media during cultivation of *Rhodotorula toruloides* in filtered POME, unfiltered POME, and YPD media

The pH reduction in filtered POME indicated that glucose was consumed in the media, resulting in the formation of acid chemicals by the yeast, as reported by Yu et al. (2018). Due to glucose depletion and the yeast's inability to produce acids, the pH rapidly rose after 6 hours of cultivation. The slight decline of pH in filtered POME after 36 hours of cultivation could be due to the lactic acid fermentation by the existing bacteria in the wastewater media (Gientka et al., 2017), as evidenced by the rapid increase in biomass concentration at 36 hours (Figure 2).

According to Gientka et al. (2017), the growth development of yeast is favoured in a slightly acidic environment, whereas an alkalisation environment may induce stress on the yeast. On the other hand, the yeast in this study yielded a higher biomass concentration as the pH increased. It might be due to the varied types of media and yeast strain employed, which could trigger yeast growth rate and pH changes (Jiru et al., 2017).

Lipid Production of *Rhodotorula toruloides*

As the growth of *Rhodotorula toruloides* was greater in filtered POME, the lipid production and COD removal ability of the yeast were further investigated in this media. Figure 4 depicts the lipid content of yeast grown in filtered POME and YPD media for 36 hours of cultivation. *Rhodotorula toruloides* grown in the POME medium had a greater lipid content (4.51%) than those produced in control media (0.94%). The yeast cultured in the POME medium had a high lipid content, corresponding to the high biomass yield depicted in Figure 3. Lipid formation in oleaginous bacteria normally occurs during secondary metabolic growth in a carbon excess and nitrogen limitation environment (Matsakas et al., 2017). The lipid accumulation was triggered by assimilating the excess substrate and converted to fat for storage (Li et al., 2007). An earlier study reported *Rhodotorula glutinis* lipid production up to 21% (w/w) grown in POME (Leiva et al., 2014). Oleaginous yeast *Lipomyces starkeyi*, on the other hand, generated approximately 7.4% (w/w) lipid content and enhanced the lipid production to 21.3% (w/w) by diluting POME with deionised water in a 1:1 ratio (Islam et al., 2018). The presence of a high number of inhibitors in undiluted POME hindered the lipid production and growth rate of *L. starkeyi*. The cultivation of microalgae *Chlorella* sp on POME also

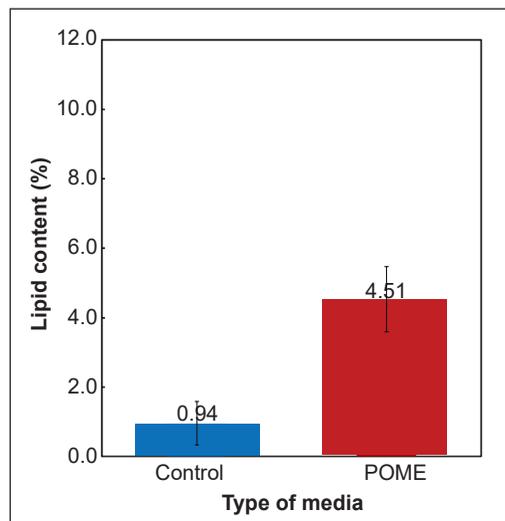


Figure 4. Lipid production of *Rhodotorula toruloides* grown in YPD media and filtered POME for 36 hours of cultivation at 30°C

agreed that enhanced lipid production could be obtained using diluted POME media (Kamyab et al., 2017). The lipid production of *Rhodotorula toruloides* probably can be enhanced further in future by optimising the concentration of POME.

COD Removal of *Rhodotorula toruloides*

Figure 5 depicts the COD values of POME samples before and after *Rhodotorula toruloides* cultivation in the POME. The COD content of untreated POME was determined at 33 923 mg/L. After the treatment, the COD value was significantly reduced to 19 290 mg/L. The data shows that the COD removal was about 43%. Based on the result, yeast has significantly utilised the nutrients present in POME during the biological process, such as respiration, for its growth. *Rhodotorula toruloides* converted the organic component found in POME into lipids, decreasing the COD level. A previous study showed that the cultivation of *Yarrowia lipolytica* on POME achieved approximately 88.4% of COD removal (Louhasakul et al., 2020), about 45% higher than the present study. It was due to the addition of a biosurfactant that aided the cells in utilising the carbon compounds, specifically the hydrocarbons compounds in POME (Louhasakul et al., 2020). Thus, this explained the small amount of COD reduction in this current study due to the application of pure POME.

Higher COD removal (72–95%) was reported using a 50% diluted POME concentration. Islam et al. (2018) found that 50% diluted POME supports a greater growth rate of oleaginous yeast due to fewer phenolic compounds present in POME, resulting in increased COD reduction. As yeast thrives in nitrogen-limited circumstances, a high quantity of phenolic compounds may impede yeast growth and metabolic activity inside the POME owing to a high ammonium concentration (Islam et al., 2018). According to Saenge et al. (2011), COD removal can also be enhanced by 65.1% by adding nitrogen sources like yeast extract into POME. Even though yeast extract is a good nitrogen source, its high cost makes it uneconomical for industrial operations. *Rhodotorula sp.* could convert organic compounds in POME to lipid and carotenoids, which is beneficial for COD reduction (Saenge et al., 2011).

Rhodotorula toruloides grew best when cultivated in filtered POME with a 152 mg/ml biomass concentration and a lipid content of 4.51%. This study shows the ability of the yeast to utilise the nutrients present in POME for growth. The filtered POME

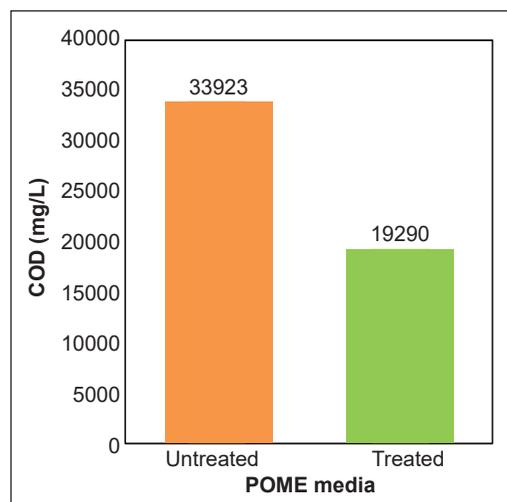


Figure 5. COD values of filtered POME before and after the cultivation of *Rhodotorula toruloides*

sample was found favourable for the cultivation of *Rhodotorula toruloides* as growth media. However, suspended solids and other colloidal particles in unfiltered POME may reduce the accessibility of available nutrients for yeast growth. By benchmarking the findings with past reported studies, as summarised in Table 2, this study shows that *Rhodotorula toruloides* grown in filtered POME were comparable to or even better than the reported oleaginous microorganisms cultivated in various cheap media for lipid production and COD removal. The study's findings provide beneficial insight into utilising agricultural wastes as cheap media for practical industrial applications.

Table 2

Comparative study of various types of microorganisms cultivated in different types of media based on their lipid content and COD removal efficiency

Microorganisms	Types of media	Lipid content % (w/w)	COD removal (%)	Reference
<i>Rhodotorula toruloides</i>	POME	4.51	43.0	This study
<i>D. etchellsii</i> BM 1	62.4% expired soft drinks + 37.6% olive mill wastewater	28.1	41.3	Arous et al. (2016)
<i>C. tropicalis</i>	Olive mill wastewater + 132 C/N ratio	78.7	13.9	Dias et al. (2021)
<i>C. tropicalis</i>	Olive mill wastewater + Urea	13.3	36.3	Dias et al. (2021)
<i>T. coremiiforme</i>	acetone-butanol-ethanol fermentation wastewater + oleic acid (C16:1)	1.06	64.14	Xiong et al. (2016)
<i>T. cutaneum</i>	acetone-butanol-ethanol fermentation wastewater + oleic acid (C16:1)	0.77	69.46	Xiong et al. (2016)
<i>Rhodotorula toruloides</i>	Cellulosic ethanol fermentation wastewater	17.32	38.72	Yao et al. (2020)

CONCLUSION

The efficiency of POME as growth media for *Rhodotorula toruloides* was evaluated and compared with yeast standard media. As evident by the characteristic analysis and growth profile of yeast, the reduced number of pollutants available in POME was found favourable for the yeast growth. Filtered POME enhanced biomass production about 4.1 times in comparison with non-filtered POME. The cultivation of *Rhodotorula toruloides* on POME has successfully reduced COD content by about 43 % and simultaneously increased the lipid production by 1.8-fold. The findings of this study indicate that POME can be used as a cost-effective substrate for the cultivation of *Rhodotorula toruloides*, which would be beneficial for waste treatment and biofuel production. For future study, *Rhodotorula toruloides* can be co-cultivated with other oleaginous microorganisms for enhanced lipid production and waste removal.

ACKNOWLEDGEMENT

Universiti Malaysia Sabah funded the project under the Niche Grant Scheme (SDN0024-2019). Special thanks to Assoc. Prof. Dr Kenneth Francis Rodrigues and Langkon Palm Oil Mill for providing valuable research samples.

REFERENCES

- Abdurahman, N. H., Rosli, Y. M., Azhari, N. H., & Tam, S. F. (2011). Biomethanation of palm oil mill effluent (POME) by membrane anaerobic system (MAS) using POME as a substrate. *International Journal of Chemical and Molecular Engineering*, 5(3), 220-225.
- Abdulsalam, M., Man, H. C., Idris, A. I., Yunus, K. F. & Abidin, Z. Z. (2018). Treatment of palm oil mill effluent using membrane bioreactor: Novel processes and their major drawbacks. *Water*, 10(9), 1165. <https://doi.org/10.3390/w10091165>
- Amat, A. N. A., Tan, Y. H., Lau, W. J., Lai, G. S., Ong, C. S., Mokhtar, N. M., Sani, N. A. A., Ismail, A. F., Goh, P. S., Chong, K. C., & Lai, S. O. (2015). Tackling colour issue of anaerobically treated palm oil mill effluent using membrane technology. *Journal of Water Process Engineering*, 8, 221-226. <https://doi.org/10.1016/j.jwpe.2015.10.010>
- Arous, F., Frikha, F., Triantaphyllidou, I. E., Aggelis, G., Nasri, M., & Mechichi, T. (2016). Potential utilization of agro-industrial wastewaters for lipid production by the oleaginous yeast *Debaryomyces etchellsii*. *Journal of Cleaner Production*, 133, 899-909. <https://doi.org/10.1016/j.jclepro.2016.06.040>
- Arous, F., Jaouani, A., & Mechichi, T. (2019). Oleaginous microorganisms for simultaneous biodiesel production and wastewater treatment. In M. P. Shah and S. Rodriguez-Couto (Eds). *Microbial Wastewater Treatment* (pp.153-174). Elsevier. <https://doi.org/10.1016/b978-0-12-816809-7.00008-7>
- Bala, J. D., Lalung, J., Al-Gheethi, A. A. S., Hossain, K., & Ismail, N. (2018). Microbiota of palm oil mill wastewater in Malaysia. *Tropical Life Sciences Research*, 29(2), 131-163. <https://doi.org/10.21315/tlsr2018.29.2.10>
- Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8), 911-917. <https://doi.org/10.1139/o59-099>
- Cheirsilp, B., & Louhasakul, Y. (2013). Industrial wastes as a promising renewable source for production of microbial lipid and direct transesterification of the lipid into biodiesel. *Bioresource Technology*, 142, 329-337. <https://doi.org/10.1016/j.biortech.2013.05.012>
- Cheng, J., Zhu, X., Ni, J., & Borthwick, A. (2010). Palm oil mill effluent treatment using a two-stage microbial fuel cells system integrated with immobilized biological aerated filters. *Bioresource Technology*, 101(8), 2729-2734. <https://doi.org/10.1016/j.biortech.2009.12.017>
- Chin, M. J., Poh, P. E., Tey, B. T., Chan, E. S., & Chin, K. L. (2013). Biogas from palm oil mill effluent (POME): Opportunities and challenges from Malaysia perspective. *Renewable and Sustainable Energy Reviews*, 26, 717-726.

- Cordeiro, R., S., Vaz, I. C., Magalhaes, S., & Barbosa, F. A. (2017). Effects of nutritional conditions on lipid production by cyanobacteria. *Anais da Academia Brasileira de Ciências*, 89(3), 2021-2031. <https://doi.org/10.1590/0001-3765201720150707>
- Department of Environment. (2011). *Revised Standard Methods (1985) for analysis of rubber & palm oil mill effluent (3rd edition)*. Ministry of Natural Resources and Environment. https://enviro2.doe.gov.my/ekmc/wp-content/uploads/2016/10/CETAK_REVISIED_PINDA-4_14.12.2011-28129.pdf
- Dias, B., Lopes, M., Ramôa, R., Pereira, A. S., & Belo, I. (2021). *Candida tropicalis* as a promising oleaginous yeast for olive mill wastewater bioconversion. *Energies*, 14(3), Article 640. <https://doi.org/10.3390/en14030640>
- Foo, K. Y., & Hameed, B. H. (2010). Insight into the applications of palm oil mill effluent: A renewable utilization of the industrial agricultural waste. *Renewable and Sustainable Energy Reviews*, 14(5), 1445-1452. <https://doi.org/10.1016/j.rser.2010.01.015>
- Gientka, I., Kieliszek, M., Jermacz, K., & Błażejczak, S. (2017). Identification and characterization of oleaginous yeast isolated from kefir and its ability to accumulate intracellular fats in deproteinated potato wastewater with different carbon sources. *BioMed Research International*, Article 6061042. <https://doi.org/10.1155/2017/6061042>
- Hahn-Hägerdal, B., Karhumaa, K., Larsson, C. U., Gorwa-Grauslund, M., Görgens, J., & van Zyl, W. H. (2005). Role of cultivation media in the development of yeast strains for large scale industrial use. *Microbial Cell Factories*, 4, Article 31. <http://dx.doi.org/10.1186/1475-2859-4-31>. <https://doi.org/10.1016/j.rser.2013.06.008>
- Islam, M. A., Yousuf, A., Karim, A., Pirozzi, D., Khan, M. R., & Wahid, Z. A. (2018). Bioremediation of palm oil mill effluent and lipid production by *Lipomyces starkeyi*: A combined approach. *Journal of Cleaner Production*, 172, 1779-1787. <https://doi.org/10.1016/j.jclepro.2017.12.012>
- Jiru, T. M., Groenewald, M., Pohl, C., Steyn, L., Kiggundu, N., & Abate, D. (2017). Optimization of cultivation conditions for biotechnological production of lipid by *Rhodotorula kratochvilovae* (syn, *Rhodospiridium kratochvilovae*) SY89 for biodiesel preparation. *3 Biotech* 7(2), Article 145. <https://doi.org/10.1007/s13205-017-0769-7>
- Kaman, S. P. D., Tan, I. A. W., & Lim, L. L. P. (2016). Palm oil mill effluent treatment using coconut shell-based activated carbon: Adsorption equilibrium and isotherm. *MATEC Web of Conferences*, 87, Article 03009. <https://doi.org/10.1051/mateconf/20178703009>
- Kamyab, H., Chelliapan, S., Shahbazian-Yassar, R., Din, M. F. M., Khademi, T., Kumar, A., & Rezaia, S. (2017). Evaluation of lipid content in microalgae biomass using palm oil mill effluent (POME). *JOM*, 69, 1361-1367. <https://doi.org/10.1007/s11837-017-2428-1>
- Karim, A., Islam, M. A., Mishra, P., Muzahid, A. J. M., Yousuf, A., Khan, M. M. R., & Faizal, C. K. M. (2021). Yeast and bacteria co-culture-based lipid production through bioremediation of palm oil mill effluent: A statistical optimization. *Biomass Conversion and Biorefinery*. <https://doi.org/10.1007/s13399-021-01275-6>
- Kitcha, S., & Cheirsilp, B. (2011). Screening of oleaginous yeasts and optimization for lipid production using crude glycerol as a carbon source. *Energy Procedia*, 9, 274-282. <https://doi.org/10.1016/j.egypro.2011.09.029>

- Kot, A. M., Błażejczak, S., Kieliszczak, M., Gientka, I., & Bryś, J. (2019). Simultaneous production of lipids and carotenoids by the red yeast *Rhodotorula* from waste glycerol fraction and potato wastewater. *Applied Biochemistry and Biotechnology*, 189, 589-607. <https://doi.org/10.1007/s12010-019-03023-z>
- Lee, C. H., Chong, D. Y. L., Hemmati, S., Elneghi, M. M., Foo, D. C. Y., How, B. S., & Yoo, C. (2020). A P-graph approach for the synthesis of national-wide bio-hydrogen network from palm oil mill effluent. *International Journal of Hydrogen Energy*, 45(35), 17200-17219. <https://doi.org/10.1016/j.ijhydene.2020.04.179>
- Leiva-Candia, D. E., Pinzi, S., Redel-Macías, M. D., Koutinas, A., Webb, C., & Dorado, M. P. (2014). The potential for agro-industrial waste utilization using oleaginous yeast for the production of biodiesel. *Fuel*, 123, 33-42. <https://doi.org/10.1016/j.fuel.2014.01.054>
- Li, Y., Zhao, Z. K., & Bai, F. (2007). High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme and Microbial Technology*, 41(3), 312-317. <https://doi.org/10.1016/j.enzmictec.2007.02.008>
- Liang, M., & Jiang, J. (2013). Advancing oleaginous microorganisms to produce lipid via metabolic engineering technology. *Progress in Lipid Research*, 52(4), 395-408. <https://doi.org/10.1016/j.plipres.2013.05.002>
- Liu, Y., Koh, C. M. J., Yap, S. A., Du, M., Hlaing, M. M., & Ji, L. (2018). Identification of novel genes in the carotenogenic and oleaginous yeast *Rhodotorula toruloides* through genome-wide insertional mutagenesis. *BMC Microbiology*, 18, Article 14. <https://doi.org/10.1186/s12866-018-1151-6>
- Lokman, N. A., Ithnin, A. M., Ramli, M. N., Bahar, N. D., & Yahya, W. J. (2019). Effect of filtered palm oil mill effluent (POME) via close-loop in-line microbubbles treatment chamber for biochemical oxygen demand (BOD) treatment. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 62(1), 10-19.
- Louhasakul, Y., Cheirsilp, B., Intasit, R., Maneerat, S., & Saimmai, A. (2020). Enhanced valorization of industrial wastes for biodiesel feedstocks and biocatalyst by lipolytic oleaginous yeast and biosurfactant-producing bacteria. *International Biodeterioration & Biodegradation*, 148, Article 104911. <https://doi.org/10.1016/j.ibiod.2020.104911>
- Ma, N. L., Khoo, S. C., Loke, L. K., Lam, S. S., & Tan, S. H. (2019). Growth media derived from solid waste for orchid *Dendrobium kingianum* culture. *Malaysian Applied Biology*, 48(1), 73-78.
- Ma, N. L., Khoo, S. C., Peng, W., Ng, C. M., Teh, C. H., Park, Y. K., Lam, S. S. (2020). Green application and toxic risk of used diaper and food waste as growth substitute for sustainable cultivation of oyster mushroom (*Pleurotus ostreatus*). *Journal of Cleaner Production*, 268, 122272. <https://doi.org/10.1016/j.chemosphere.2021.131477>
- Marjakangas, J. M., Lakaniemi, A., Koskinen, P. E., Chang, J., & Puhakka, J. A. (2015). Lipid production by eukaryotic microorganisms isolated from palm oil mill effluent. *Biochemical Engineering Journal*, 99, 48-54. <https://doi.org/10.1016/j.bej.2015.03.006>
- Matsakas, L., Giannakou, M., & Vörös, D. (2017). Effect of synthetic and natural media on lipid production from *Fusarium oxysporum*. *Electronic Journal of Biotechnology*, 30, 95-102. <https://doi.org/10.1016/j.ejbt.2017.10.003>

- Nam, W. L., Phang, X. Y., Su, M. H., Liew, R. K., Ma, N. L., Rosli, M. H. N., Lam, S. S. (2018). Production of bio-fertilizer from microwave vacuum pyrolysis of palm kernel shell for cultivation of oyster mushroom (*Pleurotus ostreatus*). *Science of The Total Environment*, 624, 9-16. <https://doi.org/10.1016/j.scitotenv.2017.12.108>
- Nascimento, I. A., Marques, S. S. I., Cabanelas, I. T. D., Pereira, S. A., Druzian, J. I., Souza, C. O. de, Vich, D. V., Carvalho, G. C. de, & Nascimento, M. A. (2013). Screening microalgae strains for biodiesel production: lipid productivity and estimation of fuel quality based on fatty acids profiles as selective criteria. *Bioenergy Research*, 6(1):1-13. <https://doi.org/10.1007/s12155-012-9222-2>
- Qin, L., Liu, L., Zeng, A. P., & Wei, D. (2017). From low-cost substrates to Single Cell Oils synthesized by oleaginous yeasts. *Bioresource Technology*, 245(Part B), 1507-1519. <https://doi.org/10.1016/j.biortech.2017.05.163>
- Saenge, C., Cheirsilp, B., Suksaroge, T. T., & Bourtoom, T. (2011). Efficient concomitant production of lipids and carotenoids by oleaginous red yeast *Rhodotorula glutinis* cultured in palm oil mill effluent and application of lipids for biodiesel production. *Biotechnology and Bioprocess Engineering*, 16, 23-33. <https://doi.org/10.1007/s12257-010-0083-2>
- Sajjad, A. A., Teow, Y. H., & Hussain, A. W. M. (2018). Sustainable approach of recycling palm oil mill effluent (POME) using integrated biofilm/membrane filtration system for internal plant usage. *Jurnal Teknologi*, 80(4), 165-172. <https://doi.org/10.11113/jt.v80.11054>
- Shahbandeh, M. (2021, January 27). Palm oil: Global production volume 2012/13-2020/21. *Statista*. <https://www.statista.com/statistics/613471/palm-oil-production-volume-worldwide/>
- Soleimaninanadegani, M., & Manshad, S. (2014). Enhancement of biodegradation of palm oil mill effluents by local isolated microorganisms. *International Scholarly Research Notices*, 2014, Article 727049. <https://doi.org/10.1155/2014/727049>
- Xiong, L., Huang, C., Chen, X. F., Hu, W. X., Li, X. M., Qi, G. X., Wang, C., Lin, X. Q., Li, H. L., & Chen, X. D. (2016). Comparison of fermentation by mono-culture and co-culture of oleaginous yeasts for ABE (acetone-butanol-ethanol) fermentation wastewater treatment. *Journal of Environmental Chemical Engineering*, 4(4), 3803-3809. <https://doi.org/10.1016/j.jece.2016.08.025>
- Yao, S., Xiong, L., Chen, X., Li, H., & Chen, X. (2020). Comparative study of lipid production from cellulosic ethanol fermentation wastewaters by four oleaginous yeasts. *Preparative Biochemistry & Biotechnology*, 51(7), 669-677. <https://doi.org/10.1080/10826068.2020.1852416>
- Yu, D., Wang, X., Fan, X., Ren, H., Hu, S., Wang, L., Shi, Y., Liu, N., & Qiao, N. (2018). Refined soybean oil wastewater treatment and its utilization for lipid production by the oleaginous yeast *Trichosporon fermentans*. *Biotechnology for Biofuels*, 11, Article 299. <https://doi.org/10.1186/s13068-018-1306-6>



The Relationship Between Bangladesh's Financial Development, Exchange Rates, and Stock Market Capitalization: An Empirical Study Using the NARDL Model and Machine Learning

Rehana Parvin

*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Minden, Penang, Malaysia
Department of Quantitative Sciences, International University of Business Agriculture and Technology,
Dhaka-1230, Bangladesh*

ABSTRACT

This research looks at the interplay between financial development, exchange rates, and the stock market in Bangladesh from 1995 to 2019 and employs the Nonlinear Autoregressive Distributed Lag (NARDL) model. The machine learning technique uses the iterative classifier optimizer to beat other classifiers in stock market capitalization prediction. According to our NARDL findings, changes in financial development and exchange rates positively impact stock market capitalization in Bangladesh. Negative changes in financial development and the currency rate, on the other hand, have mixed long-term and short-term consequences for the stock market. The dynamic multiplier graphs show that the response of the stock market capitalization to positive changes in financial development and exchange rates is nearly comparable to the response to negative changes. According to the Wald test, there are asymmetries among variables. We urge governments to remove barriers to development, upgrade infrastructure, expand the stock market's capacity, and restore market participants' confidence in the Bangladesh stock market.

Keywords: Dynamic multiplier, iterative classifier, NARDL, prediction, Wald test

ARTICLE INFO

Article history:

Received: 08 September 2021

Accepted: 23 March 2022

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.11>

E-mail address:

rehana.parvin@iubat.edu

INTRODUCTION

The stock market's performance substantially impacts a country's economic development by bringing surplus fundholders and users together. Creating an enabling climate for investor confidence is crucial to the stock market's growth and development in a market economy. Bangladesh's economy

is heavily reliant on the stock exchange. As a result, it is crucial to understand which microeconomic issues influence stock performance. Financial markets, like stock markets, are both parts of the financial system. A stronger and more stable financial system can boost overall investment and savings rates (Tuyon & Ahmad, 2016, Alquraan et al., 2016). The stock market can satisfy commercial firms' continuous financial requirements if a favorable climate for developing confidence in both stock market operations and investors. Therefore, creating an enabling climate for investor confidence in a market economy is critical to the stock market's growth and development (Kolapo & Adaramola, 2012).

Financial and economic data forecasting is crucial for academics and decision-makers to forecast their own and competitors' future status (Elliott & Timmermann (2016). Financial forecasting assists in making better business decisions without jeopardizing financial stability. Furthermore, it establishes reasonable expectations and provides a better picture of the company's future. Forecasts are important in decision-making since they help improve the efficiency of the process. As a result, businesses use capital budgeting as a tool to prepare for and manage such huge investments. However, identifying which features can be used to forecast a company's market value might be difficult. In addition to capital accumulation, long-term growth per capita fosters economic expansion. Therefore, academics and decision-makers must be able to forecast financial data to forecast their own and competitors' future performance. However, it is not always clear which features can be utilized to predict a company's market capitalization. We employed support vector machines, Bayesian networks, random forests, and iterative classifier optimizers to apply machine-learning methods to basic financial statement data. Market capitalization is a term used in the decision-making process to estimate a company's value. Various studies have been conducted to predict financial data using machine learning approaches, which backs up our findings (Chatzis et al., 2018; Uthayakumar et al., 2020). Patel et al. (2015) employed a two-stage machine learning technique for forecasting financial data. The authors used Support Vector Regression (SVR) in the first step and then fusion prediction models such as SVR-ANN, SVR-RF, and SVR-SVR in the second. They discovered that a two-stage fusion model reduced total prediction errors.

The foreign exchange rate has several effects on the stock market. For starters, domestic currency depreciation, or an increase in the foreign exchange rate, increases short-term exchange rate risk for international portfolio investors. As a result, stock market values will fall. Second, currency depreciation boosts long-term market performance by encouraging foreign portfolio investment (Drucker, 1978). Third, domestic currency depreciation may make local equities more desirable to importers, resulting in higher export revenues for the home country. Fourth, the market performance of multinational firms' international subsidiaries may be influenced by exchange rate fluctuations.

The devaluation of the domestic currency may result in higher anticipated inflation, casting doubt on the firm's future performance and triggering a short-term reduction in

stock prices. Some researchers have looked at the uneven impact of the exchange rate on the stock index. In contrast, others have discovered that stock index volatility has a symmetrical effect on exchange rate changes. According to Sheikh et al. (2020), the global financial crisis disrupted nonlinear co-integrating relationships between stock market performance and currency rates.

Some research looked at the relationship between macroeconomic fundamentals and stock market response using a linear relationship between the variables. The exchange rate and stock market performance have a positive link (Habiba & Zhang, 2020; Okere et al., 2021). On the other hand, some research shows that exchange rates and the stock market have a negative relationship (Delgado, et al., 2018; Singhal et al., 2019; Adeniyi & Kumeka, 2020). Kumar and Misra (2019) and Churchill et al. (2019) provide more evidence of an inverted U-shape relationship between oil prices, exchange rates, and the stock market.

Banks have power over Bangladesh's financial system. Direct financing through the issuance of shares is rapidly becoming more common, but direct financing through corporate bonds is virtually non-existent in this country. Previous studies have found that the financial sector contributes to economic growth by lowering the cost of obtaining information, completing transactions, and encouraging savings mobilization. The financial industry in Bangladesh contributes to economic growth by cutting the cost of receiving information, executing transactions, and encouraging savings mobilization. Financial depth indicators like private credit stock and market capitalization as a percentage of GDP are widely used to assess financial development. However, there is little evidence to back up the notion that financial progress and stock market performance are linked (Alanyali et al., 2013; Musallam, 2018). It is also considered the impact of financial development on investment and productivity. We find little evidence that monetary expansion 'leads' economic growth, either directly or indirectly. Importantly, our findings are applicable to research into which metrics should be used to assess financial development. Drehmann and Juselius (2014) discovered that comparing the private credit to GDP ratio to its trend is a good indicator. All the research investigated the ability of bank credit to predict crisis occurrences and other indicators. Our research adds to the body of knowledge by demonstrating that banks are more exposed to stock market shocks with a high private credit-to-GDP ratio. To the author's knowledge, no study has been conducted globally that takes financial development and stock market performance into account.

Only a few research have been undertaken in Bangladesh to strengthen the stock market. According to Hasan and Zaman (2017), there is no significant correlation between oil prices and stock market performance. Exchange rates and the stock market, on the other hand, have a positive and considerable association. Sarwar and Hussan (2016), to confirm the significant positive link between oil prices and stock market prices in Bangladesh, used the regression estimation method. Inflation and the stock market index have a one-way short-

run causal link, according to Ahmed et al. (2015). According to Khan and Yousuf (2013), exchange rates negatively link with stock prices, while the CPI does not affect stock prices. Even though the study mentioned above was conducted in Bangladesh in previous literature based on the stock market, it is insufficient to describe stock performance. To the best of the authors' knowledge, no research has been conducted in Bangladesh to examine how unequal microeconomic variables affect stock capitalization. We conducted this research to fill a gap in the literature.

Rahman et al. (2017) employ machine learning to find patterns in data to predict stock prices. Stock exchanges typically create a massive volume of organized and unstructured heterogeneous data. It is feasible to rapidly evaluate more complex, heterogeneous data and get more accurate findings using machine-learning algorithms. Most research has looked at the relationship between macroeconomic conditions and stock market performance from a linear perspective. Linear models, because of structural changes and short-term volatility, cannot state variables.

According to Raza et al. (2016), nonlinearity, such as structural fractures and asymmetric behavior resulting from bankruptcy or severe credit events, regularly influences market dynamics, especially when the sample period includes financial crises like 2007 and 2008. As a result, earlier research has shown contradictory findings. On the other hand, these variables show regular oscillations and nonlinear behavior overlooked in prior research. Previous studies have shown mixed results because of this. Furthermore, the linear model's variables will not effectively explain the stock market, resulting in a decrease in stock capitalization. The criteria utilized in this study were chosen to fill in the gaps described above and create accurate results, and they will cover both short and long-term changes in the sector and structural changes. It will be the first study to look at the asymmetric impact of microeconomic determinants on stock capitalization in Bangladesh. The nonlinear ARDL technique, proposed by Shin et al. (2014), allows us to assess asymmetries between research variables in both the short and long term.

MATERIALS AND METHOD

Data Description

This study uses annual data from 1995 through 2019 to explore the nonlinear dynamic links between Bangladesh's exchange rate, financial development, and stock market capitalization. The World Development Indicators (WDI) offered stock market capitalization (percentage of Gross Domestic Product, GDP) figures, while the Bangladesh Bank provided exchange rate (US Dollar/Bangladeshi Taka) data. There is no specific definition for the variable "financial development." Instead, researchers use various factors as proxies for financial or market development. Following Pan et al. (2019), we utilize the value of private bank credits divided by GDP as a financial development statistic as follows:

$$\text{Financial development} = \frac{\text{Private bank credits}}{\text{GDP growth (annual \%)}}$$

We used the natural logarithm to alter two variables: stock market capitalization and exchange rates. In this study, “STOCK” stands for stock market capitalization, “EXR” stands for exchange rates, and “FD” is for financial development.

METHODOLOGY

NARDL Model

We used a nonlinear autoregressive distribution lag (NARDL) model, as Shin et al. (2014) described, to study the long- and short-run nonlinear interactions between the variables. The following linear Equation 1 was proposed to investigate the influence of exchange rate and financial development on stock market capitalization in Bangladesh:

$$LNSTOCK_t = \alpha_0 + \alpha_1 LNEXR_t + \alpha_2 FD_t + u_t \quad (1)$$

STOCK, EXR, and FD represent stock market capitalization, exchange rate, and financial development.

Our model will be as follows, based on Ibrahim’s (2015) and Lacheheb & Sirag’s (2019) prior work and considering the asymmetric link between exchange rate, financial development, and stock capitalization (Equation 2):

$$LNSTOCK_t = \beta_0 + \beta_1 LNEXR_t^+ + \beta_2 LNEXR_t^- + \beta_3 FD_t^+ + \beta_4 FD_t^- + e_t \quad (2)$$

Where the long-run parameters relate to β_1 . Positive changes EXR^+ and FD^+ and negative changes EXR^- and FD^- respectively account for the nonlinear influence of our research variables. Positive and negative changes in their partial sums in the exchange rate and financial development are represented by the following Equations 3-6:

$$EXR_t^+ = \sum_{i=1}^t \Delta EXR_t^+ = \sum_{i=1}^t \max(\Delta EXR_i, 0) \quad (3)$$

$$EXR_t^- = \sum_{i=1}^t \Delta EXR_t^- = \sum_{i=1}^t \min(\Delta EXR_i, 0) \quad (4)$$

$$FD_t^+ = \sum_{i=1}^t \Delta FD_t^+ = \sum_{i=1}^t \max(\Delta FD_i, 0) \quad (5)$$

$$FD_t^- = \sum_{i=1}^t \Delta FD_t^- = \sum_{i=1}^t \min(\Delta FD_i, 0) \quad (6)$$

Under an unrestricted error correction representation, Equation 2 can be included in the following NARDL Equation 7:

$$\begin{aligned} \Delta LNSTOCK_t &= \beta + \sum_{i=1}^q \mu_i \Delta LNSTOCK_{t-i} + \sum_{i=1}^p \gamma_1^+ \Delta LNEXR^+_{t-i} + \\ &\sum_{i=1}^p \gamma_2^- \Delta LNEXR^-_{t-i} + \sum_{i=1}^p \gamma_3^+ \Delta FD^+_{t-i} + \sum_{i=1}^p \gamma_4^- \Delta FD^-_{t-i} + \theta_1 LNSTOCK_{t-1} + \\ &+ \theta_2 LNEXR^+_{t-1} + \theta_3 LNEXR^-_{t-1} + \theta_4 FD^+_{t-1} + \theta_5 FD^-_{t-1} + \varepsilon_t \end{aligned} \tag{7}$$

Where q and p indicate the lag order and $\beta_1 = \theta_2/\theta_1, \beta_2 = \theta_3/\theta_1, \beta_3 = \theta_4/\theta_1$ and $\beta_4 = \theta_5/\theta_1$ are long-run asymmetric effects. Accordingly, $\sum_{i=1}^4 \gamma_i$'s are measuring the short-run asymmetric effects.

There are various steps to using the NARDL model (Lacheheb & Sirag, 2019). First, the unit root tests Augmented Dickey-Fuller (ADF) and Phillips–Perron (PP) were used. Second, we used the Brock, Dechert, and Scheinkmakn (BDS) test to see if the variables had any asymmetric relationships. Thirdly, we checked the null hypothesis of $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$ jointly. Fourth, we establish both long-run and short-run asymmetry relationships among the variables using the Wald test. Fifth, we also employed an asymmetric cumulative dynamic multiplier (CDM).

We can evaluate the asymmetric effect by obtaining the cumulative dynamic multiplier of a unit change in $EXR^+_{t-1}, EXR^-_{t-1}, FD^+_{t-1}$, and FD^-_{t-1} respectively on $LNSTOCK$ (Equations 8-9);

$$m_h^+ = \sum_{j=0}^h \frac{\partial LNSTOCK_{t+j}}{\partial LNEXR^+_{t-1}}, m_h^- = \sum_{j=0}^h \frac{\partial LNSTOCK_{t+j}}{\partial LNEXR^-_{t-1}}, h=1, 2, 3, \dots \tag{8}$$

$$m_h^+ = \sum_{j=0}^h \frac{\partial LNSTOCK_{t+j}}{\partial FD^+_{t-1}}, m_h^- = \sum_{j=0}^h \frac{\partial LNSTOCK_{t+j}}{\partial FD^-_{t-1}}, h=1, 2, 3, \dots \tag{9}$$

Note that as $h \rightarrow \infty, m_h^+ \rightarrow \beta_1, \beta_3$ and $m_h^- \rightarrow \beta_2, \beta_4$.

Machine Learning Algorithm

In addition, we looked at the five basic machine-learning techniques for predicting stock capitalization: Support Vector Machine (SVM), Bayesian Network (BN), Adaptive boost, Iterative classifier optimizer, and Random Forest (RF).

Support Vector Machine (SVM). By expressing the instances as a set of two-type points in an N-dimensional space and producing an (N-1) dimensional hyperplane, the Support Vector Machine (SVM) splits them into two groups. SVM seeks to draw a straight line that splits those points into two categories and is as far away as possible from all of them.

Bayesian Network (BN). The Bayesian Classifier is based on Bayes’ theorem and the naive assumption that each pair of attributes is independent. We employ Weka’s implementation of Multilayer Perception. We compared three Weka techniques to continuous characteristics:

modeling them as a single normal, modeling them with kernel estimation, and discrediting them with supervised discretization. The Bayesian Network is an uncertainty modeling approach that employs nodes to represent variables and arcs to show direct relationships between them. When new information becomes available, the BNs model allows for dynamic updating of probabilistic assumptions about the variables.

Random Forest (RF). Random Forest (RF) generates many category trees during training and testing using many decision trees. It produces the class that is the mode of the classes as an output (classification). The decision tree learns simple decision rules based on data attributes. The more complicated the choice criteria are, the more accurate the model becomes, and the deeper the tree becomes. Overfitting of decision trees is prevented when using random decision forests. We employ Weka's Random Forest implementation.

Boosting. Freund (1997) proved that combining multiple weak models using "boosting" techniques could create a single, robust model. As a result, one of the most powerful prediction concepts in the last 20 years is an extraordinarily precise model (Hou et al., 2018). When predicting the initial weak model, boosting approaches employ the same data set to run the prediction models. Then, use prediction error rates to adjust sample and model weights to predict the next model combination, with the same weight assigned to each data set. As a result, boosting methods must be used in a specific order, which slows prediction times. However, compared to bagging methods, prediction accuracy improves.

RESULT AND DISCUSSION

Unit Root Test

When examining the order of integrating variables in time series data, the unit root test, also known as the stationary test, is the most critical criterion. In order to perform our empirical research, we used the Augmented Dickey-Fuller (ADF) and Phillips–Perron (PP) tests. The research findings are summarized in Table 1. The Schwarz information criteria (SIC) were used to determine the optimal lag structure, which included intercept and linear time trend at the level but did not include time trend in the first difference. The

Table 1
Unit root test

Variables	I(0)		I(1)	
	ADF	PP	ADF	PP
<i>LNSTOCK</i>	-1.020846	-1.813570	-6.883701**	-6.996329**
<i>LNEXR</i>	-4.719059**	-5.905791**	-3.682831**	-3.363641**
<i>FD</i>	-1.900001	-1.900001	-4.689254**	-4.687621**

Note. ** refer significant at 5% levels.

exchange rates are stationary at the level according to both tests, suggesting that variables $I(0)$ are, whereas all study variables are stationary at the first difference, indicating that variables $I(1)$. Therefore, we can use bound testing to approximate Equation 7 when there are no $I(2)$ variables.

Brock, Dechert and Scheinkmakn (BDS) Test

It is necessary to determine whether variables are linearly or nonlinearly connected before using the ARDL or NARDL model. The Brock, Dechert, and Scheinkmakn (BDS) tests were performed to check for asymmetric associations between the variables. The hypothesis that the residual terms of the variable follow a six-dimensional independent and identical distribution is accepted, as shown in Table 2. These data suggest that the variables are nonlinear, implying that the NARDL model should be used rather than the ARDL model.

Table 2
BDS nonlinearity test results

BDS Statistics Embedding Dimension=m					
Series	m=2	m=3	m=4	m=5	m=6
STOCK	0.138821**	0.183220**	0.251222**	0.281217**	0.291229**
EXR	0.131362**	0.198341**	0.274511**	0.291421**	0.302531**
FD	0.124421**	0.202532**	0.281245**	0.294507**	0.285850**

Note. Superscript ** indicates the acceptance of the residual alternative hypothesis at 5%

Bound Test for Co Integration

According to Bahmani-Oskooee and Bohl (2000), the long-run connection is determined by the model’s optimal lag section. The Akaike information criterion (AIC), the Schwarz information criterion (SIC), the sequentially modified LR test statistic (LR), the final prediction error (FPE), and the Hannan-Quinn information criterion (HQ) are used to determine the best lag order. The best lag was “2.” Table 3 shows the model estimation results for symmetric and asymmetric co-integration. The bound test verifies that there is no co-integration in a linear ARDL way because the F-statistic result of 1.529913 is less than the required lower limit of 3.79 at 5%. However, the F-statistic value of 6.549283 exceeds the upper critical constraint of 4.01 by 5%, suggesting that the nonlinear ARDL specification is co-integrated.

Table 3
Bounds test results for co-integration

Model specification	F-Statistic	95% lower bound	95% upper bound	Conclusion
Linear ARDL	1.529913	3.79	4.85	No co-integration
Nonlinear ARDL	6.665112	2.86	4.01	Co-integration

Note. The critical values are from Narayan (2005).

NARDL Short Run and Long-Run Estimates

As demonstrated in Table 4, a 1% increase in the exchange rate increases stock capitalization by 5.17% in the long run and by 3.57% in the short term. The higher profit created by greater export revenues because of the depreciation of the Bangladeshi currency will result in a rise in stock prices. Stock prices are predicted to grow as foreign portfolio investment in stocks of companies expecting higher export income rises to profit from unexpected profit opportunities. However, in the long run, a 1% decline in the exchange rate reduces stock capitalization by 6.51%—the weakening of the Bangladeshi currency wills detriment the stock market. Stock prices are likely to climb as foreign portfolio investment in companies with higher expected export income grows to profit from unexpected profit opportunities. The depreciation of the Bangladesh Taka is expected to impede stock market growth in the long run. Import costs will rise, reducing earnings and driving stock prices to fall. Furthermore, international investors may opt to sell their shares in those companies, resulting in a drop in the stock market and share prices.

Table 4
NARDL Short-run and long-run estimates

Variable	Short-run estimates		Variable	Long-run estimates	
	Coefficient	Std. Error		Coefficient	Std. Error
$\Delta LNEXR^+$	3.56648**	6.994976	$LNEXR^+$	5.136664**	2.701164
$\Delta LNEXR^-$	0.23165**	7.264923	$LNEXR^-$	-6.513424***	4.057760
ΔFD^+	0.520730**	0.146976	FD^+	0.696279**	0.140615
ΔFD^-	0.593837**	0.263425	FD^-	-0.029442	0.380227

Note. ***, ** refers significant at 10% and 5% levels of significance, respectively. The terms “+” and “-” refer to the positive and negative changes, respectively. EXR_W & FD_W Indicates the Wald test result for each variable.

Furthermore, both in the short and long run, we discover a positive and significant relationship between financial development and stock market capitalization in Bangladesh. Our research shows that a 1% increase in financial development results in a 0.520% increase in short-term stock capitalization and a 0.696% increase in long-term stock capitalization. Our research also shows that a 1% decrease in financial development results in a 0.5938% increase in short-term stock capitalization. On the other hand, our findings suggest that a negative change in financial development has no long-term relevance. If the financial industry expands and capitalizes on increased volumes, it positively impacts per capita output. More banking sector development is positively associated with a stock market capitalization in global market stocks. People who have access to credit are more likely to spend more, boosting the economy’s revenue. As a result, GDP (gross domestic product) rises, causing productivity to rise quicker. If the financial sector expands and capitalizes on increasing volumes, it will positively influence per capita output. Credit encourages

people to spend more, resulting in an increase in revenue for the economy. Credit is used to purchase productive assets, increasing revenue and stock market value.

The Wald Test and Diagnostic Test Results

Some diagnostic tests were also performed to support the NARDL model’s dependability, as shown in Table 5. To determine error normalcy, researchers utilized the Jarque-Bera (J-B) test, the Ramsey RESET test, and the Autoregressive Conditional Heteroskedasticity (ARCH) up to order 2 for heteroskedasticity, and the serial autocorrelation LM test up to level 2 for serial autocorrelation. All diagnostic tests show that the NARDL model is accurate, meaning that it is trustworthy. The speed of adjustment (SOA) is a metric that assesses how rapidly organizations narrow the gap between their leverage last year and the leverage they want this year. According to the findings of this calculation, the rate of adjustment is around 0.88. As a result, the adjustment time is 1.135 years. After a shock, the deviance returns to normal after about a year. This rapid adjustment states that companies pursue target capital structures. When their leverage ratios vary from these targets, they make financial decisions to close the gap between the previous year’s leverage and the current period’s target leverage. The Wald test was performed to confirm the nonlinearities between the variables under examination. According to the Wald test results, asymmetries across variables occur at 5%.

Table 5
The Wald test and diagnostic test results

R^2	0.923248
CointEq(-1)*	-0.881251**
J-B [prob]	0.597012
R-R [prob]	0.3526
LM(1) [prob]	0.1769
LM(2) [prob]	0.1428
ARCH(1) [prob]	0.8207
ARCH(2) [prob]	0.4932
$EXR_w[X^2, prob]$	[25.20510, 0.0001]
$FD_w[X^2, prob]$	[19.30123, 0.0000]

Note. J-B and R-R refer Jarque-Bera test for error normality and the Ramsey-RESET test for model specification, respectively. LM test is for serial correlation, and the ARCH test is for autoregressive conditional Heteroskedasticity, up to the lag order given in the parenthesis. ** refers to significant at 5% levels of significance. Finally, EXR_w & FD_w Indicates the Wald test result for each variable.

The Cumulative Multipliers Dynamic Estimates

Figure 1 depicts the cumulative dynamic asymmetric multiplier findings. The solid (dashed) black lines depict how stock capitalization changes as EXR and FD move in opposite directions. On the other hand, the asymmetric line is a light-dash red line that falls between the lower and upper bands of the 95% confidence interval. As shown in Figure 1, all study variables exhibit positive and negative shocks that fall within 95% confidence lines, indicating that our NARDL model is stable. The difference between positive and negative components is depicted by an asymmetric curve, which denotes the dynamic multiplier linked to changes in EXR and FD. The dynamic multiplier graphs illustrate that

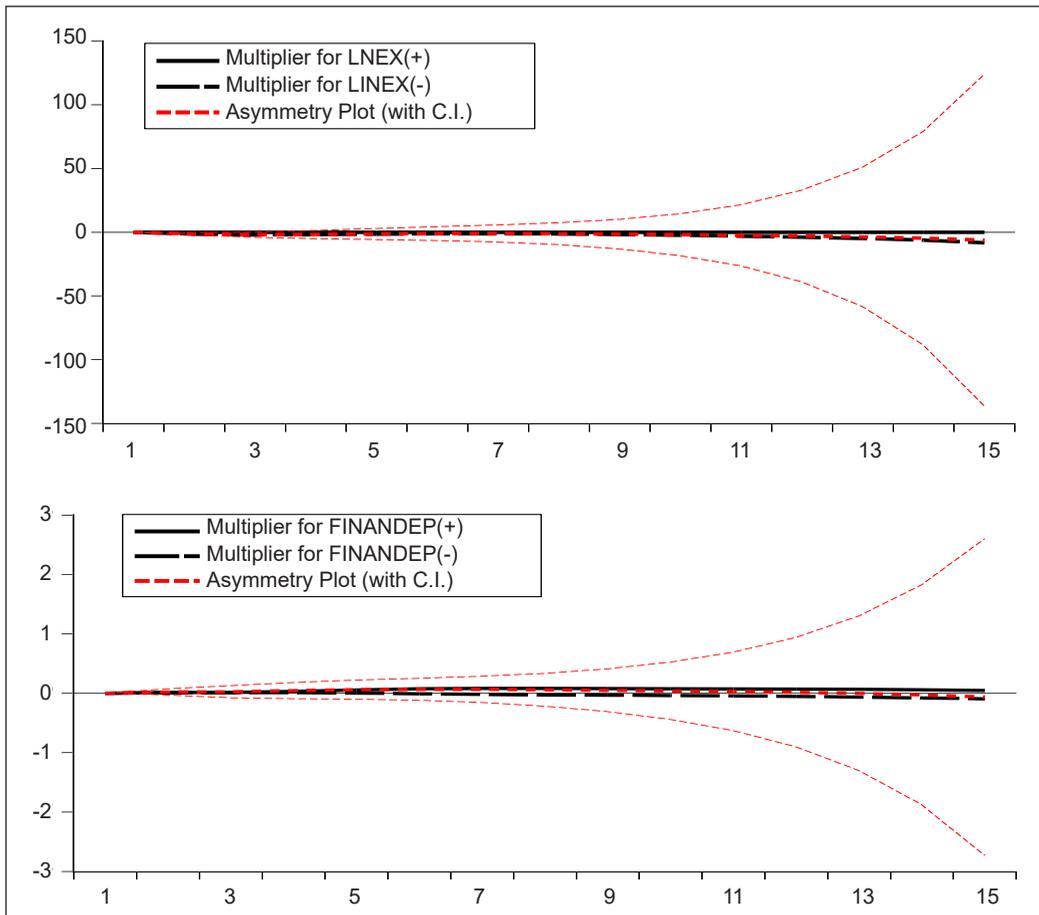


Figure 1. Cumulative dynamic multipliers (CDM)

the stock market capitalization responds nearly identically to positive changes in financial development and exchange rates as it does to negative changes.

NARDL Model Stability Check

Every statistical study’s robustness must be tested for parameter stability. Brown et al. (1975) suggested applying the stability test for the CUSUM and CUSUMSQ parameters after determining the short- and long-term coefficients. Figure 2 shows the results of the CUSUM and CUSUM Square tests, suggesting that the model is quite stable.

Prediction of Stock Market Capitalization in Machine Learning Approaches

Policymakers must forecast and estimate macroeconomic parameters such as stock market capitalization. Table 6 shows the results of machine learning classifiers. The Root Mean Squared Error (RMSE) is a form of error that is rather common. The most widely used

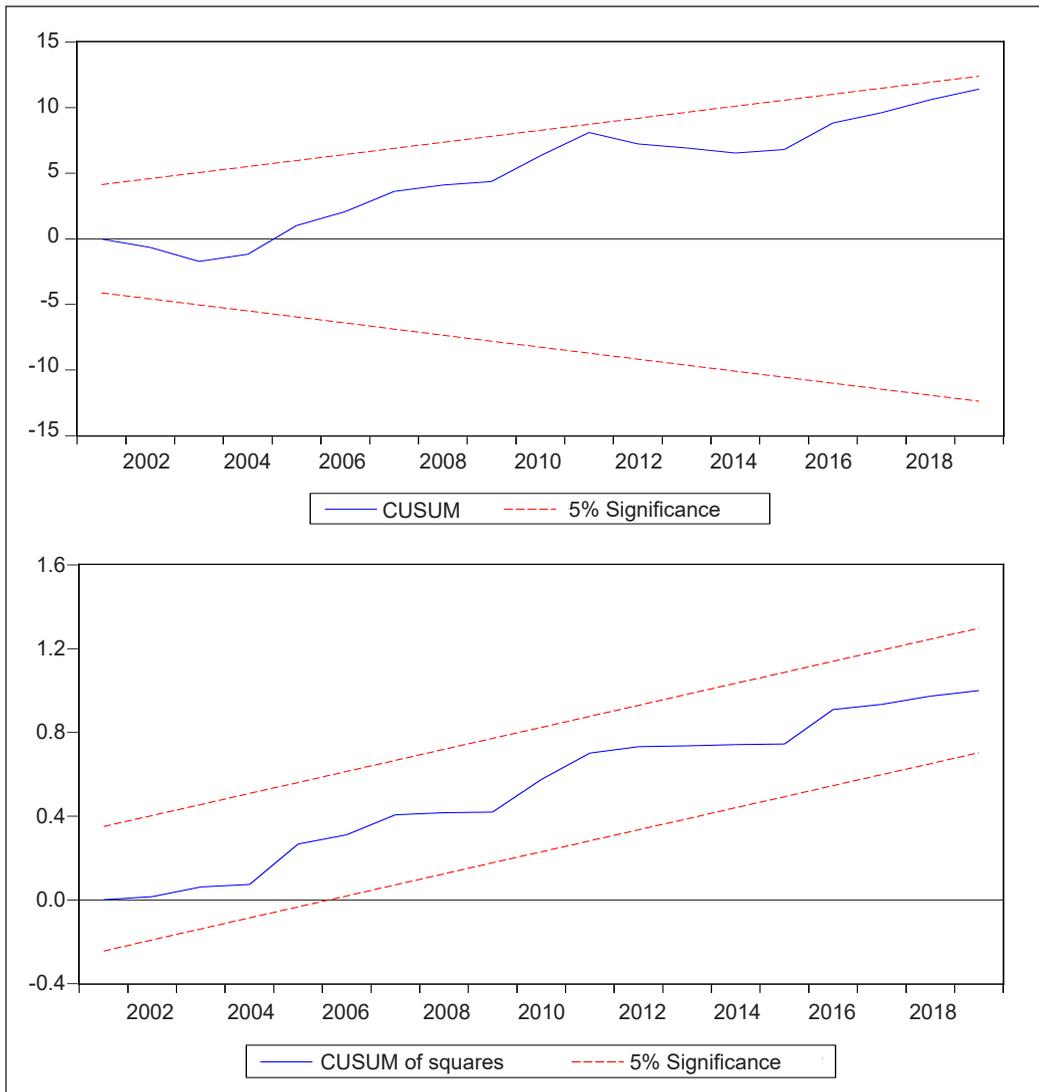


Figure 2. CUSUM and CUSUM of Square test for model stability

criteria are Mean Absolute Error and Mean Absolute Percentage Error, abbreviated as MAE and MAPE, respectively. The value of the criteria mentioned above is also compared in this study. The model with the lowest criteria value mentioned above would be the most suited. Ten-fold cross-validation is used to test the accuracy of all classifiers. The precision of BN and SVM is the same (80.11%).BN and SVM both have the same level of accuracy (80.11%). Adaptive boost has a greater accuracy rate (81.31%) than RF (75.56%). Overall, the iterative optimizer classifier beats other classifiers in stock capitalization prediction, with the lowest RMSE (0.2032), MAE (0.2112), and MAPE (10.112) values and the best accuracy (82.52%).

Table 6
Machine learning techniques for stock market capitalization prediction

Classifiers	Accuracy	RMSE	MAE	MAPE
SVM	80.11%	0.2321	0.2672	10.254
BN	80.11%	0.2412	0.2216	10.123
RF	75.56%	0.2541	0.2632	11.592
Adaptive boost	81.31%	0.2434	0.2442	11.586
Iterative classifier optimizer	82.52%	0.2032	0.2112	10.112

CONCLUSION

In Bangladesh, we used the NARDL model to examine the relationship between financial development, exchange rates, and stock market capitalization. From 1995 to 2019, the stock market has had strong, asymmetric effects on exchange rates and financial development. In the short run, both positive and negative changes in exchange rates and financial development have a large and favorable impact on stock capitalization. Beneficial changes in both variables, on the other hand, have a considerable and positive impact on stock capitalization. In contrast, negative changes in exchange rates have a long-term negative impact on stock capitalization. The dynamic multiplier graphs show that when a positive change occurs, the stock capitalization response is nearly identical to when a negative change occurs. Therefore, the iterative classifier optimizer is the most effective strategy for predicting stock capitalization with the lowest RMSE, MAE, and MAPE values and the highest accuracy.

Bangladesh's government may pass important financial market legislation. Furthermore, the government may monitor state-owned firms' credit performance, and researchers are studying the flow of financial resources from input to output to improve financial efficiency and, as a result, stock market performance. Finance for the private sector is a significant aspect in encouraging private investment due to a shortage of foreign capital inflows. As a result, the Bangladesh Bank should maintain the current monetary policy. The private sector credit ceiling might be raised to encourage private investment in Bangladesh while keeping a careful eye on credit utilization. The empirical findings suggest that policymakers should focus on long-term strategies. Therefore, the Bangladesh Bank should stick to its monetary policy, and the loan cap for the private sector might be raised to boost private investment in Bangladesh while keeping a careful eye on credit utilization.

A future study on the entire South Asian region should be completed using panel data based on the NARDL model to investigate nonlinear links between financial development-oil-gold-exchange rates and regional stock indexes.

ACKNOWLEDGEMENT

The author is thankful to Miyan Research Institute, International University of Business Agriculture and Technology (IUBAT).

REFERENCES

- Adeniyi, O., & Kumeka, T. (2020). Exchange rate and stock prices in Nigeria: Firm-level evidence. *Journal of African Business*, 21(2), 235-263. <https://doi.org/10.1080/15228916.2019.1607144>
- Ahmed, S. F., Islam, K. M., & Khan, M. (2015). Relationship between inflation and stock market returns: Evidence from Bangladesh. *DIU Journal of Business and Economics*, 9(1), Article 14.
- Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific Reports*, 3(1), Article 3578. <https://doi.org/10.1038/srep03578>
- Alquraan, T., Alqisie, A., & Al Shorafa, A. (2016). Do behavioral finance factors influence stock investment decisions of individual investors? (Evidences from Saudi Stock Market). *American International Journal of Contemporary Research*, 6(3), 159-169.
- Bahmani-Oskooee, M., & Bohl, M. T. (2000). German monetary unification and the stability of the German M3 money demand function. *Economics Letters*, 66(2), 203-208. [https://doi.org/10.1016/S0165-1765\(99\)00223-2](https://doi.org/10.1016/S0165-1765(99)00223-2)
- Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(2), 149-163. <https://doi.org/10.1111/j.2517-6161.1975.tb01532.x>
- Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications*, 112, 353-371.
- Churchill, S. A., Inekwe, J., Ivanovski, K., & Smyth, R. (2019). Dynamics of oil price, precious metal prices and the exchange rate in the long-run. *Energy Economics*, 84, Article 104508. <https://doi.org/10.1016/j.eneco.2019.104508>
- Delgado, N. A. B., Delgado, E. B., & Saucedo, E. (2018). The relationship between oil prices, the stock market and the exchange rate: Evidence from Mexico. *The North American Journal of Economics and Finance*, 45, 266-275. <https://doi.org/10.1016/j.najef.2018.03.006>
- Drehmann, M., & Juselius, M. (2014). Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3), 759-780. <https://doi.org/10.1016/j.ijforecast.2013.10.002>
- Drucker, P. F. (1978). Managing the third sector. *The Wall Street Journal*, 3, 26-26.
- Elliott, G., & Timmermann, A. (2016). Forecasting in economics and finance. *Annual Review of Economics*, 8, 81-110. <https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-080315-015346>

- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Habiba, U. E., & Zhang, W. (2020). The dynamics of volatility spillovers between oil prices and stock market returns at the sector level and hedging strategies: Evidence from Pakistan. *Environmental Science and Pollution Research*, 27, 30706-30715. <https://doi.org/10.1007/s11356-020-09351-6>
- Hasan, M. A., & Zaman, A. (2017). Volatility nexus between stock market and macro-economic variables in Bangladesh: An extended GARCH approach. *Scientific Annals of Economics and Business*, 64(2), 233-243.
- Hou, Q., Bing, Z. T., Hu, C., Li, M. Y., Yang, K. H., Mo, Z., Xie, X. W., Liao, J. L., Lu, Y., Horie, S., & Lou, M. W. (2018). RankProd combined with genetic algorithm optimized artificial neural network establishes a diagnostic and prognostic prediction model that revealed C1QTNF3 as a biomarker for prostate cancer. *EBioMedicine*, 32, 234-244.
- Ibrahim, M. H. (2015) Oil and food prices in Malaysia: A nonlinear ARDL analysis. *Agricultural and Food Economics*, 3, Article 2. <https://doi.org/10.1186/s40100-014-0020-3>.
- Khan, M. M., & Yousuf, A. S. (2013). *Macroeconomic forces and stock prices: Evidence from the Bangladesh stock market*. Munich Personal RePEc Archive.
- Kolapo, F. T., & Adaramola, A. O. (2012). The impact of the Nigerian capital market on economic growth (1990-2010). *International Journal of Developing Societies*, 1(1), 11-19.
- Kumar, G., & Misra, A. K. (2019). Liquidity-adjusted CAPM - An empirical analysis on Indian stock market. *Cogent Economics & Finance*, 7(1), Article 1573471. <https://doi.org/10.1080/23322039.2019.1573471>
- Lacheheb, M., & Sirag, A. (2019). Oil price and inflation in Algeria: A nonlinear ARDL approach. *The Quarterly Review of Economics and Finance*, 73, 217-222. <https://doi.org/10.1016/j.qref.2018.12.003>
- Musallam, S. R. (2018). Exploring the relationship between financial ratios and market stock returns. *Eurasian Journal of Business and Economics*, 11(21), 101-116.
- Okere, K. I., Muoneke, O. B., & Onuoha, F. C. (2021). Symmetric and asymmetric effects of crude oil price and exchange rate on stock market performance in Nigeria: Evidence from multiple structural break and NARDL analysis. *The Journal of International Trade & Economic Development*, 30(6), 1-27. <https://doi.org/10.1080/09638199.2021.1918223>
- Pan, X., Uddin, M. K., Han, C., & Pan, X. (2019). Dynamics of financial development, trade openness, technological innovation and energy intensity: Evidence from Bangladesh. *Energy*, 171, 456-464. <https://doi.org/10.1016/j.energy.2018.12.200>
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4), 2162-2172. <https://doi.org/10.1016/j.eswa.2014.10.031>
- Rahman, A. S. A., Abdul-Rahman, S., & Mutalib, S. (2017, November). Mining textual terms for stock market prediction analysis using financial news. In A. Mohamed, M. Berry & B. Yap (Eds.), *Soft Computing in Data Science* (pp. 293-305). Springer. https://doi.org/10.1007/978-981-10-7242-0_25

- Raza, N., Shahzad, S. J. H., Tiwari, A. K., & Shahbaz, M. (2016). Asymmetric impact of gold, oil prices and their volatilities on stock prices of emerging markets. *Resources Policy, 49*, 290-301. <https://doi.org/10.1016/j.resourpol.2016.06.011>
- Sarwar, S., & Wasim, H. (2016) Oil prices and Asian emerging stock markets: Pakistan and Bangladesh. *European Journal of Economic Studies, 2*(16), 353-357. <https://doi.org/10.13187/es.2016.16.353>
- Sheikh, U. A., Asad, M., Ahmed, Z., & Mukhtar, U. (2020). Asymmetrical relationship between oil prices, gold prices, exchange rate, and stock prices during global financial crisis 2008: Evidence from Pakistan. *Cogent Economics & Finance, 8*(1), Article 1757802. <https://doi.org/10.1080/23322039.2020.1757802>
- Shin, Y., Yu, B., & Greenwood-Nimmo, M. (2014). Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In R. Sickles & W. Horrace (Eds.), *Festschrift in honor of Peter Schmidt* (pp. 281-314). Springer. https://doi.org/10.1007/978-1-4899-8008-3_9
- Singhal, S., Choudhary, S., & Biswal, P. C. (2019). Return and volatility linkages among international crude oil price, gold price, exchange rate and stock markets: Evidence from Mexico. *Resources Policy, 60*, 255-261. <https://doi.org/10.1016/j.resourpol.2019.01.004>
- Tuyon, J., & Ahmad, Z. (2016). Behavioural finance perspectives on Malaysian stock market efficiency. *Borsa Istanbul Review, 16*(1), 43-61. <https://doi.org/10.1016/j.bir.2016.01.001>
- Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanprabu, S. K. (2020). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management, 50*, 538-556. <https://doi.org/10.1016/j.ijinfomgt.2018.12.001>

Organ Dose and Radiation Exposure Risk: A Study Comparing Radiation Dose Using Two Software Packages

Abdullah Ali M Asiri

Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia

ABSTRACT

With the rapid development of X-ray equipment, assessing the patient's radiation dose has become an important issue. This study uses DoseCal and PCXMC software to estimate the effective dose (ED) for 510 adult patients undergoing abdomen anteroposterior (AP) and chest anteroposterior/posteroanterior (AP/PA) X-ray examinations in Najran, Saudi Arabia. This study reported our experience with DoseCal and PCXMC software in calculating the ED and organ doses in abdomen and chest X-ray diagnostics. The mean ED values calculated using DoseCal were 0.051, 0.115, and 0.045 mSv for Abdomen AP, chest AP, and PA, respectively. Further, the mean ED calculated using PCXMC is 0.062, 0.132, and 0.047 mSv for Abdomen AP, chest AP, and PA, respectively. The dose results calculated by PCXMC were higher than DoseCal; however, we strongly recommended the dose surveyors utilize PCXMC because it uses the most recent tissue weighting factors (WTs) and offers a risk calculation.

Keywords: Abdomen, chest, DoseCal, effective dose, PCXMC

INTRODUCTION

Individuals are exposed to radiation from various sources, mostly natural and some that are artificial. These sources may

include nuclear power plants and diagnostic or therapeutic medical applications. Radiography equipment is one of the human-made sources. As the number of diagnostic radiography applications grows, the health risks increase, and it becomes necessary to know the accurate doses received by patients undergoing such examinations.

The radiation dose that the patient absorbs during X-ray examinations is commonly evaluated using entrance skin

ARTICLE INFO

Article history:

Received: 21 September 2021

Accepted: 07 March 2022

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.12>

E-mail address:

aaalasm@nu.edu.sa

dose (ESD) or effective dose (ED). An indirect method via X-ray output factors can be used to calculate the patient doses. Most dose surveyors have used software to perform the computational process. This software has become an important tool for reducing direct or in-vivo measurements for patients undergoing X-ray examinations. The radiation doses computed using these computer software packages may differ for the same patient model. Comparing estimated dose values to reference dose levels can help with dose optimization and dose audit in diagnostic protocols. Several publications have reported wide variations in patient doses arising from specific X-ray examinations at different places (Abdelhalim, 2010; Alsayyari et al., 2017; Mettler et al., 2008; Saeed, 2017; Osei & Darko, 2013; Osman et al., 2013; Rubai et al., 2018; Taha et al., 2016). The wide variations in patient dose in most of these studies may be attributed to the clinical condition, examination technique, radiographer skill, peak tube voltage (kVp), exposure current-time product (mAs), and focus-to-skin distance (FSD).

In Saudi Arabia, studies on radiation dose in routine X-ray examinations are scanty; therefore, in the present research, the authors were interested in evaluating the ED, comparing it with others, and obtaining risk factors that will be particularly useful for clinicians. This study uses DoseCal and PCXMC software to evaluate the EDs in different organs for abdomen anteroposterior (AP) and chest anteroposterior/posteroanterior (AP/PA) X-ray examinations in Najran University Hospital (NUH), Najran, Saudi Arabia. The additional aims include comparing the two software and estimating the radiogenic risk to patients during the abdomen and chest examinations. DoseCal software was produced by the Radiological Protection Centre of Saint George's Hospital, London. In contrast, PCXMC is a Monte Carlo tool kit developed by Radiation and Nuclear Safety Authority (STUK), Helsinki, Finland.

In 1998, the program PCXMC was published by Servomaa and Tapiovaara (1998) and later was distributed from the STUK website. PCXMC performs Monte Carlo calculation according to the exposure parameters defined by the user using the hermaphrodite phantom models of Cristy and Eckerman (1987). This phantom family describes adult and pediatric patients and includes several pediatric ages such as newborns, 1, 5, 10, and 15-year-old. In addition, this software calculates the ED based on both tissue weighting factors (WTs) of the International Commission on Radiological Protection (ICRP) publication 103 (ICRP, 2007) and ICRP publication 60 (ICRP, 1991).

Later, Kyriou et al. (2000) published the program DoseCal which calculates the ED based on ICRP60 (ICRP, 1991). DoseCal calculates the organ and tissue doses for adult and pediatric MIRD5 phantom using the conversion coefficients reported by Jones and Wall (1985) and Hart et al. (1994). DoseCal has become very popular because it is easy to use and gives quick results. In addition, DoseCal could be obtained free of charge and used to assess the ESDs, dose area product (DAP), EDs, and organ and tissue absorbed doses (Ds) for X-ray examinations according to the exposure conditions defined by the user.

As opposed to DoseCal, which calculates the ED based on the output measurement and calculation of ESD, PCXMC gives the user the option to input the dose in different quantities such as incident air kerma (in mGy), DAP (mGy.cm²), entrance exposure (mR), exposure area product (R.cm²). The factors provided by the ICRP 103 have been modified slightly from those in ICRP 60, and this certainly will influence the results of ED calculated by DoseCal. The use of different sets of radiation weighting factors in this study could illustrate the range of disparity in the results of ED for patients undergoing chest and abdomen examinations. To the best of our knowledge, there is no comparison between PCXMC and DoseCal software; however, previous studies reported that each software is reliable in ED measurements and presents a few errors (Azevedo et al., 2006; Servomaa & Tapiovaara, 1998).

MATERIALS AND METHODS

Measurements and Calculation with DoseCal

This work was conducted in NUH using two X-rays units (Toshiba DRX-3724HD and GE AL01F). Both X-ray units had a 3 mm Aluminum filter, and all the studies were performed with grids. Data were collected on patient doses for 11 months. The tube outputs (OP) of the two X-ray units were measured at 80 kV at 1 m normalized to 10 mAs using an Unfors Xi dosimeter with an accuracy better than 5% (Unfors Inc., Billdal, Sweden). The OP data were entered in the DoseCal software version 2.31 using a computer equipped with a Windows-XP.

Once the patient's age, sex, weight, and exposure parameters such as kVp, the mAs, the FSD, and filtration are known and entered into the DoseCal software, the ESD can be calculated from Equation 1 as reported by Davies et al. (1997).

$$ESD = OP \left(\frac{kVp}{80} \right)^2 \times mAs \times \left(\frac{100}{FSD} \right)^2 \times BSF \quad (1)$$

where BSF is the backscatter factor calculated automatically by the DoseCal software.

Kyriou et al. (2000) reported that the ED was calculated using Equation 2.

$$ED = ESD \times Cf(D) \quad (2)$$

Where Cf (D) is the conversion factor used to change ESD to ED based on the National Radiological Protection Board (NRPB) tables (Hart et al., 1991).

Simulation with PCXMC

The PCXMC carried out the dose calculation after defining the examination data and performing a Monte Carlo simulation using a PC equipped with a 2.53 GHz processor.

Once the recorded geometrical parameters of the abdomen and chest examinations such as X-ray field limits on the patients and FSD, age, weight, height, maximum keV, and several photons are known and entered into the PCXMC software, the simulation step can be started. The number of photons and the maximum energy used was 20,000 and 150 keV, respectively. The physical processes in the simulation begin with photons being emitted from an isotropic point source and are followed by random interactions with phantom, including Raleigh scattering, Compton scattering, or photoelectric absorption. Finally, the photons' histories generated and calculated the energies deposited in various organs and used for dose calculations.

In the dose calculation step, the X-ray spectrum was defined according to the X-ray tube potential, anode angle, and total filtration used in this study. The ESD calculated by DoseCal, according to the output measurement mentioned previously, was divided by the BSF provided by DoseCal before being entered into the PCXMC. Once the simulation process is complete, the ED, organ doses, and their estimated statistical precision are displayed in PCXMC.

The PCXMC calculated the ED using Equation 3.

$$ED = \frac{1}{2} \sum W_T [H_T(Female) + H_T(Male)] \quad (3)$$

where H_T is the average equivalent dose in a tissue or organ.

For patients undergoing both abdomen and chest X-ray examinations, the model developed by the Committee on the Biological Effects of the Ionizing Radiations (BEIR) was used to assess the risk of exposure-induced cancer death (REIC) (Tapiovaara & Siiskonen, 2008). Once a patient's age, gender, and mortality are known and entered into the PCXMC software, the radiogenic risk can be estimated using Equation 4.

$$REIC = \int_T^{\infty} [\mu_c(t|e, D) - \mu_c(t)] S(t|e, D) dt \quad (4)$$

where μ_c is the mortality rate; t is the age; c is the death cause; e is the exposure; D is the dose; S is the conditional probability, and $\mu_c(t)$ is the background mortality rate. T is equal $e + L$, where L is the latency period in years.

Statistical parameters such as the mean, error percentage, and variation factor, including abdomen AP and chest AP/PA X-ray examinations, have been calculated using SPSS version 14 (SPSS Inc, Chicago, IL).

RESULTS

A total of 510 patients who were referred for abdomen and chest X-ray examinations at NUH were included in this study. Gender distribution shows that 279 (54.7%) were males

while 231 (45.3%) were females. The mean of the anthropometric data and exposure parameters are shown in Table 1. The data outside the brackets in Table 1 represent the mean, and the data inside the brackets represent the minimum and maximum of subjects.

Table 1

Patient anthropometric data and exposure parameters of abdomen and chest radiographic examinations

	Patient age (year)	Gender		Patient weight (kg)	Radiographic data		
		Male	Female		Tube voltage (kVp)	mAs	FSD (cm)
Abdomen AP	42(18-63)	115	95	82(44-105)	64(60-85)	16(11-43)	100 (70-110)
Chest AP	44(19-88)	85	63	67(45-97)	64(50-80)	28(6-160)	100 (70-105)
Chest PA	39(25-76)	79	73	64(42-90)	64(50-85)	18(6-46)	100 (70-105)

The distribution of the mean ED calculated by PCXMC for abdomen AP and chest AP/PA X-ray examinations using W_T of ICRP publication 60 (ICRP, 1991) and ICRP publication 103 (ICRP, 2007) are summarized in Table 2. It was observed that the mean of ED values for chest AP and PA X-ray examinations calculated with ICRP publication 103 (ICRP, 2007) are higher than the values calculated with ICRP publication 60 (ICRP, 1991), with a factor range between 1.2 and 1.3 and lower in abdomen AP with a factor of 7.0.

Table 2

The mean of ED (mSv) calculated in this study using PCXMC

Examinations	Using WT of ICRP 60	Using WT of ICRP 103
Abdomen AP	0.062	0.088
Chest AP	0.132	0.168
Chest PA	0.047	0.056

Table 3

The ED (mSv) calculated in this study using DoseCal and PCXMC

	DoseCal			PCXMC*		
	Abdomen AP	Chest AP	Chest PA	Abdomen AP	Chest AP	Chest PA
Min	0.021	0.073	0.017	0.094	0.078	0.023
Max	0.340	0.852	0.103	0.860	0.950	0.111
Mean	0.051	0.115	0.045	0.062	0.132	0.047
Error (%)	0.40	0.20	0.30	0.30	0.20	0.30
Sample size	210	148	152	210	148	152

*Using WT of ICRP 60 (ICRP, 1991)

Table 4

Comparison between the mean of ED calculated in this study and previously published data

	Abdomen AP	Chest AP	Chest PA
This study using PCXMC*	0.062	0.132	0.047
This study using PCXMC**	0.088	0.168	0.056
This study using DoseCal	0.051	0.115	0.045
B. F. Wall et al. (2011)	0.430	-	0.014
Nahangi and Chaparian (2015)	0.113	-	0.043

*Using WT of ICRP 60 (ICRP, 1991)

**Using WT of ICRP 103 (ICRP, 2007)

Table 5

The D to organs and tissues using DoseCal and PCXMC software for abdomen AP examinations

Variation factor	D error (%)**	D (mGy)**	D (mGy)*	Organ/Tissue
1.000	1.6	0.0124	0.0124	Adrenals
1.100	0.8	0.0011	0.0010	Breast
1.005	2.1	0.1248	0.1254	Gall bladder
1.002	1.2	0.1332	0.1334	Stomach
1.008	1.5	0.1001	0.1009	Small intestine
1.001	1.6	0.1262	0.1263	Upper Large intestine
1.000	4.3	0.0806	0.0806	Lower Large intestine
1.068	0.5	0.0029	0.0031	Heart
1.006	0.6	0.0182	0.0183	Kidneys
13.000	1.7	0.0013	0.0001	Thyroid
1.000	0.4	0.0714	0.0714	Liver
1.294	0.2	0.0017	0.0022	Lung
1.005	22.2	0.0755	0.0751	Ovaries
1.006	1.2	0.0499	0.0502	Pancreas
1.360	3.2	0.0423	0.0311	Skin
1.007	2.1	0.0276	0.0278	Spleen
1.000	2.1	0.0138	0.0138	Testicles
1.000	0.6	0.0003	0.0003	Thymus
1.011	1.1	0.1810	0.1791	Urinary bladder
1.000	1.6	0.1018	0.1018	Uterus
3.233	1.0	0.0097	0.0031	Oesophagus
****	****	***	0.0573	Trunk region
****	****	***	0.0019	Leg region
1.323	0.1	0.0221	0.0167	Skeleton
1.063	0.1	0.0119	0.0112	Active (red) marrow

* using DoseCal software

** using PCXMC software

*** Organ/Tissue not available in phantom

**** indicates data not available.

For clarification, Table 3 presents a statistical distribution of the ED values of abdomen AP and chest AP/PA using DoseCal and PCXMC software. The mean values of EDs calculated by DoseCal for abdomen AP and chest AP/PA X-ray examinations were lower than those calculated by PCXMC, with a factor range between 1.2 and 1.0.

Table 6
The D to organs and tissues using DoseCal and PCXMC software for chest PA examinations

Variation factor	D error (%)**	D (mGy)**	D(mGy)*	Organ/Tissue
1.019	1.5	0.0463	0.0472	Adrenals
1.333	3.4	0.0003	0.0004	Brain
1.002	0.8	0.4801	0.4792	Breast
****	****	***	0.0021	Eye lenses
1.013	1.2	0.0843	0.0854	Gall bladder
1.001	2.8	0.2181	0.2183	Stomach
1.000	5.7	0.0043	0.0043	Small intestine
1.015	1.9	0.0067	0.0068	Upper Large intestine
1.000	4.9	0.0008	0.0008	Lower Large intestine
1.002	0.8	0.2978	0.2985	Heart
1.025	0.6	0.0241	0.0247	Kidneys
13.254	3.2	0.3075	0.0232	Thyroid
1.004	2.4	0.1852	0.1859	Liver
1.008	0.6	0.2202	0.2219	Lung
1.167	21.5	0.0006	0.0007	Ovaries
1.014	1.1	0.1211	0.1228	Pancreas
1.092	1.4	0.0694	0.0758	Skin
1.044	1.1	0.0710	0.0741	Spleen
0.000	0.9	0.0007	0.0000	Testicles
1.001	3.5	0.4886	0.4892	Thymus
1.500	1.5	0.0003	0.0002	Urinary bladder
2.000	0.3	0.0002	0.0004	Uterus
1.134	1.1	0.0852	0.0966	Oesophagus
****	****	***	0.0581	Trunk region
****	****	***	0.0000	Leg region
2.026	0.1	0.2233	0.1102	Skeleton
1.081	0.3	0.0493	0.0456	Active (red) marrow

* using DoseCal software

** using PCXMC software

*** Organ/Tissue not available in phantom

**** indicates data not available.

Table 7
The D to organs and tissues using DoseCal and PCXMC software for chest AP examinations

Variation factor	D error (%)**	D (mGy)**	D (mGy)*	Organ/Tissue
1.003	1.4	0.1829	0.1835	Adrenals
1.000	2.9	0.0007	0.0007	Brain
1.042	0.7	0.0273	0.0262	Breast
****	****	***	0.0003	Eye lenses
1.085	1.2	0.0201	0.0218	Gall bladder
1.037	2.7	0.0298	0.0309	Stomach
1.048	5.8	0.0022	0.0021	Small intestine
1.381	1.2	0.0021	0.0029	Upper Large intestine
1.000	3.2	0.0004	0.0004	Lower Large intestine
1.031	0.5	0.0453	0.0467	Heart
1.007	0.3	0.1282	0.1291	Kidneys
16.316	3.1	0.7032	0.0431	Thyroid
1.019	2.3	0.0683	0.0696	Liver
1.007	0.2	0.1672	0.1683	Lung
1.000	21.1	0.0003	0.0003	Ovaries
1.015	1.3	0.0619	0.0628	Pancreas
1.192	1.1	0.0521	0.0437	Skin
1.004	1.2	0.1563	0.1570	Spleen
1.000	1.3	0.0000	0.0000	Testicles
1.329	3.2	0.0143	0.0190	Thymus
1.000	1.2	0.0000	0.0000	Urinary bladder
1.000	0.1	0.0003	0.0003	Uterus
1.212	1.2	0.0523	0.0626	Oesophagus
****	****	***	0.0792	Trunk region
****	****	***	0.0000	Leg region
1.861	0.2	0.2051	0.1119	Skeleton
1.195	0.4	0.0563	0.0471	Active (red) marrow

* using DoseCal software

** using PCXMC software

*** Organ/Tissue not available in phantom

In Table 4, a comparison is given between the mean EDs obtained in this study and the data reported by Wall et al. (2011) and Nahangi and Chaparian (2015). It was observed that the mean ED values of chest PA in this study are comparable to data reported by Nahangi and Chaparian (2015). In contrast, mean EDs of abdomen PA calculated by PCXMC and Dosecal are lower than data reported by Wall et al. (2011), with a factor range between 4.9 to 8.4.

Tables 5-7 compare the average D values for organs and tissues between DoseCal and PCXMC software for the abdomen AP, chest PA, and AP examinations. PCXMC

can calculate doses for 45 organs and tissue; however, only 27 organs and tissues available in DoseCal or at X-ray exposure risk are presented in Tables 5-7. The gall bladder, stomach, upper large intestine, and urinary bladder in the abdomen examinations and the breast and thymus in chest examinations receive the highest dose. The mean values of the REIC (per million) calculated by PCXMC for abdomen and chest examinations are shown in Figure 1 for male and female patients.

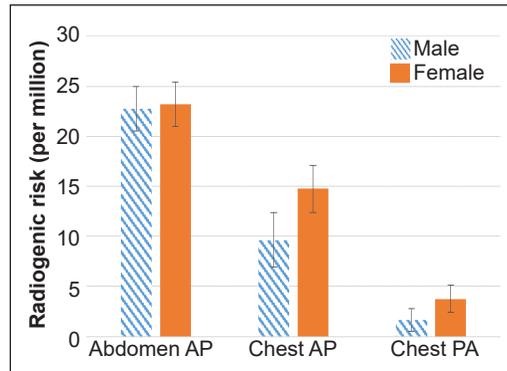


Figure 1. The mean REIC (per million) from abdomen and chest examinations of male and female patients

DISCUSSION

This article evaluated the ED, organ doses, and radiogenic risks for patients undergoing chest and abdominal diagnostic X-ray examinations using DoseCal and PCXMC software. The estimated EDs ranged from 0.045–0.168 mSv, and the median weight and age for all patients were 76 kg and 42 years, respectively. The variation in the mean of ED (mSv) presented in Table 2 could be attributed to the different tissue-weighting factors used in the ICRP publication 60 (ICRP, 1991) and ICRP publication 103 (ICRP, 2007).

The ED results shown in Table 4 for abdomen PA are lower than the data reported by Wall et al. (2011). Likewise, several researchers (Mettler et al., 2008; Osei & Darko, 2013; Taha et al., 2016), who have carried out radiation dose surveys, have also reported variations in patient dose arising from abdomen X-ray examinations. It could be attributed to several reasons: examination technique, the technologist's skill, clinical condition, equipment performance, film–screen combination speed, mAs, kVp, and FSD. For example, Nahangi and Chaparian (2015) pointed out that the mean kV and mAs used for abdomen examinations are 67 kV and 55 mAs, which could explain the lower value of the ED obtained in this study.

In Tables 5-7, the average D values for most organs and tissues between DoseCal and PCXMC software vary up to a factor range between 1.0 and 3.2. In contrast, some organs, such as the thyroid, showed a significant variation. For example, the D in thyroid for chest AP examinations varies by a factor up to 16.3, and this could be attributed to the difference in the positioning of the field size or variation in the location of thyroid in the MIRD5 phantom and phantom models of Cristy and Eckerman (1987). In addition, some modifications carried out in the phantom models, such as the correction of depths of some organs, have been made in the PCXMC. These modifications have been described elsewhere (Wall et al., 2011) and could affect the results of the D of the organs calculated by each software.

In this study, the radiogenic risks in male and female patients were expressed as values of REIC per million. It can help the physician justify the X-ray examinations and compare them with other risks. The calculated REIC for abdomen AP and chest PA radiography showed a relatively good agreement with the findings of the Nahangi and Chaparian (2015). It can be observed that the mean of REIC values for the abdomen is higher than chest AP/PA, and this could be attributed to the higher radiosensitivity of the organs in the abdomen. As shown in Figure 1, the difference in the REIC between males and females was statistically correlated for chest examinations, similar to Nahangi and Chaparian's (2015) findings. This correlation in the chest examinations could be explained due to the difference in radiosensitivity of some organs, such as breasts which are different in patients according to gender.

The major limitation of this study was the modicum number of the X-ray examinations included in this study. Essentially, a variety of the x-ray examinations used could affect the findings of the variation between DoseCal and PCXMC software. However, the comparison performed in this work showed that the computed doses for most of the organs relatively correlate well in both software.

CONCLUSION

This study investigated the patient doses using DoseCal and PCXMC software for patients undergoing abdomen and chest diagnostic radiology examinations only. It is necessary to reproduce bigger reliability in DoseCal and PCXMC by involving many X-ray investigations. The conclusion was that both software produced, to some extent, similar results, except one generated ED based on ICRP 60 and the other on ICRP 103. As the DoseCal used old data of ICRP 60, it may result in inaccurate risk factor calculations unless it was corrected by the new W_T recommended by ICRP 103. Calculating ED with old factors may give a difference of up to 30% and 21% for abdomen and chest examination respectively. We highly recommend that the dose surveyors utilize PCXMC because it uses the most recent W_T s and provides a risk calculation.

The REIC values obtained in this study from abdomen and chest X-ray examinations for male and female patients can be an indicator helping physicians to judge radiation risks and encourage them to be concerned about knowing the REIC values.

ACKNOWLEDGMENT

The authors want to thank the co-operation of the NUH Directorate, the radiologists, Prof. M K Saeed for his valuable comments, and Mrs. Rose-Ann Telebangco Quilaton from the Radiology Department for participating in this study.

REFERENCES

- Abdelhalim, M. A. K. (2010). Patient dose levels for seven different radiographic examination types. *Saudi Journal of Biological Sciences*, 17(2), 115-118. <https://doi.org/10.1016/J.SJBS.2009.12.013>
- Alsayyari, A. A. S., Omer, M. A. A., Alkhidir, N. A. M., & Auod, A. S. A. (2017). Study of common requested radiographs and relative exposure dose in Qassim State. *International Journal of Medical Imaging*, 5(2), 14-18. <https://doi.org/10.11648/J.IJMI.20170502.12>
- Azevedo, A. C., Osibote, O. A., & Boechat, M. C. B. (2006). Paediatric X-ray examinations in Rio de Janeiro. *Physics in Medicine and Biology*, 51(15), Article 3723. <https://doi.org/10.1088/0031-9155/51/15/008>
- Cristy, M., & Eckerman, K. F. (1987). *Specific absorbed fractions of energy at various ages from internal photon sources: 7, Adult male*. Oak Ridge National Lab. https://inis.iaea.org/Search/search.aspx?orig_q=RN:19012902
- Davies, M., McCallum, H., White, G., Brown, J., & Helm, M. (1997). Patient dose audit in diagnostic radiography using custom designed software. *Radiography*, 3(1), 17-25. [https://doi.org/10.1016/S1078-8174\(97\)80021-1](https://doi.org/10.1016/S1078-8174(97)80021-1)
- Hart, D., Jones, D. C., & Wall, B. F. (1991). *Normalized organ doses for medical X-ray examinations calculated using Monte Carlo techniques. NRPB-SR262*. National Radiological Protection Board.
- Hart, D., Jones, D. G., & Wall, B. F. (1994). *Estimation of effective dose in diagnostic radiology from entrance surface dose and dose-area product measurement (NRPB R-262)*. National Radiological Protection Board.
- ICRP. (1991). 1990 Recommendations of the International Commission on Radiological Protection. *Annals of the ICRP*. 21(1-3), 1-201. <https://pubmed.ncbi.nlm.nih.gov/2053748/>
- ICRP. (2007). *The 2007 Recommendations of the International Commission on Radiological Protection. ICRP publication 103. Annals of the ICRP*, 37(2-4), 9-34. <https://doi.org/10.1016/J.ICRP.2007.10.003>
- Jones, D. G., & Wall, B. F. (1985). *Organ doses from medical X-ray examinations calculated using Monte Carlo techniques (NRPB-R-186)*. National Radiological Protection Board.
- Kyriou, J. C., Newey, V., & Fitzgerald, M. C. (2000). *Patient doses in diagnostic radiology at the touch of a button*. The Radiological Protection Center, St. George's Hospital.
- Mettler, F. A., Huda, W. J., Yoshizumi, T. T., & Mahesh, M. (2008). Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology*, 248(1), 254-263. <https://doi.org/10.1148/RADIOL.2481071451>
- Nahangi, H., & Chaparian, A. (2015). Assessment of radiation risk to pediatric patients undergoing conventional X-ray examinations. *Radioprotection*, 50(1), 19-25. <https://doi.org/10.1051/RADIOPRO/2014023>
- Osei, E. K., & Darko, J. (2013). A survey of organ equivalent and effective doses from diagnostic radiology procedures. *ISRN Radiology*, 2013, Article 204346. <https://doi.org/10.5402/2013/204346>
- Osman, H., Elzaki, A., Elsamani, M., Alzaeidi, J., Sharif, K., & Elmorsy, A. (2013). Paediatric radiation dose from routine X-ray examination: A hospital based study, Taif pediatric hospital. *Scholars Journal of Applied Medical Sciences*, 5, 511-515.
- Rubai, S. S., Rahman, M. S., Purohit, S., Patwary, M. K. A., Meaze, A. M. H., & Mamun, A. A. (2018). Measurements of entrance surface dose and effective dose of patients in diagnostic radiography.

- Biomedical Journal of Scientific & Technical Research*, 12(1), 8924-8928. <https://doi.org/10.26717/BJSTR.2018.12.002186>
- Saeed, M. K. (2017). Dose measurement using Gafchromic film for patients undergoing interventional cardiology procedures. *Radiation Protection Dosimetry*, 174(1), 109-112. <https://doi.org/10.1093/RPD/NCW082>
- Servomaa, A., & Tapiovaara, M. (1998). Organ dose calculation in medical X ray examinations by the program PCXMC. *Radiation Protection Dosimetry*, 80(1-3), 213-219. <https://doi.org/10.1093/OXFORDJOURNALS.RPD.A032509>
- Taha, M. T., Kutbi, R. A., & Allehyani, S. H. (2016). Effective dose evaluation for chest and abdomen X-ray examinations. *International Journal of Science and Research (IJSR)*, 5(3), 420-422. <https://doi.org/10.21275/V5I3.NOV161884>
- Tapiovaara, M., & Siiskonen, T. (2008). *PCXMC : A Monte Carlo program for calculating patient doses in medical X-ray examinations* (STUK-A231). Finnish Centre for Radiation and Nuclear Safety.
- Wall, B. F., Haylock, R., Jansen, J. T. M., Hillier, M. C., Hart, D., & Shrimpton, P. C. (2011). *Radiation Risks from Medical X-ray Examinations as a Function of the Age and Sex of the Patient* (HPA-CRCE-028). Health Protection Agency Centre for Radiation, Chemical and Environmental Hazards.

Effective Emergency Management: Scrutinizing the Malaysia Lead Responding Agency Planning and Information Management Approach During Disaster Exercise

Khairilmizal Samsudin^{1*}, Fatin Najihah Ghazali¹, Nur Hannani Abdul Ghani¹, Mohamad Fahmi Hussin², Ainul Husna Kamarudin³ and Haikal Kasri⁴

¹*School of Health Sciences, Health Campus, Universiti Sains Malaysia, 16150 USM, Kubang Kerian, Kelantan, Malaysia*

²*Faculty of Electrical Engineering Complex, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

³*Institute of Medical Science Technology, Universiti Kuala Lumpur, 43000 UniKL, Kajang, Selangor, Malaysia*

⁴*Jabatan Bomba Dan Penyelamat Malaysia, 62250 Putrajaya, Selangor, Malaysia*

ABSTRACT

The impact of a disaster is closely related to the response and recovery of the lead responding agency in managing the disaster effectively. Malaysia has established the MNSC20 policies for managing disasters. Unfortunately, the implementation of MNSC20 can only be evaluated through lead responding agency disaster exercises. Each responding agency's responsibility is to identify and record the information as required to ensure effective emergency management; however, problems in terms of inaccurate and/or incomplete information may be many challenges that need to be addressed. Hence, this paper aims to propose several areas for improvement for planning and information management elements

in effective emergency management theory by identifying its challenges through the FRDM, which may benefit Malaysia's overall emergency management process. Four disaster exercises at the state and federal levels were observed using a structured checklist. The result indicates challenges were found in terms of proactivity in gaining information, disseminating critical information during emergencies, developing a specific incident action plan, and media management. A good area of improvement is

ARTICLE INFO

Article history:

Received: 21 September 2021

Accepted: 02 March 2022

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.13>

E-mail addresses:

khairilmizal@usm.my (Khairilmizal Samsudin)
najihahfatin@student.usm.my (Fatin Najihah Ghazali)
hannanighani@gmail.com (Nur Hannani Abdul Ghani)
fahmi478@uitm.edu.my (Mohamad Fahmi Hussin)
ainulhusna@unikl.edu.my (Ainul Husna Kamarudin)
haikal.kasri@bomba.gov.my (Haikal Kasri)

* Corresponding author

suggested based on the evaluation. It is hoped that this paper could contribute to developing better disaster management by the lead responding agency in Malaysia.

Keywords: Emergency exercise, disaster management, information management, planning

INTRODUCTION

For successful implementation, disaster management should encompass vulnerability assessment, planning (incident action plan), training or exercise, logistic capability, and financial capacity, which are the opposite scopes of preventing disaster (Bullock et al., 2020; Sandler & Schwab, 2021). As an emergency happens and the severity has caused a disaster, minimizing the losses from the direct impact of the disaster is critical, hence indicating the importance of the response and early recovery phases in the process cycle of disaster management (Bullock et al., 2020; Samsudin et al., 2018). The process cycle of disaster management includes the overall response in reducing the losses during the disaster and possible losses that arise from the immediate disaster impact.

It is also important to understand that the overall response process includes the early recovery phase (Samsudin et al., 2016c). Single or multi-entity command structure during the management of emergency, management of planning and information coming in and going out from the emergency site, matrix of communication and notifications, current situational awareness of single or multiple situations at the site, and capacity and capabilities in managing emergencies are five of the main elements in managing emergency effectively (Kamarudin et al., 2016; Hussin et al., 2018; Samsudin et al., 2016c). In an emergency, each responding entity's responsibility is to identify and record the information as required appropriately. However, Bharosa et al. (2010), Jiang and Yuan (2019), and Waring et al. (2020) mentioned that problems regarding inaccurate and/or incomplete information between entities are not to be neglected due to possible additional losses.

Hence, Disaster Planning and Information Management will affect the outcome of the disaster (Kamarudin et al., 2016; Jiang & Yuan, 2019; Waring et al., 2020). These problems include the differences in terminology used in information sharing, inability to understand other agencies' information, lack of information management, and lack of an appropriate common operational picture as the information management limiting factor during disaster management (Lestari et al., 2019; Salmon et al., 2011; Samsudin et al., 2020d). The disaster management environment in Malaysia is outlined in MNSC20, which mentions that the response and early recovery phases shall be led by specific lead responding agencies (MNSC, 2012). For physical-based hazards such as floods, fire, transportation accidents, chemical releases, and more, it shall be led by the Fire and Rescue Department Malaysia (FRDM), hence the main source of data later for this paper (Samsudin et al., 2018; MNSC, 2012).

Experts are crucial in providing views, analyses, and solutions during disasters, as presented in earlier studies towards FRDM personnel (Samsudin et al., 2016b). Moreover, data and information need to be analyzed objectively to provide the most effective ways of managing emergencies, indicating the importance of experts in assisting the responding agency in the overall and specific planning and information during the response and early recovery phases of disasters. Unfortunately, to the authors' discernment, no studies have been made on planning and information management in effective emergency management towards lead responding agencies in Malaysia. Thus, this paper aims to use six disaster exercises at the federal and district level to propose several areas for improving the planning and information management elements of an effective emergency management theory that could benefit Malaysia's overall emergency management process.

METHODOLOGY

As part of a larger disaster management research study in Malaysia, the methodology used has been validated in previous publications (Hussin et al., 2018; Samsudin et al., 2017a; Samsudin et al., 2017b). Two methodologies were used in this paper: the systematic observation and survey questionnaire towards the subject matter experts. Both methodologies have been validated in previous studies and published (Samsudin et al., 2020c; Samsudin et al., 2020d) as part of the larger research scope. The subject matter experts for this study are the officers from the lead responding agencies responsible for the decision-making process during emergencies (Fire and Rescue Department Malaysia) (Samsudin et al., 2020a; MNSC, 2012). The systematic observation was conducted during disaster exercises, namely the EXSTORM. As with any federal level exercise conducted in the country, EXSTORM is a bi-annual disaster exercise effort outlined by the Fire and Rescue Department of Malaysia (FRDM) originally to train the FRDM Special Tactical Operation and Rescue Team of Malaysia (STORM). Due to the dynamic scenario and the exercise scale, EXSTORM has gained cooperation from primary and secondary agencies in Malaysia, as outlined in MNSC20 (MNSC, 2012). Six EXSTORM exercises were observed during this research. EXSTORM is selected as one of the primary sources of data for this study due to several reasons:

- based on Malaysia's disaster management policy, the disaster level will be declared as state or federal;
- the emergency drill or exercise provides excellent opportunities for observation;
- the exercise is orchestrated with multiple events and casualties so that real situation scenarios and worst-case scenarios.

The exercise was orchestrated to involve multi-agency to increase the coordination and cooperation of multi-agencies in Malaysia. The first methodology is the qualitative observation of the disaster management process as published in earlier studies (Samsudin

et al., 2020c; Samsudin et al., 2020d). Observing a real disaster might be challenging compared to disaster exercises that can be done more frequently (Samsudin et al., 2017b). Furthermore, the exercises can provide researchers with a rich data source as EXSTORM is conducted bi-annually and provides excellent opportunities for researchers (Jiang & Yuan, 2019; Lestari et al., 2020). Even though the scenario, exercise venue, and time are different for each exercise to ensure the best possible near to real-life scenario, it is negligible as the research concentrates on the emergency management process. Hence any tactical actions by responders are not part of the collected data. A structured observations checklist aided the observation process of all EXSTORMS exercises to ensure systematic data collection. This validated checklist was developed via document review methods involving procedures, reports, guidelines from Malaysia lead responding agencies, and international standards containing an element of effective emergency management and ensuring the reliability of the results (Samsudin et al., 2020b; Samsudin et al., 2020c; Samsudin et al., 2020d). It is validated through the content and face validity method (Hussin et al., 2012). Hence the following three validation approaches were used: the verifications from lead agency officers at the site who act as exercise observers and controllers, exercise logs recorded by both exercise controller and the observed exercise participants, and finally, the emergency exercise reports by FRDM published after each EXSTORM exercise combining views from all participating entities (Samsudin et al., 2017b; Samsudin et al., 2016c; Samsudin et al., 2020d).

The second methodology involves quantitative data extracted from questionnaires published in earlier studies (Samsudin et al., 2020c; Samsudin et al., 2020d). It is a survey where the subject matter experts (FRDM officers) were required to answer sets of structured questionnaires to obtain quantitative data on the knowledge and practice of managing emergencies (Samsudin et al., 2020b; Samsudin et al., 2016b; Samsudin et al., 2020d). This survey is conducted via an online platform where the FRDM Headquarters requested participants. All respondents were decision-makers during disasters from districts, states, and federal levels as categorized in MNSC20 (MNSC, 2012). The survey questionnaire is designed to engage the respondent to provide their knowledge and practice in managing operations during disaster situations and their needed support. As mentioned earlier, the respondents for this study are the subject matter experts from FRDM who have the authority to make decisions during disasters. Cronbach Alpha statistical test conducted on the questionnaire indicates a fair value of 0.91 (Samsudin et al., 2020c; Samsudin et al., 2020d). Earlier studies concluded that an estimated total number of 2000 FRDM officers involved in the decision-making process (inclusion criteria), and by referring to the standards population sampling of 95% confidence level with a 5% margin of error, a total number of 323 respondents shall be needed to ensure the reliability of the results (Samsudin et al., 2017b; Samsudin et al., 2020c; Samsudin et al., 2020d). Therefore, it is advantageous for this study as the total recorded respondent exceeded the minimum

requirement, with an additional 26% more respondents totaling 407 views from subject matter experts recorded.

The total respondent numbers are contributed mainly by the implementations of digital surveys using Google Forms, as Google Forms provide the ability to provide users with ease of response and portability of access. Furthermore, some research has recognized Google forms as survey tools (Khan et al., 2021; Samsudin et al., 2020d; Shoaib & Abdullah, 2020; Travis, 2010). Google Form can provide multiple question formats, including checkbox, close-ended, matrix, open-ended and more, record logs of the respondent, compiling the respondent data in providing basic descriptive (Samsudin et al., 2017b; Khan et al., 2021; Shoaib & Abdullah, 2020).

For the specific purpose of this study, other elements of effective emergency management shall not be discussed, and the author shall focus on the elements of planning and information management (Kamarudin et al., 2016; Hussin et al., 2018).

RESULTS AND DISCUSSION

Bharosa et al. (2010), Cruz and Ferenchak (2020), and Sardi and Razak (2019) stated that during disasters, pertinent information is sometimes not identified, causing delays in response operations. Samsudin et al. (2016d) and Samsudin et al. (2020c) supported this, where planning and information management are critical elements in managing disaster effectively and need the expert's assistance. Similar issues were observed during the EXSTORM exercises where pertinent information was not identified, causing delayed planning and response.

Throughout the observation process during the EXSTORM exercise, non-standardized forms are used to manage the whole operation. Most of the forms were developed on sites during the operations, depending on the on-site officer experience. This practice has caused inaccurate and/or incomplete information recorded from the activity. Nevertheless, even if one agency has recorded information, the information gained will then be manually recorded by another agency officer rather than disseminated to every agency on-site. This practice is causing inaccurate and/or incomplete information received by agencies due to the difference in the terms used and delays in disseminating pertinent information such as hazards involved, victims list, and action log. In addition, during the EXSTORM exercises, it is observed that improper documentation has caused lessons from past experiences to be hard to learn and improve, adding to the information dissemination issue (Kurita et al., 2006).

It is widely understood that during emergency response and early recoveries, the priorities of any entities responding to the emergency are to minimize the loss and protect the people, assets, and the environment (Haddow et al., 2017). It can be accomplished by properly implementing appropriate emergency response and planning (Shaluf, 2008).

Experts or professionals should always be involved in any emergency planning process. However, during the exercise, it is observed that planning is mainly based on the experience of the commanding officer rather than referring to any written roles and responsibilities, supporting documentation, or even information gained during disasters. The absence of strategic planning expertise introduces limitations in managing the disaster (Moynihan, 2009; Samsudin et al., 2017a). MNSC20 stated the importance of external entities involved in managing disasters aside from leading responding agencies (MNSC, 2012). It is noted that both the external entity and lead responding agency involved in the EXSTORM exercise is inefficient. Some even missed sharing and obtaining pertinent information regarding the incident involved. In one of the exercises observed, the lead responding agency overlooked the information provided by an external entity participating in the exercise.

Quarantelli (1997) and Waring et al. (2020) stated that continually updating information during a disaster will effectively increase the effectiveness of the planning process in managing disasters. One of the exercises observed a lack of current information feeds, causing multiple incident responses to be conducted in series rather than parallel responses. Therefore, it affects the overall disaster response and recovery when dealing with the incident in series, increasing the duration of the overall operations. It is supported by Cruz and Ferenchak (2020), Reid and Van Niekerk (2008), and Sardi and Razak (2019), whereby the lack of pertinent information available and delivered on time is causing delays in response and recovery efforts. Therefore, information management is considered one of the elements of effective disaster management during the response and early recovery phase (Samsudin et al., 2016b).

Challenges and Improvement

Effective disaster management during the response and early recovery phase consists of five major elements and timeframe parameters (Kamarudin et al., 2016); this point is mentioned by Hussin et al. (2018), which summarized that the most crucial elements are planning and information management.

This paper found five major planning and information management challenges found during the observation of EXSTORM's exercises. First, the information management is based on the officer's experience managing it, causing non-standardized forms used by the responding agencies. Therefore, all planning and information management are based on the responder's experience and expectations. Standardized forms will eventually assist responding entities in managing key and pertinent past, current, and future information. The emergency can be managed more effectively via proper standardized records, including a written incident action plan that plays a significant role in recording the strategic path of the disaster operations (Samsudin et al., 2020b; Samsudin et al., 2020d).

Secondly, the dissemination of information is only done upon request from other entities, causing the strategic planning and abundance of information that came through

the command post to be not properly disseminated. Occasionally during the observations of EXSTORM, other entities manually record all the information gained by FRDM and vice versa. Without proper dissemination platforms, it is proven that records are only kept at the site of the operations before being discarded.

Thirdly, the non-availability of an incident action plan for the overall disaster management adds to the effectiveness of a written incident action plan to be prepared for each minor incident (Madigan, 2017). Recorded and managed incident action plans will eventually strategize the whole operations and eventually will be a pertinent record (IAP reference) for lesson learning (research and development) and training exercises for new responders.

Fourthly, information identifications are based on current needs only without identifying the future possibility of disaster events which can cause a possible unexpected escalation of events contributing to more losses and damages managed (Sardi & Razak, 2019; Waring et al., 2020).

Finally, the media are managed based on officer experience rather than a written guide that should be made available and easily accessible. This challenge is supported by a study indicating that poor media management is part of human factors for effective emergency management (Hussin et al., 2018).

Planning and Information Management

Planning and information management play an important role in managing disasters effectively during the response and recovery phase. The objective of emergency management is to bring the response time as early as possible, and this can be achieved with proper planning and information management (Subramaniam et al., 2012). Proper planning and information management include the time taken for initial size-up and dissemination of size-up information. The survey indicates that 92.1% of respondents conduct an initial size-up within 30 minutes with a maximum of two hours. Regarding the dissemination of the size-up information, the survey indicates that 88.4% of respondents disseminated the information to other agencies involved during the disaster within two hours, with a maximum of six hours. This survey result shows that dissemination of information, especially towards other agencies, is not prioritized, as proven during the EXSTORM's observation.

Size-up from responders will lead to the attainment of additional information regarding the disasters. Surveys result indicates that responder prioritizes their method of acquiring information from victims (82.7%), representatives (81.9%), and witnesses (77.3%) compared to other sources, such as academic references (73.6%), paper data (70.1%), and information technology (69.9%). Survey results support that 82.8% of respondents agree that attained information is recorded. Although 84.7% of respondents state that they need

to know the other agency’s information, only 56% of respondents disseminate information to another agency. At the same time, other respondents distribute the information to another agency when ordered or requested (21.6%) or never disseminate the information to another agency (22.4%). It is proven when only 59.4% of respondents agree that information acquired is accessible to another agency.

Compared with the dissemination of information and accessibility of information within an agency, survey results show that 72.3% of respondents disseminate it internally, and 73.8% agree that this information is readily available. Unfortunately, the survey questionnaire did not indicate the depth of the information that can be accessed internally. Therefore, the information that can be accessed internally is the only summary of the disaster and not the detailed information. Table 1 shows the respondents’ responses to information management.

Table 1
Survey on single agency and multi-agency information management

Information management	Yes (%)	Order/ Request (%)	No (%)
Recorded	82.8	9.4	7.8
Disseminated internally	72.3	11.1	16.6
Accessible Internally	73.8	14.1	12.1
Disseminated to another agency	56	21.6	22.4
Accessible to another agency	59.4	24.5	16.1

The method of information dissemination was also surveyed, and multiple modes of information dissemination can be used to disseminate the same information depending on the type of information. As indicated in Table 2, it is found that for single agency response, the method of information dissemination is mainly through the GIRN (MNSC, 2012) system (81.1%) followed by a face to face communication (80.3%), information technology (73.1%), audio communication (71.2%) and finally through papers (69.3%). The results are different when looking at the response for the multi-agency method of information dissemination, where the priority is given to face to face (80.1%), followed by paper (67.3%), information technology (65.9%), audio (62.6%), and finally GIRN system (59.9%). The survey results indicate that paper documentation is the least used method compared to the face-to-face method in disseminating information, as observed during EXSTROM’s exercises. EXSTORM’s observations supported these findings, where minimum use of paper records was observed. At the same time, although the survey indicates the use of information technology as a method of information dissemination, unfortunately, the responders only use social media to disseminate their information which is also observed during the EXSTORM.

Table 2
Surveys on the method of information dissemination within a single agency and multi-agency

Method of information dissemination	Used (scale 4) and most used (scale 5) by Single Agency (%)	Used (scale 4) to most used (scale 5) by multi-agency (%)
Face to face	80.3	80.1
Papers	69.3	67.3
Information technology	73.1	65.9
Audio	71.2	62.6
GIRN	81.1	59.9

The use of standardized forms in recording disaster planning and information, such as incident action plans, meteorological data, area mapping, and victim information (Samsudin & Hussain, 2016), will eventually assist responders in managing planning and information.

It is important that during the response and the recovery phases of disasters, planning and information be managed, as stated by Kamarudin et al. (2016) and as discussed during the observation of EXSTORM. As presented in Table 3, the survey questions result indicates that 86.1% of responders agree that standardized forms within the agency are used during disasters. However, only 58% of respondents agree on the easy accessibility of the forms during disasters, and 64.9% agree on the easy availability of the forms after a disaster. As discussed earlier, the responders understand that the information recorded on the forms is only the summary of the disaster and not the detailed information. Regarding standardized forms used during a multi-agency response, only 56.1% of respondents stated that standardized forms are used. However, to the researcher’s discernment, there are no standardized forms mentioned in any multi-agency response as there is a Frelack of supporting governance for MNSC20 (Samsudin et al., 2016d).

During the observation of EXSTORM, it is summarized that one of the challenges in managing disaster is near to no existence of standardized forms used either by the single agency or multi-agency response. Hence, the survey results show that 89% of respondents stated the importance (important and very important) for responders to have a digital form to assist them in managing planning and information during the response and early recovery phase of disasters, as illustrated in Figure 1. The importance of having digital forms is supported by descriptive statistics indicating the mean at 4.47 and a standard deviation of 0.75.

Table 3
Survey on standardized forms used in managing disasters during responses and early recovery phase

Standardized forms	Yes (%)	No (%)
Used within agency	86.1	13.9
Used during multi-agency	56.1	43.9
Accessible during disasters	58	42
Records accessible after disasters	64.9	35.1
Accessible to another agency	59.4	16.1

Types of planning and information management that are deemed to be important (scale 4) and very important (scale 5) by respondents include victim’s information (68.6%), hazards involved (72%), workforce availability (69.5%), logistical capability (69.8%), location blueprints, (65.9%), surroundings information (64.3%), communication matrix (60.2%), area maps (55.3%), and meteorological data (54.1%). Other respondents (below 10%) stated either information might be moderately important (scale 3) to not important (scale 1). These results indicate that responders’ planning and information management needs are prioritized as follows:

- What can be seen at the site of the disaster: hazards involved and victim’s information
- Offensive and capability at disasters site: workforce availability and logistical capability
- Community information: surroundings information
- Supporting information: communication matrix, area maps, and meteorological data

It is also observed during EXSTORM that media release by responders was not prepared and was not supported by any documented records. However, as indicated in

Table 4, survey results indicate differently, with 79.4% of responders stating that there are guidelines for media release and the media release is initially prepared (78.6%) and recorded (72.1%). Researchers assume this response is based on what should have been done rather than responders’ practice during the actual situation.

A survey on planning and information management also shows that multi-agency roles and responsibilities are known by other responders (88.6%), and these roles and responsibilities are clear to another agency (82.7%). In addition, 74.9% of respondents know other agency actions, are well informed of supports needed (85.8%), and know what support needs to be given to another agency (87.3%). These statistics are believed to be based on respondents’ understanding of MNSC20, as the document indicates the roles and responsibilities of all agencies involved during disasters, hence showing the above response (Samsudin et al., 2016a; MNSC, 2012). Furthermore, EXSTORM

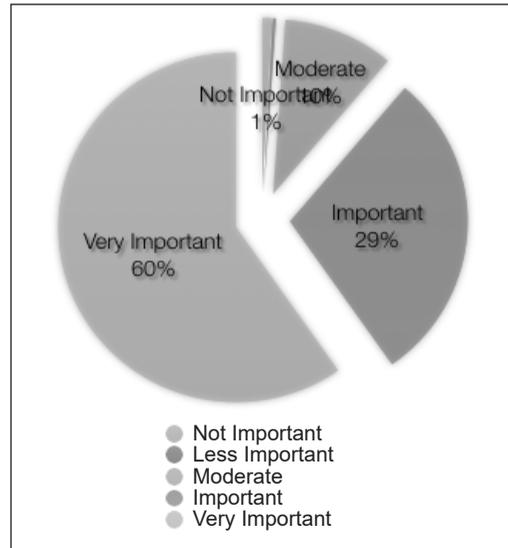


Figure 1. Survey on the importance of digital forms in planning and information management

Table 4
Survey on the media release

Media release	Yes (%)	No, and not sure (%)
Guidelines	79.4	20.6
Prepared	78.6	21.4
Recorded	72.1	27.9

found that some actions by other agencies are not supported by other agencies involved in disasters. These findings suggest that a system that records and manages all agency actions during disasters will eventually inform the other agency of the support they need. Table 3 shows the survey on the multi-agency response. The system should also be developed based on the needs and requirements of the Malaysian disaster management environment (Samsudin, 2018; Samsudin et al., 2020c).

CONCLUSION

An overall observation of the six EXSTROMS indicates challenges in three major areas. First is the heavy reliance of emergency managers (lead responding agency officers) on experience without much support from pertinent information in the analysis and decision-making process during the response and early recovery phase. Secondly, the effectiveness in managing information that comes in and goes out from the designated lead responding agencies command post is lacking. Finally, unavailability of experts in analyzing and managing information hence missing important responses and early recovery process. The authors also concluded that, from the observations of EXSTORM, effective emergency management could be further improved, especially if the elements of planning and information management are further focused on. Firstly, officers in the command post need to have clear written roles and responsibilities, which can be easy to refer to during single or multi-agency emergencies to ensure comprehensive planning and information management. Secondly, it is crucial to have and implement standardized forms and records throughout the emergency management process to ensure effective and comprehensive records and dissemination of information, especially during multi-agency responses. Finally, as each emergency has its unique challenges, the authors believed that third parties subject matter experts should be identified (Samsudin et al., 2016b; MNSC, 2012), especially for identifying and analyzing critical pieces of information and records. The involvement of third parties subject matter experts can also ensure the safety of on-site emergency responders. All the mentioned recommendations in this article were also supported by Samsudin et al. (2020d). Although the novelty of this research is the method and tools used for the observations during disaster exercise, the specific sections of the observation checklist have made it possible to identify the challenges and actively propose the area for improvement. The authors believe that a better disaster management process can be achieved in Malaysia by further improving the planning and information management approach.

ACKNOWLEDGMENTS

The authors want to thank the Universiti Sains Malaysia School of Health Science, UiTM Faculty of Electrical Engineering, and the Fire and Rescue Department Malaysia

for collaborating on this study. This study is supported by the Malaysia Fundamental Research Grant Scheme (FRGS/1/2021/SS02/USM/03/1) and partially USM grant (304/PPSK/6315498).

REFERENCES

- Bharosa, N., Lee, J., & Janssen, M. (2010). Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, 12, 49-65. <https://doi.org/10.1007/s10796-009-9174-z>
- Bullock, J. A., Haddow, G. D., & Coppola, D. P. (2020). *Introduction to emergency management seventh edition*. Elsevier.
- Cruz, M. C., & Ferenchak, N. N. (2020). Emergency response times for fatal motor vehicle crashes, 1975-2017. *Transportation Research Record*, 2674(8), 504-510. <https://doi.org/10.1177/0361198120927698>
- Haddow, G. D., Bullock, J. A., & Coppola, D. P. (2017). *Introduction to emergency management sixth edition*. Elsevier. <https://doi.org/10.1016/B978-0-12-803064-6.00011-1>
- Hussin, M. F., Khairilmizal, S., Basri, S. H., Yassin, A. I. M., Kamarudin, A. H., Salleh, M. K. M., Jusoh, M. H., Sulaiman, A. A., & Saadun, J. (2018). The role of human factors in emergency management: A Malaysian company perspective. *International Journal of Engineering & Technology*, 7(3), 490-493.
- Hussin, M. F., Wang, B., & Hipnie, R. (2012). The reliability and validity of basic offshore safety and emergency training knowledge test. *Journal of King Saud University- Engineering Sciences*, 24(2), 95-105. <https://doi.org/10.1016/j.jksues.2011.05.002>
- Jiang, Y., & Yuan, Y. (2019). Emergency logistics in a large-scale disaster context: Achievements and challenges. *International Journal of Environmental Research and Public Health*, 16(5), Article 779. <https://doi.org/10.3390/ijerph16050779>
- Kamarudin, A. H., Basri, S. H., Hussin, M. F., & Khairilmizal, S. (2016). Effective emergency management: A Malaysian public listed oil and gas company perspective. In *International Medical Science Technology Conference (IMSTC2016)*. UniKL MESTECH.
- Khan, M. A., Kamal, T., Illiyan, A., & Asif, M. (2021). School students' perception and challenges towards online classes during Covid-19 pandemic in India: An econometric analysis. *Sustainability*, 13(9), Article 4786. <https://doi.org/10.3390/su13094786>
- Kurita, T., Nakamura, A., Kodama, M., & Colombage, S. R. N. (2006). Tsunami public awareness and the disaster management system of Sri Lanka. *Disaster Prevention and Management*, 15(1), 92-110. <https://doi.org/10.1108/09653560610654266>
- Lestari, P., Paripurno, E. T., & Nugroho, A. R. B. (2019). Table top exercise disaster communication model in reducing disaster risk. *Jurnal Penelitian Komunikasi*, 22(1), 17-30. <https://doi.org/10.20422/jpk.v22i1.587>
- Lestari, P., Ritonga, R., Ruliana, P., & Barus, C. C. B. (2020). Disaster communication uses field training exercise simulation as an important aspect of disaster risk reduction. *Jurnal Komunikasi: Malaysian Journal of Communication*, 36(1), 166-186. <https://doi.org/10.17576/JKMJC-2020-3601-10>

- Madigan, M. L. (2017). *Hazmat guide for first responder*. CRC Press. <https://doi.org/10.1201/9781315230795>
- MNSC. (2012). *Policy and mechanism of disaster in Malaysia*. Malaysia National Security Council.
- Moynihan, D. P. (2009). The network governance of crisis response: Case studies of incident command systems. *Journal of Public Administration Research and Theory*, 19(4), 895-915. <https://doi.org/10.1093/jopart/mun033>
- Quarantelli, E. L. (1997). Problematical aspects of the information/communication revolution for disaster planning and research: Ten non-technical issues and questions. *Disaster Prevention and Management*, 6(2), 94-106. <https://doi.org/10.1108/09653569710164053>
- Reid, P., & Van Niekerk, D. (2008). A model for a multi-agency response management system (MARMS) for South Africa. *Disaster Prevention and Management*, 17(2), 244-255. <https://doi.org/10.1108/09653560810872541>
- Salmon, P., Stanton, N., Jenkins, D., & Walker, G. (2011). Coordination during multi-agency emergency response: Issues and solutions. *Disaster Prevention and Management*, 20(2), 140-158. <https://doi.org/10.1108/09653561111126085>
- Samsudin, K. (2018). *Database design for fire and rescue department malaysia integrated disaster management system* (Doctoral dissertation). Universiti Teknologi MARA, Malaysia. <https://ir.uitm.edu.my/view/doctype/thesis/>
- Samsudin, K., & Hussain, A. R. (2016). *SMEP 5203: Emergency management*. Centre for Instructional Design and Technology, Open University Malaysia. https://www.academia.edu/29985999/SMEP_5203_Emergency_Management
- Samsudin, K., Hussin, M. F. F., Yassin, A. I. M., Husna, K. A., Sulaiman, A. A., Jusoh, M. H., & Kasri, M. H. (2016a). Evolution of disaster and disaster management policy in Malaysia. *Advanced Science Letters*, 22(12), 4209-4212. <https://doi.org/10.1166/asl.2016.8107>
- Samsudin, K., Hussin, M. F., Kamarudin, A. H., Yassin, A. I. M., Hamid, W. A. S. H. W. A., Jusoh, M. H., Sulaiman, A. A., & Saadun, J. (2016b). The need for an integrated disaster management system for lead responding agency in Malaysia during response and recovery phase. *International Journal on Information*, 19(8A), 3307-3311.
- Samsudin, K., Hussin, M. F., Kamarudin, A. H., Yassin, A. I. M., Hamid, W. A. S. H. W. A., Saadun, J., & Hussain, A. R. (2016c). Model of an incident potential index and integrated disaster management online system for lead responding agency in Malaysia during response and recovery phase. In A. S. Arshad, A. H. A. Hai, & M. R. Samsudin (Eds.), *Technology and Applications for Disaster Management 2nd International Conference*, (pp. 28-29). Astronautic Technology (M) SDN. BHD.
- Samsudin, K., Hussin, M. F., Yassin, A. I. M., Hussain, A. R., Husna, K. A., Jusoh, M. H., Sulaiman, A. A., Saadun, J., & Kasri, M. H. (2016d). Policy on disaster management in Malaysia: The need of supporting governance. *Advanced Science Letters*, 22(12), 4213-4215. <https://doi.org/10.1166/asl.2016.8108>
- Samsudin, K., Hussin, M. F. F., Husna, K. A., Yassin, A. I. M., Jusoh, M. H., Sulaiman, A. A., Hamid, W. A. S. H. B. W., & Saadun, J. (2017a). Criteria for an integrated disaster management system for lead responding agency in Malaysia. *Advanced Science Letters*, 23(5), 4278-4280. <https://doi.org/10.1166/asl.2017.8248>

- Samsudin, K., Hussin, M. F., Yassin, A. I. M., Kamarudin, A. H., Jusoh, M. H., Sulaiman, A. A., Saadun, J., Kasri, M. H., & Hussain, A. R. (2017b). Design and development of strategic disaster management database for lead responding agency in Malaysia during response and early recovery phases. *International Journal of Electrical & Electronic Systems Research (IEESR)*, 10, 1-8.
- Samsudin, K., Hussin, M. F., Yassin, A. I. M., Salleh, M. K. M., Kamarudin, A. H., Jusoh, M. H., Sulaiman, A. A., Saadun, J., Kasri, M. H., & Hussain, A. R. (2018). Incident potential index model complimenting Malaysia disaster management environment. *International Journal of Engineering & Technology*, 7(3), 486-489.
- Samsudin, K., Hussin, M. F., Ghazali, N. F. N., Ghani, N. H. A., Kamarudin, A. H., Sansuddin, N., Khan, Z. I., & Hussein, K. (2020a). Association between workload and psychological well-being in malaysia elite firefighter. *International Conference on Science and Social Research*, 21(2), 374-381. <https://doi.org/10.37268/mjphm/vol.21/no.2/art.1067>
- Samsudin, K., Hussin, M. F., Kamarudin, A. H., Kasri, H., Hussain, A. R., & Yassin, A. I. M. (2020b). Effective emergency management: The roles of knowledge and practice of command structure for lead responding agencies. *Global Business and Management Research: An International Journal*, 12(1), 33-41.
- Samsudin, K., Hussin, M. F., Kamarudin, A. H., Yassin, A. I. M., Kasri, M. H., & Hussain, A. R. (2020c). Expert system development process of integrated disaster management system for lead responding agency in Malaysia during response and early recovery phases. *International Journal of Electrical & Electronic Systems Research (IEESR)*, 16, 47-53.
- Samsudin, K., Yassin, A. I. M., Kamarudin, A. H., Kasri, M. H., Saadun, J., & Hussin, M. F. (2020d). Dissecting the challenges of the lead responding agency during disaster management exercises. *Global Business and Management Research: An International Journal*, 12(1), 42-51.
- Sandler, D., & Schwab, A. K. (2021). *Hazard mitigation and preparedness*. Routledge. <https://doi.org/10.4324/9781003123897>
- Sardi, M. F., & Razak, K. A. (2019). Assessment of effectiveness of emergency response time during landslide event in Malaysia. *ASM Science Journal*, 12, 1-19. <https://doi.org/10.32802/ASMSCJ.2019.360>
- Shaluf, I. M. (2008). Technological disaster stages and management. *Disaster Prevention and Management*, 17(1), 114-126. <https://doi.org/10.1108/09653560810855928>
- ShoaiB, M., & Abdullah, F. (2020). Risk reduction of COVID-19 pandemic in Pakistan. *Social Work in Public Health*, 35(7), 557-568. <https://doi.org/10.1080/19371918.2020.1806172>
- Subramaniam, C., Ali, H., & Shamsudin, F. M. (2012). Initial emergency response performance of fire fighters in Malaysia. *Public Sector Management*, 25(1), 64-73. <https://doi.org/10.1108/09513551211200294>
- Travis, L. (2010). One of many free survey tools: Google docs. *Journal of Electronic Resources in Medical Libraries*, 7(2), 105-114. <https://doi.org/10.1080/15424065.2010.482902>
- Waring, S., Moran, J. L., & Page, R. (2020). Decision-making in multiagency multiteam systems operating in extreme environments. *Journal of Occupational and Organizational Psychology*, 93(3), 629-653. <https://doi.org/10.1111/joop.12309>

A Topic Modeling and Sentiment Analysis Model for Detection and Visualization of Themes in Literary Texts

Kah Em Chu¹, Pantea Keikhosrokiani^{1*} and Moussa Pourya Asl²

¹*School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

²*School of Humanities, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

ABSTRACT

Despite the growing emergence of new computer analytic software programs, the adoption and application of computer-based data mining and processing methods remain sparse in literary studies and analyses. This study proposes a text analytics lifecycle to detect and visualize the prevailing themes in a corpus of literary texts. Two objectives are to be pursued: First, the study seeks to apply a Topic Modeling approach with selected algorithms of LDA, LSI, NMF, and HDP that can effectively detect the recurring topics about the major themes developed in the dataset. Second, the project aims to apply a Sentiment Analysis model that can analyze the polarity of writers' discourse on the detected thematic topics with the algorithms of Vader and TextBlob. The implementation of Topic Modeling has detected six thematic topics of sex, family, revolution, imprisonment, intellectual, and death. The adoption of the Sentiment Analysis model also revealed that the feelings attached to all the identified themes are largely negative sentiments expressed towards socio-political issues.

Keywords: Iranian diaspora, life writing, sentiment analysis, text mining, topic modeling

INTRODUCTION

Collecting and processing textual evidence is widely recognized as the principal strategy to detect and evaluate the underlying themes of literary writings. The aim is to identify the prevailing view of the world or human nature by closely examining the story's plot, characterization, and the dominant conflicts within the text. Literary scholars traditionally use manual qualitative content analysis to perform this kind of examination

ARTICLE INFO

Article history:

Received: 21 September 2021

Accepted: 04 March 2022

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.14>

E-mail addresses:

pantea@usm.my (Pantea Keikhosrokiani)

kahem@student.usm.my (Kah Em Chu)

moussa.pourya@usm.my (Moussa Pourya Asl)

* Corresponding author

which poses certain methodological challenges (Ying et al., 2022). Due to the rising level of abstraction and the high degree of subjective interpretation used in the process, the credibility and authenticity of such analyses have always been a matter of critical debate and controversy (Graneheim et al., 2017; Ying et al., 2021). In recent years, advances in computer analytic software programs have made collecting and analyzing text corpus much easier by replacing manual processing with systematic and automatic procedures (Firmin et al., 2017; Misuraca et al., 2021). However, despite the emergence of new methods and analytic tools, the adoption and application of these new computer-based methods remain sparse in literary studies and analyses.

This study proposes a text analytics lifecycle for detecting and visualizing the dominant themes in a corpus of literary texts. Therefore, the main objectives of this study are as follows:

1. to apply topic modeling techniques for detecting the topics related to imprisonment from 28 Iranian diasporic life writings,
2. to apply sentiment analysis for analyzing the polarity of diasporic writers' discourse on imprisonment with the selected algorithms.

Topic Modeling (Shi et al., 2018) and Sentiment Analysis (Alaei et al., 2019) are analytical techniques in the proposed digital model. Topic modeling with selected algorithms of Latent Dirichlet Allocation (LDA), Latent Semantic Indexing (LSI), Non-negative Matrix Factorization (NMF), and Hierarchical Dirichlet Process (HDP) can effectively detect the recurring topics concerning the major themes developed in the texts. On the other hand, sentiment analysis can help us to analyze the polarity of writers' discourse on the major themes with the algorithms of Vader and TextBlob. However, even though both models enable researchers to produce more detailed textual analysis, the implementation of topic modeling and sentiment analysis on literary texts is scarce. This study aims to apply topic modeling and sentiment analysis to literary texts. Hence, the study would benefit literary scholars with a more accurate analytical tool and methodology and provide data scientists with a comparison of how different topic modeling and sentiment analysis algorithms perform on a text corpus.

The paper is structured as follows. After briefly reviewing the existing literature on Topic Modeling and Sentiment Analysis approaches, we elaborate on the proposed analytical strategy. An implementation on a text corpus related to different books is then presented to show how the strategy operates. Finally, the paper ends with a discussion of the results and concludes with theoretical and practical implications of the approach.

LITERATURE REVIEW

Natural Language Processing (NLP) is considered the interaction between computers and human language, which can be used for examining a text and generating insight from it. NLP

is widely used to study the opinion and sentiment of the target corpus. Sentiment analysis uses NLP and text analytics to identify, extract, analyze, and study subjective information. Topic modeling is another NLP technique for discovering the abstract “topics” that occur in a collection of documents. For instance, sentiment analysis is used to study the polarity of the opinion of Twitter users about the TV series “Game of Thrones.” In contrast, topic modeling was used to visualize the weightage of the content related to the selected topic. For example, the topic can be a character or a place in the “Game of Thrones” TV Series (Scharl et al., 2016). This section includes the relevant literature related to topic modeling, sentiment analysis, and related studies.

Topic Modeling

Topic modeling refers to the weightage of a target theme in the whole corpus. For example, it can study the percentage of references to country names in a text (Costa, 2018). Topic modeling has a wide variety of visualization such as line charts, bar charts, pie charts, word clouds, and heatmaps. Topic modeling is used to group words for a set of texts. Because it automatically categories words without a specified list of labels, this is referred to as unsupervised learning (Sukhija et al., 2016). After feeding the data model, sets of words will appear from which the main topic can be assumed. However, it is challenging to understand the proper topic merely by looking at a combination of words and numbers. Therefore, topic modeling provides visualization, considered one of the most effective methods of understanding the data. Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) are popular topic modeling algorithms. LDA is famous for visualization using pyLDAvis. LSA and LDA are Vector Space Models (VSM) that represent text documents as vectors in a high-dimensional space (Rehurek & Sojka, 2010).

Table 1 is a comparison of different Topic Modeling Methods that shows a summary of the concept and advantages, and disadvantages of five topic modeling approaches (Hornick, 2017; Gabrilovich & Markovitch, 2006; Mazzola et al., 2018; Řehůřek, 2019; Rehurek & Sojka, 2010; Shi et al., 2018). As shown in Table 1, ESA and LSA are tightly connected, and in principle, ESA cannot outperform the peak performance of LSA. Therefore, ESA is not used as part of the comparison of different topic modeling algorithms for this study.

In this project, LSA, LDA, NMF, and HDP will be implemented, and the performance of each model will be compared. The four algorithms are selected because they are open-source resources and suitable for large documents.

Sentiment Analysis

Sentiment analysis is a kind of opinion mining that is usually performed on comments of users about an object or a topic (Ying et al., 2021; Malik et al., 2021; Keikhosrokiani & Asl, 2022) to determine its sentiment orientation or whether the comments of a selected

Table 1
 Comparison of different topic modeling methods

Method	Criteria	Concept	Advantages	Disadvantages
Latent Semantic Analysis (LSA)	Exploits co-occurrence between terms to the project documents into a low-dimensional space. The inference is made using linear algebra Singular Value Decomposition (SVD).	Can be combined with another algorithm and can produce a relatively fast model. It is fast and popular.	May not perform well when working with short documents. It is based on dimensional reduction of the original dataset, whereas the dimension factors' determination is subjective.	
Latent Dirichlet Allocation (LDA)	A fully generative model is based on the bag of words paradigm, and word document counts. Documents are assumed to have been generated according to per-document topic distribution and per-topic word distribution.	Efficient when the corpus is large. Noise reduction is possible. It is popular, and there are many examples.	The topics discovered by using LDA are implicit and hard to interpret as they are defined only using their keywords, but not labels or abstract descriptions. Sometimes the keywords from different topics overlap and do not yield a proper topic name: the explanation is fuzzy.	
Non-negative Matrix Factorization (NMF)	Discover topics by decomposing the document-term matrix into two low-rank factor matrices	Efficient when the corpus is large. Does not require a predefined number of topics.	May not perform well when working with short documents. Less interpretative as it lacks explicit probabilistic meaning of each factor.	
Explicit Semantic Analysis (ESA)	Compute the "semantic relatedness" between the documents and humans' defined topics to improve text document categorization	It is a knowledge base, hence, it can be assigned with human-readable labels to the topic. ESA can discover relevant topics even when the topic is overlapping.	The previous ESA research used Wikipedia as a knowledge repository, more suitable for an expert system model. The conceptual motivation for ESA, recent work has observed unexpected behavior. ESA and LSA are tightly connected, and in principle, ESA cannot outperform the peak performance of LSA.	
Hierarchical Dirichlet Process (HDP)	A nonparametric Bayesian approach to clustering grouped data	Does not require a predefined number of topics	The maximum number of topics can be unbounded and learned from the data rather than specified in advance. It is more complicated to implement and unnecessary when a bounded number of topics is acceptable.	

topic are positive, negative, or neutral (Ding et al., 2008). According to Vinodhini and Chandrasekaran (2012), sentiment analysis can be conducted at three levels document level, sentence level, and attribute level. Two main approaches can be applied for sentiment analysis: semantic orientation and machine learning. Semantic orientation is unsupervised. Therefore, prior training in data is not needed. However, Machine learning refers to supervised learning and unsupervised learning. For supervised machine learning, prior training in data is essential (Vinodhini & Chandrasekaran, 2012). A hybrid machine learning method and semantic approach are also widely applied (Lodin & Balani, 2017).

Supervised Machine Learning methods can achieve 80–84% accuracy for Sentiment Analysis (Lodin & Balani, 2017). The popular Machine Learning methods in Sentiment Analysis are Support Vector Machine (SVM), Naïve Bayes, and Maximum Entropy (Abdelrahman & Keikhosrokiani, 2020; Teoh & Keikhosrokiani, 2020). SVM is one of the most popular machine learning algorithms as it is a high-performing algorithm with a little tuning. When there is a clear margin of distinction between classes, SVM performs well. In high-dimensional spaces, SVM is more effective. When the number of dimensions is more than the number of samples, SVM is more effective. SVM uses a small amount of memory. Machine Learning models have higher accuracy for sentiment analysis, but they also require more time to train the model. Thus, semantic analysis is more suitable for real-time applications (Vinodhini & Chandrasekaran, 2012). An unsupervised machine learning method applied for Sentiment Analysis is Clustering (Alaei et al., 2019).

In a study, Ding et al. (2008) proposed a lexicon-based method to use the opinion-bearing words to determine whether the comments are positive or negative. Opinion-bearing words are often used to express positive or negative opinions, for example, amazing and ugly. The algorithm counts the number of positive and negative opinions bearing words near the study's subject or topic. If there are more positive than negative sentiments, then the opinion on the topic is positive, or vice versa. However, this method is ineffective in context-dependent opinion-bearing words (Ding et al., 2008). Lexicon-based techniques are based on word dictionaries, where each word relates to a certain sentiment, and the overall sentiment is calculated. The lexicon-based methods are restricted by their lexicon, and the sentiment values assigned to the words in the dictionary neglect the context. In different contexts, the same adjective could have a different sentiment. Lexicon-based approaches are limited by their lexicons, specifically the static prior sentiment values of words or concepts in all circumstances. In order to overcome this limitation, various techniques have been proposed to study the semantics (Lodin & Balani, 2017). One of the solutions to overcome Lexicon-based techniques is to assign an updated sentiment strength to words in the lexicon. However, it still needs to be trained from manually annotated corpora. Another issue with lexicon-based techniques is that they rely entirely on the presence of phrases that express sentiment overtly, but the sentiment of a term is often implicitly reflected by the semantics of its context (Lodin & Balani, 2017).

The semantic analysis was introduced to improve the lexicon-based approach. Compared to the lexicon-based approach, the semantic analysis uses a dictionary of domain-specific terms, and the polarity of the terms is required (Alaei et al., 2019). There are two methods for semantic analysis: contextual semantic methods and conceptual semantic methods. Contextual semantic methods, also known as statistical semantics or corpus-based methods, determine the semantics based on the co-occurrence patterns of terms. Conceptual semantic methods, sometimes called dictionary-based approaches, use an external semantic knowledge base or a dictionary with the domain specified terms and sentiments attached to the words with NLP techniques to capture the sentiment. Examples include SenticNet, SentiStrenght, and WordNet (Ding et al., 2008; Lodin & Balani, 2017; Alaei et al., 2019). In some studies, a hybrid approach uses both machine learning and lexicon-based approaches in the same model. As a result, the accuracy is higher than the Naïve Bayes model (Appel et al., 2016; Lodin & Balani, 2017). Another study by Mumtaz and Ahuja (2018) proposed a model in which the accuracy of the hybrid system is found to be around 93%, which is higher than the pure lexical and SVM algorithm.

In this project, a lexicon-based or rule-based algorithm will be used for sentiment analysis as the data are in text form with no train and test set available to create a Machine Learning model. Preparing a labeled dataset from pure text data is time-consuming. Furthermore, the machine learning classifier is domain-based and needs to be retrained if future works use a dataset from a different domain. Hence, TextBlob and VADER libraries written in Python for rule-based sentiment analysis will be used. For sentiment analysis, three approaches are available, as shown in Table 2 (Devika et al., 2016; Ding et al., 2008; Lodin & Balani, 2017).

Table 3 summarizes the comparison of the related studies based on the data source, methods and tools, and their visualization.

MATERIALS AND METHODS

In order to achieve the main goal of this study, Cross-Industry Standard Process for Data Mining (CRISP-DM) Methodology is improvised as a new proposed text analytics lifecycle in this project. There are six stages in CRISP-DM Methodology: (1) Business Understanding, (2) Data Understanding and Preparation, (4) Modeling, (5) Evaluation, and (6) Deployment. For this project, data understanding and data preparation steps are combined. Then, two new steps are added: text processing and text exploration, while the deployment phase is not included in this project. Figure 1 shows the proposed digital model developed and adopted to detect and visualize themes in literary works.

The main steps of the proposed digital model are business understanding, data understanding and preparation, text processing, text exploration, modeling, and evaluation which are explained in detail. Business understanding emphasizes the specification of the

Table 2
Comparison of different sentiment analysis methods

Criteria	Method		
	Machine Learning Approach	Semantic-Based Approach	Lexicon/Rule-Based Approach
Classification	<ol style="list-style-type: none"> Supervised Learning Unsupervised Learning 	Unsupervised Learning	Unsupervised Learning
Advantages	<ol style="list-style-type: none"> Not necessary to have a dictionary High accuracy of classification demonstrated Can train with or without Semantic concept 	<ol style="list-style-type: none"> Performance of sentiment classification at the sentence level is better than word level Consider the context of the words Does not need labeled data 	<ol style="list-style-type: none"> Labeled data and the procedure of learning are not required Fast
Disadvantages	<ol style="list-style-type: none"> Classifier trained are domain-based Required significant time to train 	<ol style="list-style-type: none"> Accuracy and efficiency depend on defining rules Domain Specified 	<ol style="list-style-type: none"> Requires powerful linguistic resources, which is a scarce resource Context of the words is not considered Relying on POS Tagger
Algorithm	<ol style="list-style-type: none"> Naïve Bayes Support Vector Machine K Nearest Neighbour 	<ol style="list-style-type: none"> SenticNet SentiStrength 	<ol style="list-style-type: none"> TextBlob VADER
			Combination of Machine learning and a lexicon-based approach Higher accuracy Any method from the previous approach

Table 3
Comparison of related studies

No	Author (Year)	Topic	Data Source	Method / Tools	Visualization
1	(Scharl et al., 2016)	Analyzing the public discourse on works of fiction – Detection and visualization of emotion in online coverage about HBO's Game of Thrones	Website: Anglo-American News Media social media: Twitter, Facebook, Google+, YouTube	NOVEL Developed Westeros Sentinel, utilizes the weblYzard: web intelligence Sentiment analysis, topic modeling	Interactive dashboard shows the weightage of the occurrence of the events and the characters in the data, and the sentiment orientation.
2	(Costa, 2018)	A method for content analysis applied to newspaper coverage of Japanese personalities in Brazil and Portugal	Portuguese and Brazilian newspaper CHAVE corpus	Topic Modeling log-likelihood ratio to rank words according to their relative frequency differences in two corpora	Percentage of texts referring to other countries Percentage of texts by section Percentage of texts mentioning Japanese personalities Percentage of texts referring to Kurosawa, Oshima, and the Emperor Akihito Distribution by a week of texts referring to Hosokawa in the first semester of 1994
3	(Ding et al., 2008)	A Holistic Lexicon-Based Approach to Opinion Mining	Customer reviews of 8 products: two digital cameras, one DVD player, one MP3 player, two cellular phones, one router, and one anti-virus software	Opinion Observer (Proposed model) Sentiment Analysis	Accuracy of the model
4	(Paroubek & Pak, 2010)	Twitter as a Corpus for Sentiment Analysis and Opinion Mining	Twitter	Sentiment Classification (unigrams, bigrams, and trigrams)	The distribution of the word frequencies Impact of different parameters on the accuracy of the Sentiment Classifier

Table 3 (continue)

No	Author (Year)	Topic	Data Source	Method / Tools	Visualization
5	(Rehurek & Sojka, 2010)	Software Framework for Topic Modeling with Large Corpora	Mathematical papers from the Czech Digital Mathematics Library DML-CZ, from the NUMDAM repository, and the math part of arXiv	Latent Semantic Analysis Latent Dirichlet Allocation	None
6	(Vinodhini & Chandrasekaran, 2012)	Sentiment Analysis and Opinion Mining: A Survey	Movie review dataset Product review from amazon.com	Sentiment analysis Naïve Bayes Classifier	Performance of the sentiment classification model and sentiment analysis model.
7	(Grayson et al., 2017)	Exploring the Role of Gender in 19th-century Fiction Through the Lens of Word Embeddings	Forty-eight novels from twenty-nine 19 th -century novelists sourced from Project Gutenberg,	word2vec t-Distributed Stochastic Neighbor Embedding (t-SNE)	Word frequencies for our initial list of gender-encoded words. The cosine similarity scores between female and male-authored words in our gender-encoded list.
8	(Grayson et al., 2016)	Novel2Vec: Characterizing 19 th -Century Fiction via Word Embeddings	Twelve popular 19th-century novels were written by Jane Austen, Charles Dickens, and Arthur Conan Doyle.	Two variants of word2vec, a continuous bag-of-words strategy and a skip-gram strategy,	Word embedding visualization Context window sensitivity comparison
9	(Grayson et al., 2016)	The Sense and Sensibility of Different Sliding Windows in Constructing Co-occurrence Networks from Literature	A collection of nine novels from two 19th century British novelists—six by Jane Austen and three by Charles Dickens—sourced from Project Gutenberg	Applying different sliding window methodologies to capture character co-occurrences within the literature to build social networks	Social network of the characters in the novel
10	(Leavy et al., 2020)	Mitigating Gender Bias in Machine Learning Data Sets	A set of over 16,000 volumes of 19th-century fiction from the British Library Digital corpus	Sentiment analysis	Terms denoting emotion associated with men and women were extracted, and the levels of association

Table 3 (continue)

No	Author (Year)	Topic	Data Source	Method / Tools	Visualization
11	(Leavy et al., 2020)	Curatr: A Platform for Exploring and Curating Historical Text Corpora	35,918 English language fiction and non-fiction books dating from 1700 to 1899.	Curatr	Semantic network
12	(Leavy, 2019)	Curatr: A Platform for Semantic Analysis and Curation of Historical Literary Texts	16,426 works of fiction in the corpus	Curatr	None
13	(Suhendra et al., 2022)	Opinion Mining and Text Analytics of Literary Reader Responses	Goodreads review for KL Noir Volumes	Sentiment analysis & topic modeling	Book's rating, most salient terms
14	(Jafery et al., 2022)	Text Analytics Model to Identify the Connection Between Theme and Sentiment in Literary Works	Diasporic women from Iraq wrote six life writings were	Latent Dirichlet Allocation (LDA), sentiment analysis	Most salient terms, sentiments related to themes
15	(Al Mamun et al., 2022)	Sentiment Analysis of the Harry Potter Series Using a Lexicon-Based Approach	Harry Potter book series	Lexicon-based sentiment analysis	Sentiment frequency, sentiment analysis, AFINN sentiment dictionary, a sentiment of hero characters, a sentiment of houses
16	(Sofian et al., 2022)	Opinion Mining and Text Analytics of Reader Reviews of Yoko Ogawa's The Housekeeper and the Professor in Goodreads	Reader reviews of Yoko Ogawa's The Housekeeper and the Professor on Goodreads	Sentiment analysis & topic modeling	Book's rating, most salient terms, bigrams, trigrams

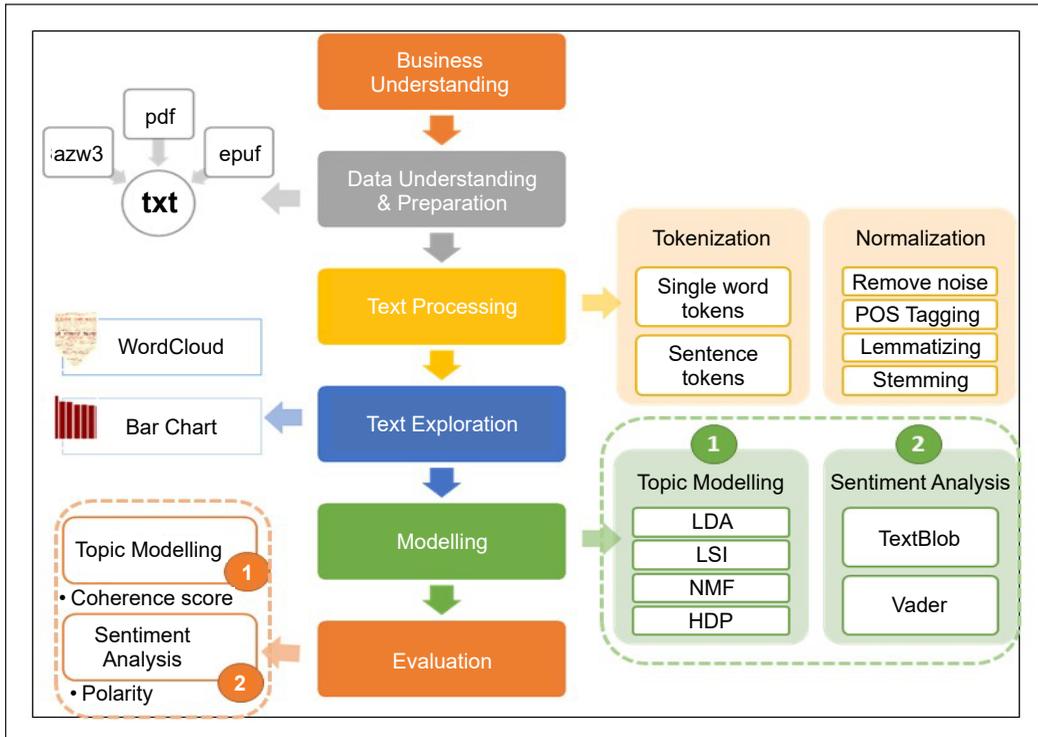


Figure 1. Proposed text analytics lifecycle

problem and methods of evaluating the achievement of the goal. The data understanding and preparation step is used to discover and reformat the secondary data, which are 28 literary books selected for this project. In the text processing step, tokenization and normalization are applied. The modeling step consists of two main models: (1) topic modeling and (2) sentiment analysis. First, topic modeling is used to develop a model with selected algorithms of LDA, LSI, NMF, and HDP that can detect the topics related to imprisonment from the 28 Iranian diasporic life writings. Then, the results of different topic modeling algorithms are compared based on coherence scores. The second part of the modeling step focuses on sentiment analysis to create a model that can analyze the polarity of diasporic writers’ discourse on imprisonment with the selected algorithms of Vader and TextBlob. Finally, the sentiment analysis models are evaluated by comparing the polarity scores. Each step of the proposed digital model is explained in detail in the following sections.

Data Understanding and Preparation

The data used in this study are secondary data from a corpus of 28 books that cover contemporary life writings by the Iranian diaspora, as shown in Table 4. All the works were published after the 1979 Islamic revolution and revolved around similar subject matters. The works are temporally, spatially, and thematically related. Together, they serve as a text

corpus for the aims of this study. Since the dataset is not labeled, unsupervised learning, particularly the lexicon/rule-based approach, is the best method (Ding et al., 2008; Lodin & Balani, 2017).

Table 4
List of Literary works used in this project

No	Book Name
1	Fatemeh Keshavarz–Jasmine and Stars_Reading More Than Lolita in Tehran -The University of North Carolina Press (2007).pdf
2	Nemat, Marina–Prisoner of Tehran_One Woman's Story of Survival Inside an Iranian Prison.pdf
3	Andalibian, Rahimeh–The Rose Hotel_A Memoir of Secrets, Loss, and Love from Iran to America.pdf
4	Anita Amirrezvani–The Blood of Flowers-Back Bay Books (2008).pdf
5	Ansary, Nina–Jewels of Allah _the untold story of women in Iran-Revela Press (2015) .pdf
6	Azadeh Moaveni–Guest House for Young Widows_Among the Women of ISIS-Random House (2019).pdf
7	Azar Nafisi–Reading Lolita in Tehran_A Memoir in Books-Random House (2003).pdf
8	Azar Nafisi–Things I've Been Silent About_Memories of a Prodigal Daughter-Random House (2008).pdf
9	Basmenji, Kaveh–Afsaneh_Short Stories by Iranian Women.pdf
10	Bijan, Donia–Maman's Homesick Pie A Persian Heart in an American Kitchen.pdf
11	Dalia Sofer–The Septembers of Shiraz (2007).pdf
12	Dina Nayeri –The Ungrateful Refugee-Canongate Books (30 May 2019).pdf
13	Ebadi, Shirin–Until we are free _ my fight for human rights in Iran-Random House (2016).pdf
14	Entekhabifard, Camelia–Camelia_ Save Yourself by Telling the Truth_ A Memoir of Iran.pdf
15	Esfandiari, Haleh–My Prison, My Home_One Woman's Story of Captivity in Iran.pdf
16	Firoozeh Dumas–Funny in Farsi_ A Memoir of Growing Up Iranian in America-Villard (2003).pdf
17	Firoozeh Dumas–Laughing without an accent_ adventures of an Iranian American, at home and abroad -Villard (2008).pdf
18	Gohar Homayounpour–Doing Psychoanalysis in Tehran-The MIT Press (2012).pdf
19	Goldin, Farideh–Wedding Song_ Memoirs of an Iranian Jewish Woman.pdf
20	Nafisi, Azar–The Republic of Imagination_ America in Three Books.pdf
21	Nourai-Simone, Fereshteh_ Farrokh, Faridoun_ Khalili, Sara–The Shipwrecked_ Contemporary Stories by Women from Iran.pdf
22	Rachlin Nahid–Persian-Girls_-A-Memoir.pdf
23	Rostampour, Maryam–Captive in Iran.pdf
24	Roxana Saberi–Between Two Worlds_ My Life and Captivity in Iran -Harper Perennial (2011).pdf
25	Shahla Talebi–Ghosts of Revolution_ Rekindled Memories of Imprisonment in Iran -Stanford University Press (2011).pdf
26	Shirin Ebadi_Rich, Nathaniel–The golden cage _ three brothers, three choices, one destiny-Kales Press (2011).pdf
27	Zanjani, Sohila_ Brewster, David–Scattered Pearls_ Three generations of Iranian women and their search for freedom.pdf
28	Zarah_ Ghahramani–My-Life-as-a-Traitor_-An-Iranian-Memoir.pdf

Text Processing

Text processing consists of two main phases: tokenization and normalization (Mayo, 2017). Tokenization is splitting longer text strings into smaller pieces or tokens (Subramanian, 2019). Normalization refers to converting numbers to their word equivalents, removing punctuation, converting all text to the same case, and lemmatization, which means returning the word to its base form. Tagging is the second step after tokenization in the typical NLP pipeline (Bird et al., 2009). Part-of-speech (POS) tagging is the process of classifying and labeling words into their parts-of-speech. Then, *tagset* refers to the collection of tags used for a task. After POS Tagging, only nouns and adjectives remain for the following process as other word tags like pronouns and determiners do not carry any insight.

Text Exploration

For an expository text, text exploration is when the text is examined critically and tries to find and prioritize the main ideas or facts, or the essential event and characters, in the case of a narrative composition. As the original text contains many topics, only the sentences which contained any keyword from the list of 324 related keywords to the topic of imprisonment are extracted. About 19 667 sentences are extracted out of the 137 319 sentences in the corpus. After extracting the sentences, the data cleaning processes of tokenization, removing stop words and irrelevant values, POS Tagging, and lemmatization are performed on the extracted text for a second time. Bar charts of top frequent words, Bigrams and Trigrams WordCloud, are created based on the cleaned text using python. These visualizations are used as descriptive data to give a general idea about the dataset. Figure 2 shows 10 frequent words, among which *woman* is the most frequent term while the *law* is the least frequent one. Within the extracted text, nouns such as woman and time

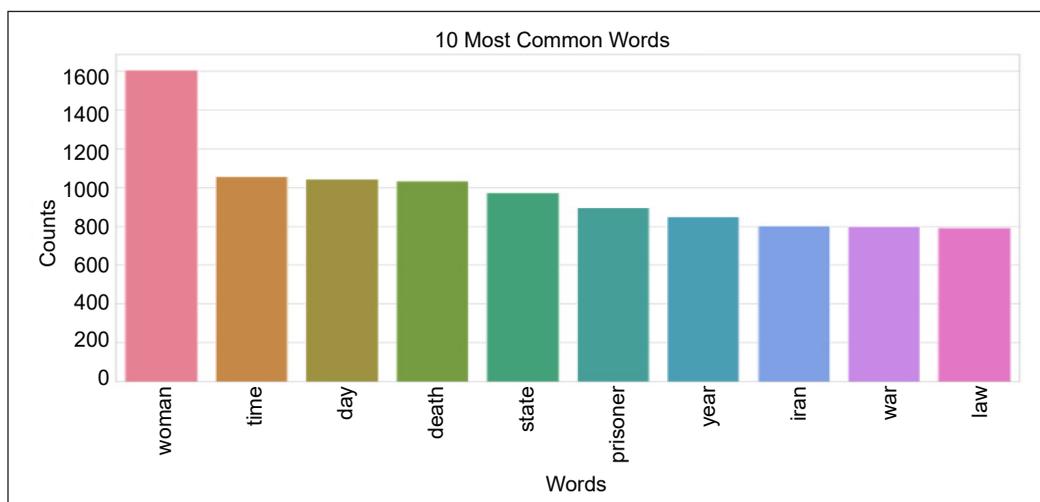


Figure 2. Bar chart of top 10 frequent terms

are still the main discussion subject, accompanied by other terms such as state, prisoner, and death. The most frequent adjectives also appear semantically neutral, leaning slightly towards positivity.

The WordClouds in Figure 3 shows the most frequently occurring bigrams and trigrams. An expert interprets the WordClouds in Iranian diasporic literary and cultural studies. Based on the WordClouds and experts' opinions, the following assumption can be made about the topics:

1. The prisoners are criminals, political prisoners, or war prisoners in Evin Prison.
2. Women demand their rights and family protection laws related to the hypothesis made with WordCloud of Unigram that women seek a new and better life.
3. The prisoners demand their rights, such as human rights and civil rights.

Similar to topic modeling, text exploration is done before sentiment analysis. In this phase, the text is first tokenized, then nouns and adjectives are extracted to create WordClouds [Figures 4(a) and 4(b)]. The WordCloud of adjectives shows that the dominant sentiment is rather positive because of the high frequency of terms like *new* and *good* and the neutrality of other adjectives like *Iranian* that indicate nationality. Here, a hypothesis is suggested that Iranian women are looking forward to having a new and better life.



Figure 3. WordClouds: (a) Most frequent bigrams; (b) Most frequent trigrams



Figure 4. WordClouds: (a) WordCloud of nouns; (b) WordCloud of adjectives

Topic Modeling

Two models are built in this project for sentiment analysis and topic modeling. The process of topic modeling is shown in Figure 5.

The first step of developing the model is to generate a dictionary and a corpus from the text preprocessed in the previous sub-section. Here, the dictionary refers to id2word, which means giving an ID to every unique word in the text. Corpus refers to the term-document frequency, where the frequency of the term in the document is calculated. The corpus only contains the ID of the word and its frequency. By passing the dictionary, the ID can be changed to the word. The example of the corpus is shown in Figure 6.

In this project, four algorithms, namely, LDA, LSI, NMF, and HDP, are used, and their performance is compared in topic modeling. LDA, LSI, and NMF require the number of topics defined to build the model. The number of topics representing the number of keyword lists is generated from the model. The performance of topic modeling models is evaluated based on coherence score. The performance is better when the coherence score is higher. In order to obtain the number of topics that yield the optimal coherence score, the model is iterated from 10 topics to 30 topics. The reason to start from 10 topics rather than from one is to ensure the sufficiency of the generated topics to cover all discussions in the corpus.

On the other hand, to prevent generating too many topics, the iteration is stopped at 30. After the iteration is completed, the number of topics with the highest coherence score is retrieved, and the generated topics are saved for analysis. Figures 7(a) to 7(c) show the chart of coherence score of the models from 10 topics to 30 topics. An Iranian diasporic literary and cultural studies expert manually evaluates the polarity accuracy. Based on the line chart of the coherence score of each model, each algorithm has different changes when the number of topics is increased. LDA models' coherence score showed an increasing trend when the number of topics increased. For LSI and NMF models, the performance decreased when

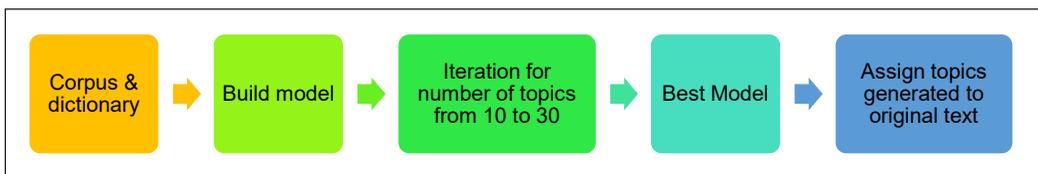


Figure 5. Topic modeling process

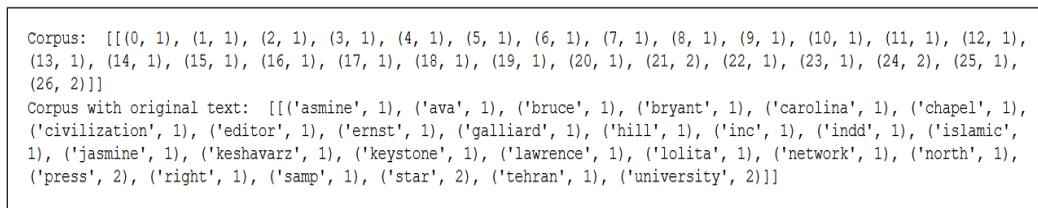
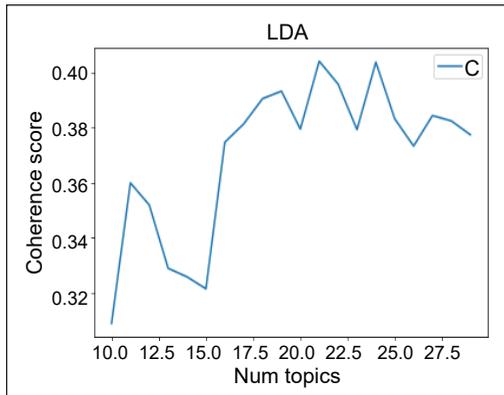
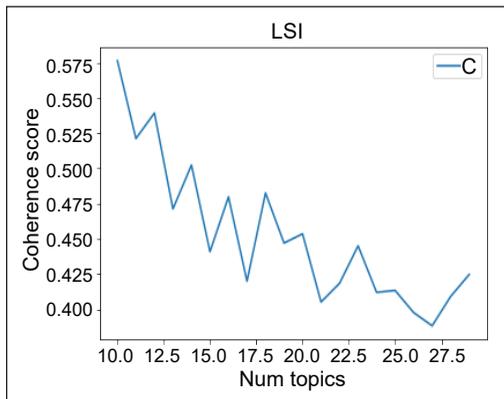


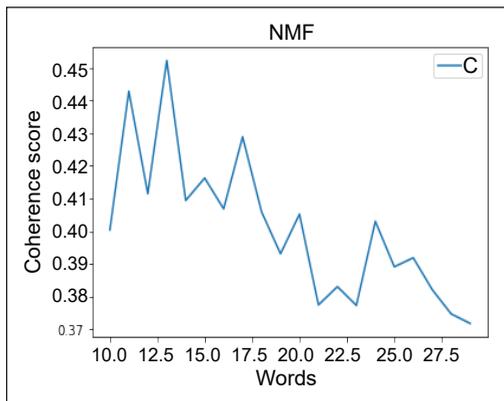
Figure 6. Example of corpus



(a)



(b)



(c)

Figure 7. Line Chart of Coherence Score: (a) Line Chart of Coherence Score for LDA Model, Best Coherence Score: 0.404, Number of Topics: 21; (b) Line Chart of Coherence Score for LSI Model, Best Coherence Score: 0.576, Number of Topics: 10; (c) Line Chart of Coherence Score for NMF Model, Best Coherence Score: 0.452, Number of Topics: 13.

the number of topics increased. HDP model does not require several topics; hence, the model is built directly. After generating the list of keywords, the keywords can be used to assign the most relevant topic for each sentence in the text.

Sentiment Analysis

This project uses two python libraries, Vader and TextBlob, to calculate the corpus' sentiment based on a rule-based approach. The development of Sentiment Analysis is more straightforward than Topic Modeling. The only step needed is to pass the sentences to the analyzer, which will return with the polarity of the sentences, as shown in Figure 8.

In Vader libraries, the analyzer calculates the score of negativity, neutrality, and positivity of the sentences and calculates the compound score based on the earlier scores. While in Text blob, the analyzer calculates the polarity and subjectivity of the sentences. Both compound score and polarity range from -1.0 to +1.0, while subjectivity ranges from 0 to 1, where 0 means the sentence is very objective, and one indicates that the sentence is very subjective. Like Topic Modeling, only the extracted text is used in Sentiment Analysis. The text data used in this project is extracted purely from the discourse; hence there are no labels on the polarity for each sentence. Instead, a sample is extracted from the model's output to evaluate the accuracy of the polarity calculated by the models. The polarity of each piece of text in the sample is manually identified to verify if the model performs as anticipated.



Figure 8. Process of sentiment analysis

Evaluation

The model's performance is measured with a coherence score for Topic Modeling. The coherence score of each model is calculated and compared. The higher the coherence score, the better the model. For Sentiment Analysis, the model's performance is analyzed by comparing the polarity (TextBlob Model) and Compound Score (Vader) with the manually assigned polarity to the text sample. The results from the model are also compared with the related literature.

RESULTS

Objective 1: Topic Modeling

As shown in Table 5, the best model for topic modeling is LSI, followed by HDP, NMF, and LDA. In LSI, the keywords of the 10 topics are repeated, but they can still be identified as different from each other by retrieving part of the text that falls on a specific topic. Six themes are derived from the 10 groups of keywords and their related texts: (1) sexism, (2) family, (3) revolution, (4) imprisonment, (5) intellectual, and (6) death. Based on the line chart of the coherence score of each model, each algorithm shows different changes when the number of topics varies. LDA models' coherence score is higher when the number of topics grows. For LSI and NMF models, the performance declines when the number of topics drops. As HDP does not need to determine the number of topics, this model is formed directly. After generating the list of keywords, the keywords are used to assign the most relevant topic for each sentence in the text.

Table 5
Results of topic modeling

Algorithm	Number of Topics	Coherence Score
LDA	21	0.404
LSI	10	0.576
NMF	13	0.452
HDP	-	0.496

Objective 2: Sentiment Analysis

Two python libraries have been used to generate the sentiment of the corpus. Comparing a sample from the results obtained from each library reveals that Vader generates more meaningful results than others, and the classification of the sentiments is more accurate than others. The top 10 positive sentences from the corpus are shown in Tables 6 and 7 using Vader and TextBlob. The positivity comes from appreciation, freedom, intellectual companion, power, wealth, and glory based on the sentences. Even though the extracted sentences are related to the notion of imprisonment, the detected positive sentiments are

Table 6
Top 10 positive sentences generated by Vader

Top Positive Vader		
Score Compound	Sentence	Justification (Theme)
0.9903	but i have grown wiser and more appreciative not only of the material comforts i unthinkingly enjoy every day a leisurely cup of coffee a moment in the sunlight the reassuring touch of shauls hand on mine but of the freedom with which i am blessed	Appreciate to have freedom
0.985	many years ago my mother asked me if i knew why it was a great deal more important to be happy than to be rich famous and beautiful because as schopenhauer informs us if we are telling a friend about a very rich famous and attractive person the first question we have to answer is but is he or she happy	Happiness is more important than rich, famous and beauty
0.9847	to the memory of our dear friend shirin alam hooli whose courage kindness and love live on in the hearts of all who knew her to the precious women who were with us in evin during our imprisonment some of whom have since been released and to all the women in evin today still waiting for the justice that only a free nation can give them	Memory in prison
0.9795	one can see why babbitt would be both attracted to the joys of freedom and frightened by its perils for freedom does have many perils and the best way to confront them is not to avoid being free but to cultivate independence of thought the kind of freedom that incidentally has been the great engine of american creativity and vitality in all fields from engineering to literature	Freedom is the source of creativity and vitality
0.9765	but of course there are all different kinds of freedom and the kind that is most precious you will not hear much talked about in the great outside world of wanting and achieving and displaying	Freedom
0.9719	but my heart was so filled with love and our shared experience that i felt him inside and tried to convince myself he was still alive	Love
0.9689	each one of us was honored with a special role in farahs life and everyone was more than his or her assigned role mahnaz was more than a sister neda more than a daughter nema more than a son hamid more than a brother jaleh more than a former comrade and best friend roshanak more than a former sisterinlaw bahram more than an intellectual companion and within that exclusive list i was left with the role of more than a childhood friend	Intellectual companion

Table 6 (continue)

Top Positive Vader		
Score Compound	Sentence	Justification (Theme)
0.9676	but it is also about wealth its great attraction as well as its destructive power the carelessness that comes with it and yes it is about the american dream a dream of power and wealth the beguiling light of daisys house and the port of entry to america	Power and Wealth
0.9666	cent warriors gained ready to give her rudabeh who is how shabby all when devoid of the love for which one of the many important who perform feats is women all norms of his own shahnameh her son are the war and win glory is for their their role as moth a different kind of courage as men they love	Glory
0.9635	before he is fully awake that alarm clock is described in great detail early in the novel we are invited to recognize that this all american businessman a defender of individualism and free trade is best defined not by any peculiarity of temperament or cherished keepsake but by his ownership of the best of the nationally advertised and quantitatively produced alarm clocks with all modern attachments making its owner proud of being awakened by such a rich device	Proud to be awake because of individualism and free trade

Table 7

Top 10 positive sentences generated by TextBlob

Top Positive TextBlob		
Polarity	Subjectivity	Sentence
1	1	she has always seemed perfect serving others having nothing but goodness in her heart
1	1	he read them with authority as if they were his as if he had composed every word like a perfect melody
1	0.3	i gave the best speech of my life there in that courtroom
1	0.3	maryam lived in one of the best an octagonal building known as eight heavens
1	0.3	60 farzanehs exceedingly unbiased viewpoint may best be exemplified in the philosophy of abbasgholizadeh we know that secular women do not share our convictions but this does not give us any problems since we are all working to promote the status of women
1	1	walid had listened to their arguments as a teenager on cassette tapes and videos impressed by their plans to bring justice to society
1	0.3	after school we were taken minas house where layla who we were entertained did their best to divert us
1	1	being i ate to hear her talk it would seem my greatest fault has been not at deaths of ourselves start door
1	1	the prosecu of my father with his greatest detrac seyeyed mehdi pirasteh the minister of the tor on bail
1	1	azarmis voice could be heard God bless this wonderful motherinlaw

related to certain themes such as freedom, love, and wealth that carry positive connotations and sentiments. The results show that book authors celebrate their freedom after moving to the west. Freedom has brought happiness, and in the eyes of some of the authors, this happiness is more important than wealth, fame, and beauty. The results reveal that freedom is the source of creativity and vitality.

Based on the top 10 negative sentences shown in Tables 8 and 9 obtained from Vader and textBlob, respectively, the dominant feeling of the Iranians is against torture, which they have experienced during the war, in prison, in the house, or in being raped. It is because women in Iran do not have equal social and political rights (Asl, 2019, 2020, 2021). In addition, the experiences of mental and physical torture have traumatized the victims.

In summarizing the Tables 8 and 9 above, it can be said that Iranian diasporic writers express negative opinions about life in prison. Here, imprisonment not only refers to

Table 8
Top 10 negative sentences generated by Vader

Top Negative Vader		
Score Compound	Sentence	Justification (Theme)
-0.9869	but then others who witness my death or hear about it will know that i died because i refused to give in to hatred and violence and theyll remember and maybe someday theyll find a peaceful way of defeating evil	Last wish of a man who dying soon
-0.9833	she was scared of the snow and disease and loneliness and closed doors and the sulking of her son in law but she wasnt scared of death provided that she wouldnt suffer any pain that is provided that she wouldnt notice that she was dying provided that she died in her sleep	Torture by son in law
-0.9826	emerging from prison opening the enormous metal gate the guard suddenly took away my blindfold and asked me tauntingly if i would recognize my parents	Torture in the prison
-0.9816	as they navigate observing the lonesomeness of the river they are constantly threatened by the danger and violence that emanate like poisonous fumes from the land and its smothery houses the feuding between the seemingly civilized and church going grangerfords and shepherdsons the coldblooded and open killing of a helpless drunk seething mob anger the tarring and feathering of the duke and the dauphin	Murder
-0.9801	did he starve to death or die from lack of water as had imam hussein the third imam of shii muslims who in 681 was killed by yazid the caliph along with seventytwo of his companions because the enemy had deprived them of water for whose thirst and death yousuf used to cry so sincerely	Torture by enemy
-0.9738	the isis chroniclers remained obsessed with religion vividly portraying every atrocity the group committed as some facet of islam rape that longtime tactic of war used to humiliate the enemy became a theology of rape sickness rather than a war crime	Rape was used to humiliate the enemy

Table 8 (continue)

Top Negative Vader		
Score Compound	Sentence	Justification (Theme)
-0.9738	the torturers were taking turns, but he was being interminably beaten either hung from the ceiling so that his feet could not touch the floor and his arms were stretching to the point of being torn from the joints or on the torture bed or on the floor where all of them would attack him as if he were a dangerous animal	Torture
-0.9733	the argument in response often went like this such brutality was certainly not desirable but the west had left the militants no choice there was no other way left to resist nonviolent protest would not sway the dictator assad whose military was torturing and killing scores in detention centers nor would it sway the united states which had invaded and occupied iraq killed countless civilians and sustained and protected arab tyrants	Torture
-0.9727	there was no denying prerevolution corruption or abuse of power by the former regime but she knew that the mullahs brutality in the name of islam far surpassed that of the shahs secret police that they would suppress women gag them under their wretched veils deny them equal rights and put children their softest target in the line of fire in an ideological war	Women got no equal right
-0.9724	this is a problem not just for the uneducated but for trauma victims who have gaps in memory and rape victims who are ashamed	Trauma of victims

Table 9
Top 10 negative sentences generated by TextBlob

Top Negative TextBlob		
Polarity	Subjectivity	Sentence
-1	1	in this case ignorance is not the worst problem either
-1	1	farrukhlaqa asks why if this woman is insane has she not been taken to an asylum
-1	1	we are in a war against evil
-1	1	the records of that terrible war are not classied information
-1	1	i drove people insane with questions i am told
-1	1	something terrible is happening in this country i can smell it in the air and it smells of blood and disaster
-1	1	we have to protect islam gods law and gods people from the evil forces that are at work against them
-1	1	i had tried to accept my situation and to understand him but i couldnt pretend i didnt know about the horrible things he had done
-1	1	youve been up for two days this is insane
-1	1	at the ceremony the scent of jasmine casablanca lilies dahlias and gardenias at the sofreh mingled with the odor of esfand seeds burned to ward o the evil eye

jails and incarceration centers but also the constraining conditions of life within family, community, and religion. For men, the feeling of imprisonment mostly emanates from their experiences in war prisons and punitive facilities, but for women, the feeling mainly originates from a bad family milieu or community injustice where the law does not protect them nor are they granted equal rights (Asl, 2019; Hadi & Asl, 2022). War prisoners suffer from similar feelings of trauma as the physical torture by the enemy has increased the psychological feelings of hopelessness in the captives. Overall, most of the sentences are negative from the extracted text, followed by neutral and positive ones, as shown in Figure 9.

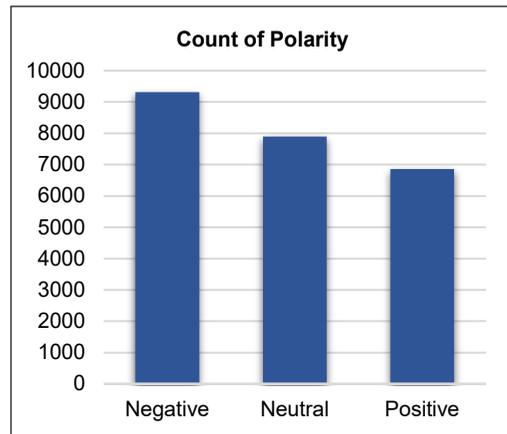


Figure 9. Sentiment bar chart

DISCUSSION

This project analyzed the underlying emotion toward the theme of imprisonment in a corpus of 28 life writings by Iranian diasporic writers. The sentiment attached to the theme has been obtained using sentiment analysis. Two sentiment analyzers from Python libraries have been used: VADER and TextBlob. The output of VADER is more insightful and interpretable. The main contribution to positive sentiment is the feeling of freedom. After the diasporic subjects leave their country of origin, which is struck by a devastating war and an increasing rate of crime and social discrimination, they experience a true sense of freedom that brings joy and happiness. On the other hand, the negative sentiments come from the destructions of the revolution and the eight-year war during which many were captured and tortured as prisoners, as well as the social-political inequalities that deny female citizens of equal rights and protection against patriarchal injustice.

From Topic Modeling, six interrelated themes have been derived from the 10 groups of keywords: sexism, family, revolution, imprisonment, intellectual, and death. A Sentiment Analysis revealed the underlying sentiment of the writers toward each of the themes. Since the 1979 revolution in Iran, the policies and practices of moral purification have generated a new sexual economy that has denied women certain social and political rights. The new sexism, for instance, has restricted women's presence and mobility in public places (Nekai, 2013; Shahrokni, 2020). The exclusionary policies have similarly reinforced the patriarchal power within the family structure to control women. Most women show their discontent with a family structure in which men are in charge of women by both marriage laws and what religious principles indoctrinate (Afary, 2009; Asl, 2018; Asl, 2021). The demand for

women's obedience, modesty, and self-sacrifice concerning their male guardians makes the family structure and milieu restricting and prison-like. The third theme revolves around the revolution and the subsequent individual and collective feelings of trauma. As a result of the revolution, the diasporic Iranian population has suffered painful losses: "the loss of family and friends, the loss of economic and social status, and the loss of their home country" (Naghibi, 2016, p. 4). While the revolution has been formative for people, it has been destructive for some others, as the losses brought about by the revolution have been completely disruptive and upsetting (Vasapollo, 2020). The role of silence and censorship as imposed practices by the state similarly denies intellectuals their right and will to voice out against injustice. Such policies have generated various forms of discontent, hatred, miseries, and despair about the incarcerating socio-political realities (Ranucci, 2019). This strong sense of entrapment and imprisonment is closely tied with the formation and expression of negative sentiments. Finally, the imminent threat of death and annihilation exerts a strong grip on the bodies and minds of Iranians. For a repressive society that seeks collective silencing of its nation, corporeal elimination of individuals is a common practice (Chiaramonte, 2013). Comparing the results of this project with related ethnographical literature makes it clear that the six themes identified in the selected 28 literary writings of diasporic Iranians evoke undesired negative sentiments.

CONCLUSION

This study aimed to propose a text analytics lifecycle for detecting and visualizing the prevailing themes in a corpus of literary texts. As shown in Figure 1, the model was developed in two stages. First, topic modeling techniques with selected algorithms of LDA, LSI, NMF, and HDP are applied to detect the main topics related to the theme of imprisonment in a corpus of 28 Iranian diasporic life writings. LSI is the best algorithm for this study's proposed Topic Modeling model based on the coherence score. By implementing Topic Modeling, six major sub-themes have been derived: sexism, family, revolution, imprisonment, intellectual, and death. Second, sentiment analysis is applied to analyzing the polarity of the diasporic writer's discourse on imprisonment with the selected algorithms. The feeling attached to each of the six topical themes has been visualized by adopting a Sentiment Analysis model. Specifically, two sentiment analyzers from Python libraries, VADER and TextBlob, were adopted (Figure 1). The main contribution to positive sentiment is the feeling of liberation and freedom. On the other hand, the negative sentiments are expressed and directed toward the 1979 revolution, the Iran-Iraq war, and the existing socio-political discrimination and injustice for the disadvantage of women. It is further concluded that the output of VADER is more insightful and interpretable.

The four algorithms compared in this paper can be utilized with different datasets to determine the best model for future works. Besides, this project focused only on published

writings from diasporic Iranian writers. In future studies, text data from social media related to the thematic topics discovered in this project can be the object of analysis. Furthermore, machine learning and deep learning techniques can be utilized to classify author opinions in these books further.

ACKNOWLEDGEMENT

The authors are thankful to School of Computer Sciences and Division of Research & Innovation, Universiti Sains Malaysia for the support from Short Term Grant (304/PKOMP/6315435), which is granted to Ts.Dr. Pantea Keikhosrokiani.

REFERENCES

- Abdelrahman, O., & Keikhosrokiani, P. (2020). Assembly line anomaly detection and root cause analysis using machine learning. *IEEE Access*, 8, 189661-189672. <https://doi.org/10.1109/ACCESS.2020.3029826>
- Afary, J. (2009). *Sexual politics in modern Iran*. Cambridge University Press.
- Al Mamun, M. H., Keikhosrokiani, P., Asl, M. P., Anuar, N. A. N., Hadi, N. H. A., & Humida, T. (2022). Sentiment analysis of the Harry Potter Series using a lexicon-based approach. In P. Keikhosrokiani & M. Pourya Asl (Eds.), *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media* (pp. 263-291). IGI Global. <https://doi.org/10.4018/978-1-7998-9594-7.ch011>
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191. <https://doi.org/10.1177/0047287517747753>
- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, 108, 110-124. <https://doi.org/https://doi.org/10.1016/j.knosys.2016.05.040>
- Asl, M. P. (2018). Practices of counter-conduct as a mode of resistance in Middle East women's life writings. *3L: Language, Linguistics, Literature®*, 24(2), 195-205. <https://doi.10.17576/3L-2018-2402-15>
- Asl, M. P. (2019). Foucauldian rituals of justice and conduct in Zainab Salbi's between two worlds. *Journal of Contemporary Iraq & the Arab World*, 13(2-3), 227-242. https://doi.10.1386/jciaw_00010_1
- Asl, M. P. (2020). Micro-Physics of discipline: Spaces of the self in Middle Eastern women life writings. *International Journal of Arabic-English Studies*, 20(2), 223-240. <https://doi.10.33806/ijaes2000.20.2.12>
- Asl, M. P. (2021). Gender, space and counter-conduct: Iranian women's heterotopic imaginations in Ramita Navai's City of Lies. *Gender, Place & Culture*, 1-21. <https://doi:10.1080/0966369X.2021.1975100>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media, Inc.
- Chiaromonte, P. (2013, January 12). *Hell on earth: Inside Iran's brutal Evin prison*. Fox News. <https://www.foxnews.com/world/hell-on-earth-inside-irans-brutal-evin-prison>
- Costa, L. F. (2018). A method for content analysis applied to newspaper coverage of Japanese personalities in Brazil and Portugal. *Digital Scholarship in the Humanities*, 33(2), 231-247. <https://doi.org/10.1093/lle/fqx050>

- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, 87, 44-49. <https://doi.org/https://doi.org/10.1016/j.procs.2016.05.124>
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 231-240). ACM Publishing. <https://doi.org/10.1145/1341531.1341561>
- Firmin, R. L., Bonfils, K. A., Luther, L., Minor, K. S., & Salyers, M. P. (2017). Using text-analysis computer software and thematic analysis on the same qualitative data: A case example. *Qualitative Psychology*, 4(3), 201-210. <https://doi.org/10.1037/qp0000050>
- Gabrilovich, E., & Markovitch, S. (2006). Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge. In *AAAI* (Vol. 6, pp. 1301-1306). American Association for Artificial Intelligence.
- Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today*, 56, 29-34. <https://doi.org/10.1016/j.nedt.2017.06.002>
- Grayson, S., Mulvany, M., Wade, K., Meaney, G., & Greene, D. (2016, September 20-21). Novel2vec: Characterising 19th century fiction via word embeddings. In *24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'16)*. Dublin, Ireland.
- Grayson, S., Mulvany, M., Wade, K., Meaney, G., & Greene, D. (2017). Exploring the role of gender in 19th century fiction through the lens of word embeddings. In J. Gracia, F. Bond, J. McCrae, P. Buitelaar, C. Chiarcos & S. Hellmann (Eds.), *Language, Data and Knowledge*, (pp. 358-364). Springer. https://doi.org/10.1007/978-3-319-59888-8_30
- Grayson, S., Wade, K., Meaney, G., & Greene, D. (2016). The sense and sensibility of different sliding windows in constructing co-occurrence networks from literature. In B. Bozic, G. Mendel-Gleason, C. Debruyne & D. O'Sullivan (Eds.), *Computational History and Data-Driven Humanities*, (pp. 65-77). Springer.
- Hadi, N. H. A., & Asl, M. P. (2022). The real, the imaginary, and the symbolic: A Lacanian reading of Ramita Navai's City of Lies. *GEMA Online Journal of Language Studies*, 22(1), 145-158. <https://doi.org/10.17576/gema-2022-2201-08>
- Hornick, M. (2017, November 17). Explicit semantic analysis (ESA) for text analytics. *Oracle Machine Learning*. <https://blogs.oracle.com/r/explicit-semantic-analysis-esa-for-text-analytics>
- Jafery, N. N., Keikhosrokiani, P., & Asl, M. P. (2022). Text analytics model to identify the connection between theme and sentiment in literary works: A case study of Iraqi life writings. In P. Keikhosrokiani & M. P. Asl (Eds.), *Handbook of research on opinion mining and text analytics on literary works and social media* (pp. 173-190). IGI Global. <https://doi.org/10.4018/978-1-7998-9594-7.ch008>
- Keikhosrokiani, P., & Asl, M. P. (Eds.). (2022). *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media*. IGI Global. <https://doi.org/10.4018/978-1-7998-9594-7>.
- Leavy, S., Meaney, G., Wade, K., & Greene, D. (2019). Curatr: A platform for semantic analysis and curation of historical literary texts. In E. Garoufallou, F. Fallucchi & E. W. De Luca (Eds.), *Metadata and Semantics Research*, (pp. 354-366). Springer. https://doi.org/10.1007/978-3-030-36599-8_31

- Leavy, S., Meaney, G., Wade, K., & Greene, D. (2020). Mitigating gender bias in machine learning data sets. In L. Boratto, S. Faralli, M. Marras & G. Stilo (Eds.), *Bias and Social Aspects in Search and Recommendation* (pp. 12-26). Springer. https://doi.org/10.1007/978-3-030-52485-2_2
- Lodin, H., & Balani, P. (2017). Rich semantic sentiment analysis using lexicon based approach. *ICTACT Journal on Soft Computing*, 7(04), 1486-1491. <https://doi.org/10.21917/ijsc.2017.0206>
- Malik, E. F., Keikhosrokiani, P., & Asl, M. P. (2021, July 4-5). Text mining life cycle for a spatial reading of Viet Thanh Nguyen's *The Refugees* (2017). In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*. Taiz, Yaman. <https://doi.org/10.1109/ICOTEN52080.2021.9493520>
- Mayo, M. (2017). *A general approach to preprocessing text data*. KDnuggets. <https://www.kdnuggets.com/2017/12/general-approach-preprocessing-text-data.html>
- Mazzola, L., Siegfried, P., Waldis, A., Kaufmann, M., & Denzler, A. (2018, September 25-27). A domain specific ESA inspired approach for document semantic description. In *2018 International Conference on Intelligent Systems (IS)*. Funchal, Portugal. <https://doi.org/10.1109/IS.2018.8710507>
- Misuraca, M., Scepi, G., & Spano, M. (2021). Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback. *Studies in Educational Evaluation*, 68, Article No. 100979. <https://doi.org/10.1016/j.stueduc.2021.100979>
- Mumtaz, D., & Ahuja, B. (2018). A lexical and machine learning-based hybrid system for sentiment analysis. In B. Panda, S. Sharma & U. Batra (Eds.), *Innovations in Computational Intelligence: Best Selected Papers of the Third International Conference on REDSET 2016* (pp. 165-175). Springer. https://doi.org/10.1007/978-981-10-4555-4_11
- Naghbi, N. (2016). *Women Write Iran: Nostalgia and Human Rights from the Diaspora*. University of Minnesota Press.
- Nekai, P. (2013). *From education to segregation: Iran's sexist policy*. PROSPECT <https://prospect-journal.org/2013/01/18/womens-access-to-higher-education-in-iran/>
- Paroubek, P., & Pak, A. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf
- Ranucci, D. (2019). *We never forgot Princeton's Xiyue Wang during his imprisonment in Iran, student says*. <https://www.nj.com/opinion/2019/12/we-must-keep-fighting-for-princeton-grad-student-xiyue-wangs-release-from-an-iranian-prison-opinion.html>
- Řehůřek, R. (2019). *Models.hdpmodel–Hierarchical Dirichlet Process*. Gensim. <https://radimrehurek.com/gensim/models/hdpmodel.html>
- Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks* (pp. 45-50). CiteSeer. <https://doi.org/10.13140/2.1.2393.1847>
- Scharl, A., Hubmann-Haidvogel, A., Jones, A., Fischl, D., Kamolov, R., Weichselbraun, A., & Rafelsberger, W. (2016). Analyzing the public discourse on works of fiction–Detection and visualization of emotion in

- online coverage about HBO's Game of Thrones. *Information Processing and Management*, 52(1), 129-138. <https://doi.org/10.1016/j.ipm.2015.02.003>
- Shahrokni, N. (2020). *Women in Place: The Politics of Gender Segregation in Iran*. University of California Press.
- Shi, T., Kang, K., Choo, J., & Reddy, C. K. (2018). Short-text topic modeling via non-negative matrix factorization enriched with localword-context correlations. In *Proceedings of the 2018 World Wide Web Conference* (pp. 1105-1114). ACM Publishing. <https://doi.org/10.1145/3178876.3186009>
- Sofian, N. B., Keikhosrokiani, P., & Asl, M. P. (2022). Opinion mining and text analytics of reader reviews of Yoko Ogawa's *The Housekeeper and the Professor* in Goodreads. In P. Keikhosrokiani & M. P. Asl (Eds.), *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media* (pp. 240-262). IGI Global. <https://doi.org/10.4018/978-1-7998-9594-7.ch010>
- Subramanian, D. (2019). *Text Mining in Python: Steps and Examples*. Medium. <https://medium.com/towards-artificial-intelligence/text-mining-in-python-steps-and-examples-78b3f8fd913b>
- Suhendra, N. H. B., Keikhosrokiani, P., Asl, M. P., & Zhao, X. (2022). Opinion mining and text analytics of literary reader responses: A case study of reader responses to KL Noir volumes in Goodreads using sentiment analysis and topic. In P. Keikhosrokiani & M. P. Asl (Eds.), *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and SocialMedia* (pp. 191-239). IGI Global. <https://doi.org/10.4018/978-1-7998-9594-7.ch009>
- Sukhija, N., Tatineni, M., Brown, N., Van Moer, M., Rodriguez, P., & Callicott, S. (2016). Topic modeling and visualization for big data in social sciences. In *2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)* (pp. 1198-1205). IEEE Publishing. <https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0183>
- Teoh, Y. Z. I., & Keikhosrokiani, P. (2020). Knowledge workers mental workload prediction using optimised ELANFIS. *Applied Intelligence*, 51, 2406-2430. <https://doi.org/10.1007/s10489-020-01928-5>
- Vasapollo, S. (2020). *Causes of the Iranian 1979 Revolution: Historical and Political Aspects*. ASERI.
- Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(6), 282-292.
- Ying, S. Y., Keikhosrokiani, P., & Asl, M. P. (2021). Comparison of data analytic techniques for a spatial opinion mining in literary works: A review paper. In F. Saeed, F. Mohammed & A. Al-Nahari (Eds.), *Innovative Systems for Intelligent Health Informatics. IRICT 2020. Lecture Notes on Data Engineering and Communications Technologies* (pp. 523-535). Springer. https://doi.org/10.1007/978-3-030-70713-2_49
- Ying, S. Y., Keikhosrokiani, P., & Asl, M. P. (2022). Opinion mining on Viet Thanh Nguyen's the sympathizer using topic modelling and sentiment analysis. *Journal of Information Technology Management*, 14(Special Issue), 163-183. <https://doi.org/10.22059/jitm.2022.84895>



Data Safety Prediction Using Bird's Eye View and Social Distancing Monitoring for Penang Roads

Lek Ming Lim¹, Majid Khan Majahar Ali^{1*}, Mohd. Tahir Ismail¹ and Ahmad Sufriil Azlan Mohamed²

¹*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

²*School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

ABSTRACT

In terms of fatalities, Malaysia ranks third among ASEAN countries. Every year, there is an increase in accidents and fatalities. The state of the road is one factor contributing to near misses. A near miss is an almost-caused accident, an unplanned situation that could result in injury or accidents. The Majlis Bandar Pulau Pinang (MBPP) has installed 1841 closed-circuit television (CCTV) cameras around Penang to monitor traffic and track near miss incidents. When installing CCTVs, the utilisation of video allows resources to be used and optimised in situations when maintaining video memories is difficult and costly. Highways, industrial regions, and city roads are the most typical places where accidents occur. Accidents occurred at 200 per year on average in Penang from 2015 to 2017. Near misses are what create accidents. One of the essential factors in vehicle detection is the "near miss." In this study, You Only Look Once version 3 (YOLOv3) and Faster Region-based Convolutional Neural Network (Faster RCNN) are used to solve transportation issues. In vehicle detection, a faster RCNN was used. Bird's Eye View and Social Distancing Monitoring are used to detect the only vehicle in image processing and observe how near misses occur. This experiment tests different video quality and lengths to compare

test time and error detection percentage. In conclusion, YOLOv3 outperforms Faster RCNN. In high-resolution videos, Faster RCNN outperforms YOLOv3, while in low-resolution videos, YOLOv3 outperforms Faster RCNN.

ARTICLE INFO

Article history:

Received: 29 September 2021

Accepted: 29 December 2021

Published: 15 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.15>

E-mail addresses:

limlekming@gmail.com (Lek Ming Lim)

majidkhanmajaharali@usm.my (Majid Khan Majahar Ali)

m.tahir@usm.my (Mohd. Tahir Ismail)

sufriil@usm.my (Ahmad Sufriil Azlan Mohamed)

* Corresponding author

Keywords: Bird's eye view, near miss, social distancing monitoring vehicle detection

INTRODUCTION

Road traffic accidents are the leading cause of death among adolescents globally. Road traffic accidents are now the eighth leading cause of death in all age groups worldwide, and they are expected to become the seventh leading cause of death by 2030 (World Health Organization, 2015).

The main concern with the transportation issue is the accuracy of recent (real-time based) and reliable data. Currently, Penang is overly reliant on manually observing data instead of automatically calculating real-time data. The baseline data, POL 37, a police report, is the only manual confidential data that most insurance companies will use to process claims for their clients. However, the data could not be used to determine the road condition because there was insufficient information due to erroneous and missing data.

Majlis Bandar Pulau Pinang (MBPP) already installed 534 Closed-circuit television (CCTV) cameras in 2015 to assist the Council in investigating the condition of road illumination, prohibited dumping activities, and activities on the prohibited ground in the hills. However, the images from all installed cameras are of poor quality. MBPP added 1841 high-resolution CCTV cameras (on both Mainland and Island) to address these concerns in 2019. One of the difficulties discussed concerning CCTV is storage, as MBPP can only preserve a video for 45 days before it is automatically removed. Despite the installation of CCTV cameras, the higher authorities lack an algorithm for calculating and detecting vehicles. Many different types of vehicles complicate the vehicle counting algorithm. No algorithm can be used to implement vehicle detection on a specific road.

Furthermore, POL 37 is a hardcopy police report, which does not assist in the automatic counting of near misses due to the conversion limitations into a visualising report. Rather than relying on hardcopy data, different means must be used. There is no autonomous near miss counting algorithm that can be employed. Due to the lack of research, it is hard to simultaneously calculate near misses from CCTV.

Near miss is one of the issues that the Penang 2030 mission will address to lessen the likelihood of a near miss. To begin, locate near misses on a specific road. Then, because near misses are the causes of accidents, the near misses report and investigation can help to improve road safety. According to the safety triangles or the Heinrich 300-29-1 model, for every 300 near misses, there are 29 minor accidents and one major accident. In addition, accidents raise carbon emissions, which harm the environment. Therefore, the ultimate goal of reporting a near miss is to resolve the incident and take preventative measures to ensure that it does not happen again.

A near miss is an unanticipated event that led to the investigation into the cause of the Malaysian accident. In each of the cases, there was the possibility of an accident occurring, but due to fortunate circumstances, the loss was avoided. According to the researchers, for every significant occurrence, there is a chance of 10 minor incidents and up to 100

near misses (Silva Consultants, 2016). According to Aldred (2016), near miss analyses could reveal information about cyclists' experiences with difficulties related to road user behaviour, culture, and cycling infrastructure. The study also concluded that, based on the experience of near misses, the number of injury incidents could be reduced compared to common types of accidents.

Nevertheless, near misses have sparked deep concern, and they have the potential to be used to investigate factors affecting pedestrian safety. There are numerous crash types in a mixed traffic flow scenario. According to Wang et al. (2020)'s research, low-visibility conditions such as heavy rainy days, foggy days, and nights with insufficient lighting could cause a near miss in traffic flow. A vision-based crash detection framework recognises various objects and crash types from images.

Previous research in Appendix A reveals that there are still some gaps in this study. There are no studies that combine the near miss and vehicle detection applications. Furthermore, most studies employ a survey or questionnaire method to ease data collection. It is due to a lack of accident reports and data. To complete the data, they can only rely on people's experiences. In this study, Social Distancing Monitoring and Bird's Eye View are used in vehicle identification to analyse images and identify near misses.

Vehicle detection is a method for estimating accidents between vehicles and can aid in observing the entire near miss process. Arinaldi (2018) has presented a traffic video system based on the visualisation method that uses important statistics. Vehicle counting, vehicle type, estimated vehicle speed, and vehicle lane change are all included in the statistics. In the previous study, most researchers chose image processing to detect vehicles. Researchers link the model or algorithm method and the software method in their studies. The key reason is that the findings will be visible in a monitoring system. Monitoring systems assist researchers in analysing data from image or video processing. A good strategy to monitor vehicle detection in this study is to use Social Distancing Monitoring and Bird's Eye View methods. Furthermore, these two methods can assist in obtaining accurate and real-time data.

Object detection and vehicle tracking employ a wide range of models and software. CNN, RCNN, Fast RCNN, Faster RCNN, and YOLO are a few examples. However, the models are not used in most near miss studies to estimate the near miss. Zohra et al. (2018) have proposed CNN to improve accuracy and reduce error in the proposed system. Ciberlin et al. (2019) have employed YOLOv3 as object detectors for detection and tracking. Ding and Yang (2019) have demonstrated YOLOv3 to locate parking lots for vehicle detection. The experiment method can improve parking lot detection accuracy while reducing error detection. Huang et al. (2020) have presented traffic flow detection using YOLOv3 and Faster RCNN on 40-second video datasets. Bull et al. (2017) have used Faster RCNN and YOLOv3 to count vehicles in Kuala Lumpur, Malaysia.

In this study, the Social Distancing Monitoring and Bird’s Eye View are used in YOLOv3 to perform image processing on vehicle detection to deal with large and high-resolution images or videos. Near miss accidents can be recorded and observed using image processing and vehicle detection. These analyses can forecast future near miss accidents and pinpoint the source of the problems.

MATERIALS AND METHODS

Convolutional Neural Network

Convolutional Neural Networks (CNN) are highly popular in the deep learning community. These CNN models are used in various applications, the most common of which are image and video processing projects (Rawat & Wang, 2017). A CNN is a computational model that also employs several layers of neurons and is composed of one or more convolutional layers that can be completely linked or pooled. Furthermore, these convolutional layers provide feature maps that record an area of the picture, divided into rectangles and sent out for non-linear processing (Qin et al., 2018).

Figure 1 shows the operation of a convolutional neural network. First, provide the input image. Next, put the image into the convolution layer. Then, perform pooling to reduce the dimensionality. Finally, examine the outcomes to see if they meet expectations. Otherwise, return to the convolution layer and display the results.

According to Behl et al. (2014), convolution is a mathematical operation performed on two functions (f and g) to generate a third function defined as ($f * g$) that explains how the form of one is changed by the other in Equation 1.

$$(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau = \int_{-\infty}^{\infty} f(t - \tau)g(\tau)d\tau \tag{1}$$

Although the symbol t as in Equation 1, does not always indicate the time domain. The convolution formula can be written as a weighted average of the function $f(\tau)$ at time t , where $g(-\tau)$ represents the weight, and just the amount t is shifted. The weighting function emphasises certain aspects of the input function as t changes. As a result, the integration

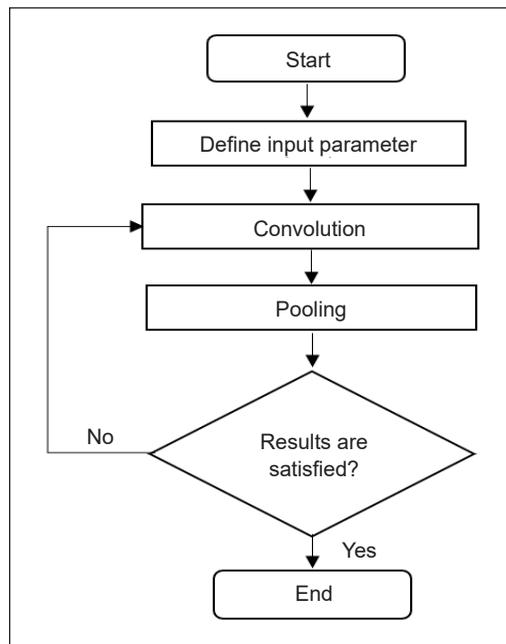


Figure 1. CNN flow chart

limitations can be shortened. Functions that are only supported on (for example, 0 for negative arguments), resulting in Equation 2:

$$(f * g)(t) := \int_0^t f(\tau)g(t - \tau)d\tau \tag{2}$$

Convolutional neural networks employ many cascaded convolution kernels in machine vision and artificial intelligence applications. CNN models have a consistent structure that consists of alternating convolutional layers and pooling layers (typically, each pooling layer is arranged after a convolutional layer), which serve as feature extraction.

According to Albelwi and Mahmood (2017), the last layers consist of a limited number of fully connected layers, with the final layer being a softmax classifier used for image classification. Over K classes, the softmax classifier estimates the posterior probability of each class label.

$$y_i = \frac{\exp(-z_i)}{\sum_{j=i}^K \exp(z_j)} \tag{3}$$

Where z_i is any real number and reflects the softmax function's input values, the bottom part of Equation 3 is the normalisation terms, required to ensure that all the function's output values add up to i .

The term "hyperparameters" refers to numerous settings that influence learning. For example, CNN employs more hyperparameters than a typical multilayer perceptron (MLP). In the CNN architecture design, the algorithm for the CNN is described by a structural hyperparameter λ that encompasses the CNN architecture design is shown as in Equation 4:

$$\lambda = \left((\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i)_{i=1, M_C}, \dots, (\lambda_j^i)_{j=1, N_f} \right) \tag{4}$$

where $\lambda \in \Psi$ determines the domain for each hyperparameter, M_C is the number of convolutional layers, N_f is the number of fully connected λ_1^i is the number of filters, λ_2^i is the filter size, λ_3^i is the pooling locations and size, and λ_4^i is the stride step.

Social Distancing Monitoring

Social distancing monitoring is a technique for measuring the distance between vehicles and predicting the likelihood of an accident occurring. Calculate the Euclidean distance between all detected boxes and filter out or flag vehicles close to each other, indicating that the vehicles are at risk by using this method.

Figure 2 shows the flowchart for Social Distancing Monitoring. Input the video into the programme as the first step. Then, use object detection to find the single vehicle in the image or frame. The pairwise distances between all detected vehicles should then

be computed. Then, using the horizontal and vertical unit lengths, calculate the distance between the two vehicles. The unit length is measured in centimetres (cm). If the distance is greater than 180 cm, green boxes would appear (safe). Also, the yellow boxes (near miss) are displayed when the distance is between 50 cm and 180 cm; otherwise, red boxes (high risk) are displayed when a distance is less than 50 cm.

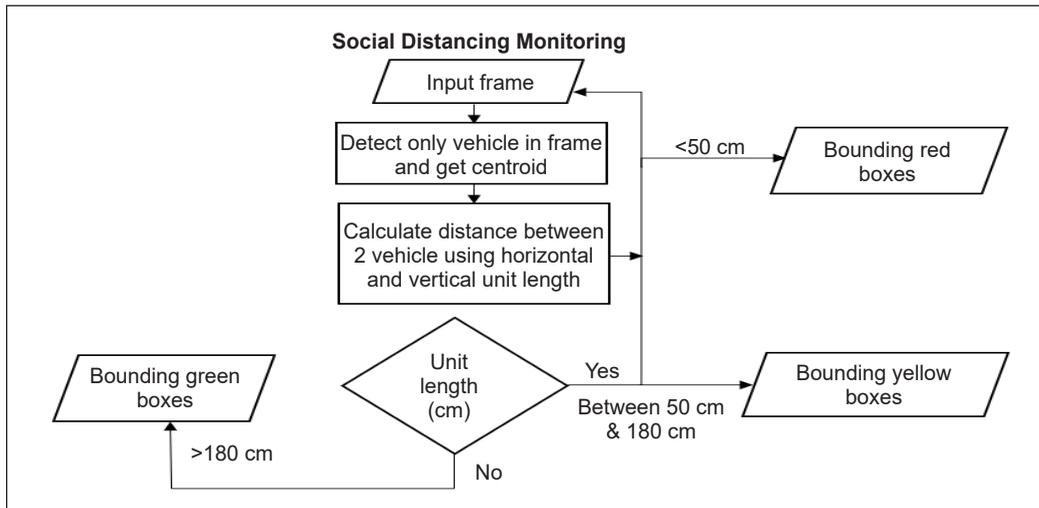


Figure 2. Social distancing monitoring algorithm

Bird's Eye View

The Bird's Eye View method calculates the near miss rate by displaying the distances between points (vehicles) in a specific frame box. The near miss is calculated based on the distance between the vehicles. When the distance between two vehicles is very close to causing an accident, it is called a near miss. The near miss distance between vehicles estimation and how close the threshold for near miss distance between vehicles estimation must be implemented.

Figure 3 shows the flowchart using the Bird's Eye View method. Input the video into the programme as the first step. Then, using object detection, find the only vehicle in the image or frame and calculate the centroid. The project was then discovered from a Bird's Eye View. Then, the horizontal and vertical unit lengths compute the distance between two vehicles. The length of a unit is measured in centimetres (cm). Green dots appear (safe) if the distance is greater than 180 cm. The yellow dots (near miss) appear if the distance is between 50 cm and 180 cm. Otherwise, the red dots (high risk) appear at less than 50 cm.

Figure 4 shows how the ImageJ application can measure the distance between each pixel in the image in real life. In this video, 209.1220 pixels equal 32 rocks, which equals 96 feet, or 2926.08 cm. Hence, the image displays 0.0715 pixel = 1 cm or 0.0715 pixel/cm with a perspective angle ranging from 10 to 15. As a result, the two points represent

the threshold of a near miss between two vehicles. When the two blue points are fixed, the number of pixels fluctuates depending on their width. For example, 12.87 pixels = 180 cm (constant distance).

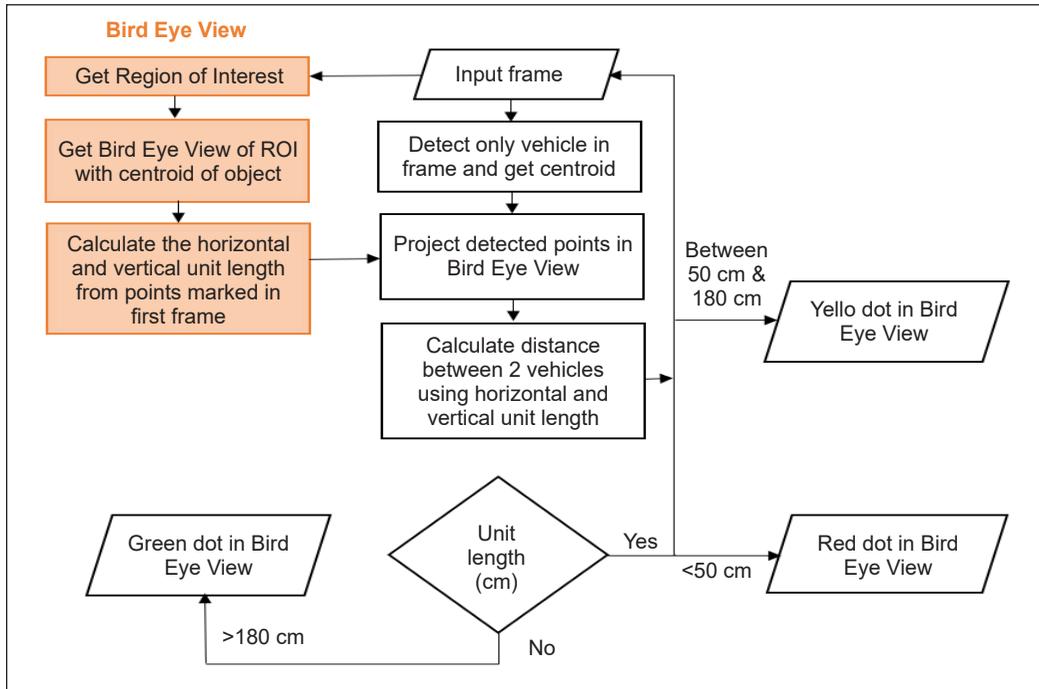


Figure 3. Bird's Eye View algorithm

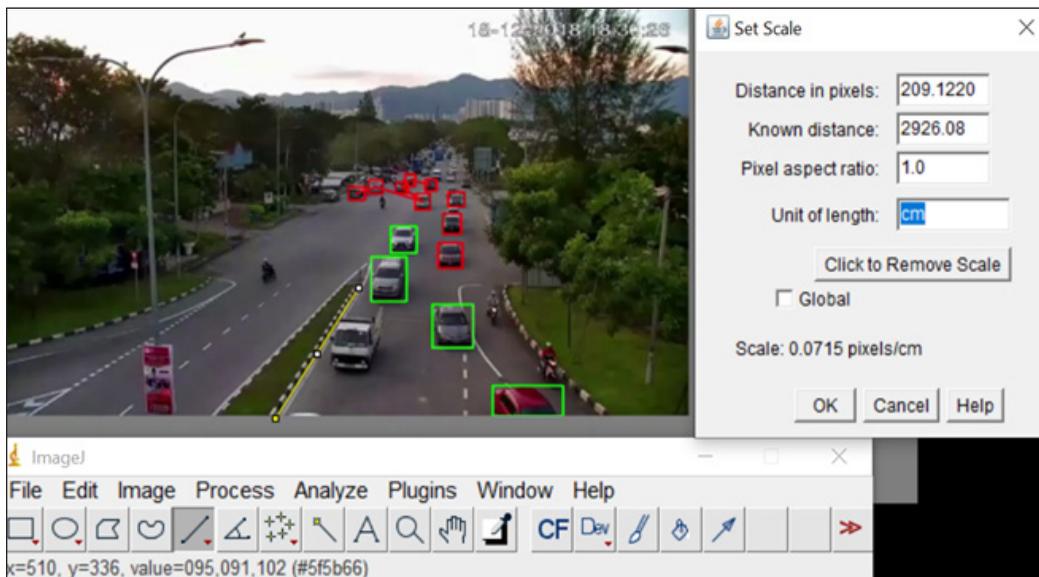


Figure 4. Calculation of unit length in the Tun Dr Lim Chong Eu Expressway video

the threshold of a near miss between two vehicles. When the two blue points are fixed, the number of pixels fluctuates depending on their width. For example, 12.87 pixels = 180 cm (constant distance).

As shown in Figure 5, the ImageJ application measured the distance between pixels in an image in real life (Rahman et al., 2012). In this video, 134.1641 pixels = 12 rocks = 36 feet = 1097.28 cm. As a consequence, the image displays 0.1223 pixels = 1 cm or 0.1223 pixels/cm with a perspective angle of 10 to 15 degrees. For example, 22.014 pixels = 180 cm (constant distance).

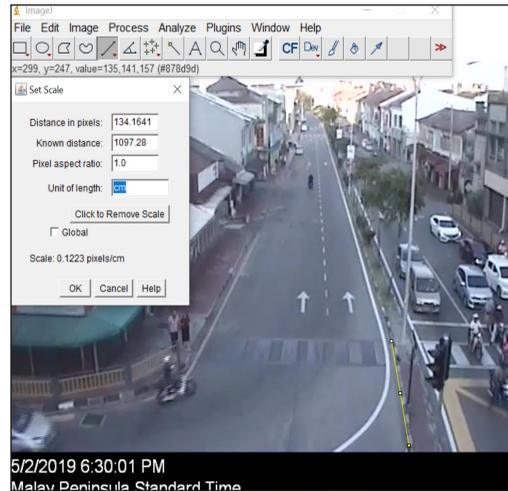


Figure 5. Calculation of unit length in the Lim Chwee Leong Road video

Faster RCNN

Convolutional neural networks (CNN) are mostly used for image classification, whereas RCNN is typically used for object detection. Faster RCNN was developed in June 2015. The selective search is replaced by a region proposal network (RPN) and a detector based on Fast RCNN for more accurate object detection while reducing the number of region proposals.

Faster RCNN is focused on and used to detect vehicles in this study. It is a deep convolutional neural network that detects and classifies objects in images. RPN and Fast RCNN are the two modules that makeup Faster RCNN. RPN generates region suggestions, while Fast RCNN is used to identify objects in the suggested regions. It employs the concept of attention in neural networks via the RPN function, and thus RPN guides the Fast RCNN as a detector to locate objects in images (Gad, 2020).

The datasets are derived from the videos. First, the video is divided into frames. Then label the vehicles in the frames. Next, train the model with the labelled file until the loss functions are less than 0.5. The training frames must account for 40% of the total frames and 60% of the testing frames for detection. Finally, the Faster RCNN is used in the experiment.

RESULTS AND DISCUSSION

YOLOv3 Results

The video Social Distancing Monitoring monitors the distance between vehicles and predicts the likelihood of accidents occurring (Vinitha & Velantina, 2020). In addition, a bird's eye view is used to show the distances between points (vehicles) in a specific frame box and estimate that near misses occur. These two methods are used in this study

to demonstrate how near misses occur and to estimate the distance between vehicles in video and real life.

YOLOv3 is used to detect vehicles for the Social Distancing Monitoring and Bird's Eye View (Ong, 2020). Other datasets, such as motorcycles, people, and trucks, are being filtered out, leaving only cars to be detected. In the video, the results showed red bounding boxes (high-risk accident), yellow bounding boxes (near miss detected), and green bounding boxes (no accident occurred).

Social Distancing Monitoring detects all vehicles in the video, whereas Bird's Eye View uses Region of Interest (ROI) to detect vehicles. Furthermore, the distance settings in Social Distancing and Bird's Eye View are identical. As a result, the outcomes of both methods are the same. The locations of the dots in Bird's Eye View, for example, can be shown in the detected vehicles in Social Distancing Monitoring. Figure 6 depicts the ROI points on Lebuhraya Tun Dr Lim Chong Eu.

Figure 7 shows only vehicle detection using YOLOv3 to detect the likelihood of collisions and near miss between vehicles. Result Tun Dr Lim Chong Eu Expressway videos were shot between 18/12/2018, 6:30:15p.m. and 18/12/2018, 6:31:34p.m. Only cars within the specified rectangle zone would be displayed in the bird's eye view. Due to the poor viewpoint, results outside the rectangular zone are ignored, and only cars within the ROI are recorded.



Figure 6. The ROI for Tun Dr Lim Chong Eu Expressway and the threshold ratio

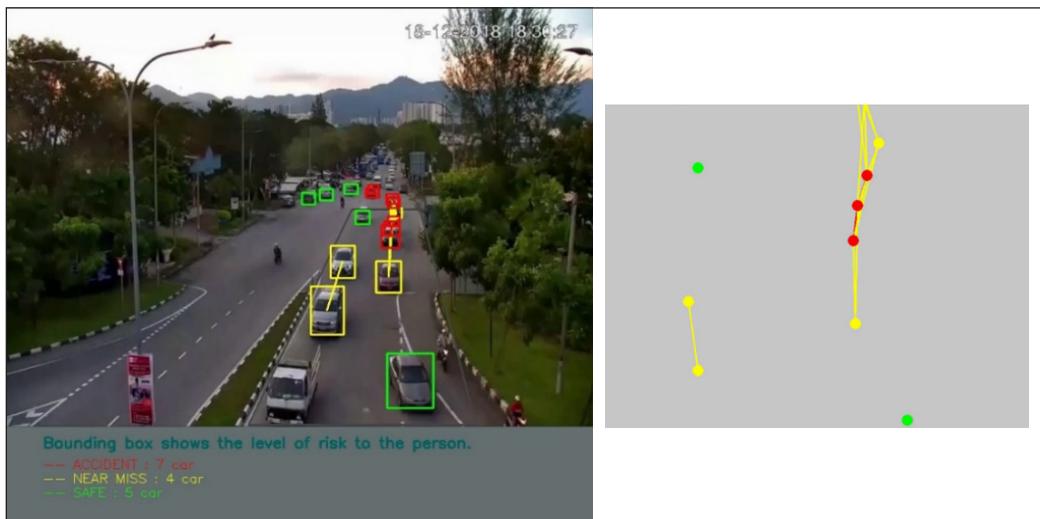


Figure 7. Tun Dr Lim Chong Eu Expressway near miss detection result

Table 1 compares different lengths of the same video in the 20s, 40s, 60s, and 80s of high-quality videos in YOLOv3. Three videos with a 10-minute duration and 30 frames per second were taken from YouTube to conduct the experiments in Sonnleitner et al. (2020)'s study. The running time increases as the length of the video increases. It is one of the reasons why a 20-second video was chosen for this study instead of an 80-second video because the computational time for the image process using YOLOv3 is too long for the computational process. Although the data in the 80-second video is more reliable than the data in the 20-second video, manually counting data is a major task for data recording—the total number of frames taken increases as the video length increases.

As the duration of videos in image processing increases, error detection, near miss detection, and accident detection can be observed and recorded in greater detail. Motorcycles, trucks, and lorries are recognised as vehicles in error detection. The percentage of near miss detection and accident detection in the 20s video is higher than in the 80s video. It demonstrates that the vehicles in the first 20 seconds of the video are too near to each other when stopping in front of a traffic light or when there is congestion. After the 20s, the vehicles began to move and maintain a safe distance from one another. As a result, near miss and accident detection percentages were reduced from 100% to 84.61% and 33.33% to 28.44%.

Table 1
Comparison of high-quality videos in YOLOv3

CCTV time	6:30:15 p.m. – 6:30:34 p.m.	6:30:15 p.m. – 6:30:54 p.m.	6:30:15 p.m. – 6:31:14 p.m.	6:30:15 p.m. – 6:31:34 p.m.	
Length of video	20s	40s	60s	80s	
Computational time	445 seconds	838 seconds	1655 seconds	2501 seconds	
Total number of frames	600	1198	1798	2398	
Error detection	Number of frames	10	106	120	226
	Percentage	$10/600 \times 100\% =$ 1.67%	$106/1198 \times 100\%$ = 8.85%	$120/1798 \times 100\%$ = 6.68%	$226/2398 \times 100\%$ = 9.42%
	Object detected	Motorcycle	Motorcycle, truck	Motorcycle, truck	Motorcycle, truck, lorry
Near Miss detection	Number of frames	600	1153	1734	2029
	Percentage	$600/600 \times 100\%$ = 100%	$1153/1198 \times$ $100\% =$ 96.24%	$1734/1798 \times$ $100\% =$ 96.44%	$2029/2398 \times$ $100\% =$ 84.61%
Accident detection	Number of frames	200	392	385	682
	Percentage	$200/600 \times 100\%$ = 33.33%	$392/1198 \times 100\%$ = 32.72%	$385/1798 \times 100\%$ = 21.41%	$682/2398 \times 100\%$ = 28.44%

Figure 8 shows the ROI points on Lim Chwee Leong Road. In contrast, Figure 9 shows only vehicle detection using YOLOv3 to detect the likelihood of collisions and near miss between vehicles. Result Lim Chwee Leong Road videos were shot between 5/2/2019, 6:30:15p.m. and 5/2/2019, 6:31:34p.m. The Bird's Eye View only shows vehicles within the specified rectangular area. Results outside the rectangular area are ignored as the viewpoint is weak and only cars within the ROI are counted.

Table 2 compares different lengths of the same video, which are in the 20s, 40s, 60s, and 80s of low-quality videos in YOLOv3. The computing time increases as the length of the video increases. It is one of the reasons why a 20-second video was chosen for this study instead of an 80-second video because the computational time for the image process using YOLOv3 is too long for the computational process. Therefore, in this study, image processing was carried out using videos with a length of 80 seconds, which is more trustworthy than videos of 20 seconds. However, the problem of counting data in detection was manually counted, which took a long time and was a significant advancement—the total number of frames taken increases with the length of the video.

As the duration of videos processed in image processing increases, error detection, near miss detection, and accident detection can be observed and recorded in greater detail. Motorcycles, tricycles, and buses are identified as vehicles in error detection. From the 20s video to the 80s video, the percentage of near miss detection and accident detection decreased from 37.63% to 27.23% and 1.17% to 0.5%, respectively.



Figure 8. Result Lim Chwee Leong Road ROI and set threshold ratio

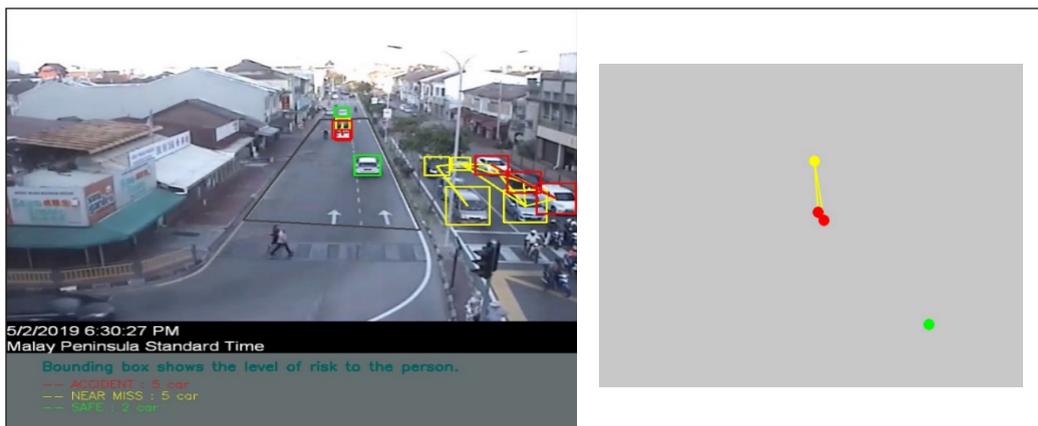


Figure 9. Result Lim Chwee Leong Road near miss detection

In this low-quality video (Lim Chwee Leong Road), near miss and accident probability is lower than in Tun Dr Lim Chong Eu Expressway's video.

Table 2
Comparison of low-quality videos in YOLOv3

CCTV time	6:30:15 p.m.– 6:30:34 p.m.	6:30:15 p.m.– 6:30:54 p.m.	6:30:15 p.m.– 6:31:14 p.m.	6:30:15 p.m.– 6:31:34 p.m.	
Length of video	20s	40s	60s	80s	
Computational time	572 seconds	1255 seconds	2083 seconds	2680 seconds	
Total number of frames	598	1198	1798	2398	
Error detection	Number of frames	2	120	154	
	Percentage	$2/598 \times 100\% =$ 0.33%	$120/1198 \times 100\% =$ 10.02%	$154/1798 \times 100\% =$ 8.57%	$154/2398 \times 100\% =$ 6.42%
	Object detected	Motorcycle	Motorcycles, tricycle, bus	Motorcycles, tricycle, bus	Motorcycles, tricycle, bus
Near Miss detection	Number of frame detection	225	406	438	509
	Percentage	$225/598 \times 100\% =$ 37.63%	$406/1198 \times 100\% =$ 33.89%	$438/1798 \times 100\% =$ 24.36%	$509/2398 \times 100\% =$ 21.23%
Accident detection	Number of frame detection	7	11	12	12
	Percentage	$7/598 \times 100\% =$ 1.17%	$11/1198 \times 100\% =$ 0.92%	$12/1798 \times 100\% =$ 0.66%	$12/2398 \times 100\% =$ 0.5%

Huang et al. (2020) collected data from three videos shot in the 1940s for their paper. The videos show various weather and scenario scenarios. For video traffic monitoring, the YOLOv3 algorithm is used. They also mention the location of the video collection and the field of vision. If the video collection is too large, this will result in missed detection. Error detection occurs when the field of view is too large. The experiment used video lengths from the 20s, 40s, 60s, and 80s in this study. Weather, other scenarios, and camera location could all play a role in future work.

Furthermore, Cepni et al. (2020) presented video collected from UAVs (crewless aerial vehicles) with a resolution of 1280x720 and video collected at a terrestrial quality of 1080x1920, which was tested in the experiment. The videos are 24 fps and a one-minute-long video. When comparing both videos, terrestrial videos outperform model accuracy and estimation. Therefore, two videos are tested in this study: high-quality (Tun Dr Lim Chong Eu Expressway) and low-quality videos (Lim Chwee Leong Road).

Figure 10 shows the test time in high-quality and low-quality videos. Low-quality videos require more computational time than high-quality videos.

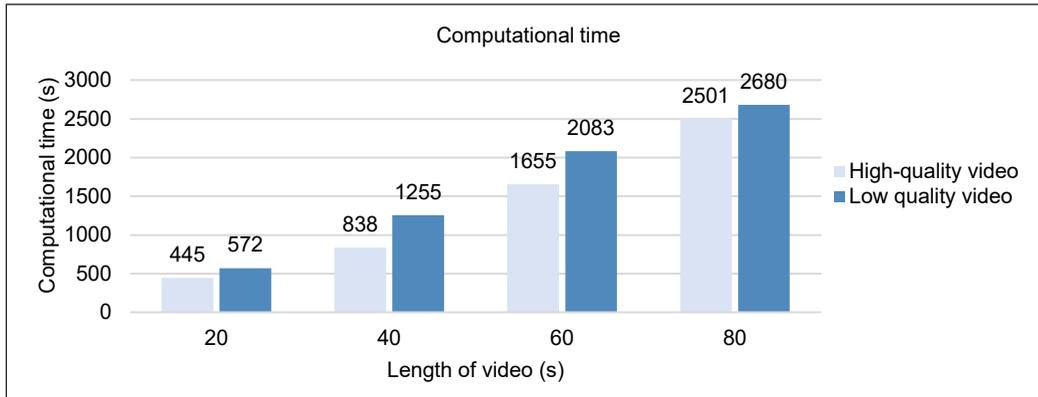


Figure 10. Comparison of computational time between high-quality videos and low-quality videos in YOLOv3

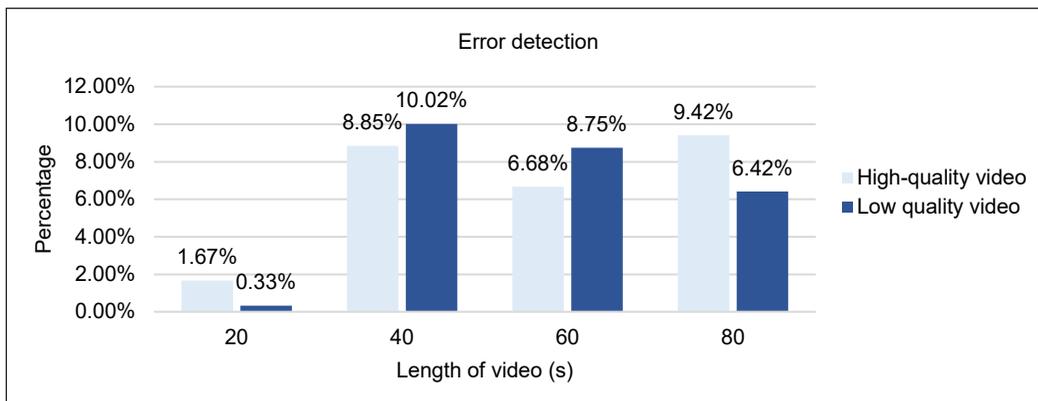


Figure 11. Comparison of error detection between high-quality videos and low-quality videos in YOLOv3

Figure 11 represents error detection in high-quality and low-quality videos. In YOLOv3, error detection occurred, such as detecting a motorcycle, a truck, and a lorry in high-quality videos and detecting motorcycles, a tricycle, and a bus as the car. A proportion of error analysis for YOLO and Fast RCNN is presented in Dixit et al.'s (2019) paper. Compared to Fast-RCNN, most of the errors in YOLO are localisation errors.

Figure 12 presents the detection of errors in high-quality and low-quality videos. This analysis supports the theory that near miss occurs during the black spot's peak hour. Aside from driver behaviour, the root cause of near misses is an enormous issue. According to Matsui et al.'s (2013) research, the development of driving safety devices necessitates detailed functions of the contact scene between the car and the pedestrian to reduce the number of fatalities and the severity of injuries in Japan. However, due to a lack of data from real-world accidents, the researchers focused on near miss situations. As a result, in the video taken from the black spot area, Social Distancing Monitoring and Bird's Eye View are used to detect the near miss.

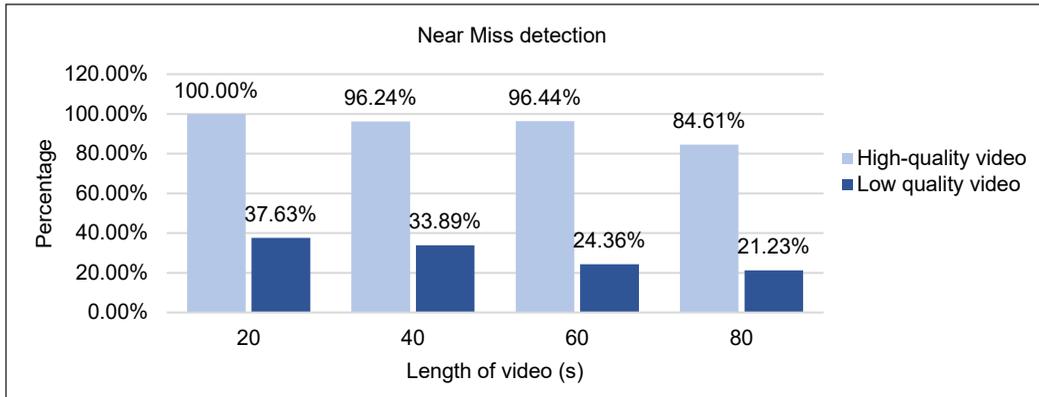


Figure 12. Comparison of near miss detection between high-quality videos and low-quality videos in YOLOv3

According to Vinita and Velantina’s (2020) research, social distance detection tools can monitor the public to keep themselves safe from other people using video from surveillance cameras. By combining Social Distancing Monitoring and Bird’s Eye View, the experiment sheds new light on vehicle detection. The methods described above also detect vehicles at a safe distance to avoid near misses or accidents and do not exceed the threshold ratio set in the algorithm.

The videos provided by MBPP vary in quality depending on location. They are not in the same location and have different video quality. In the introduction, MBPP installed 534 CCTV cameras in 2015 and 1841 CCTV cameras in 2019. Some locations installed CCTV cameras in 2015, but some places are not good quality compared to good quality places.

Faster RCNN Results

Faster RCNN also detected vehicles on Tun Dr Lim Chong Eu Expressway and Lim Chwee Leong Road. The videos on Tun Dr Lim Chong Eu Expressway are high quality, whereas the videos on Lim Chwee Leong Road are low quality. Both videos are tested in various lengths, including the 20s, 40s, 60s, and 80s.

Figure 13 captures the detection of a vehicle on Tun Dr Lim Chong Eu Expressway. Fast RCNN detects only cars in the video and uses green bounding boxes to show the percentage similarity of objects in the dataset.



Figure 13. Result of Tun Dr Lim Chong Eu Expressway vehicle detection

Table 3 compares the lengths of the video Tun Dr Lim Chong Eu

Expressway, covering the 20s, 40s, 60s, and 80s of high-quality videos in Faster RCNN. The computational time required increases in proportion to the length of the video. Therefore, compared to an 80-second video, this study chose a 20-second video duration because the computational time is too long for the vehicle detection process when using Faster RCNN. The total number of frames taken increases with the length of the video. The dataset from the video frames is used to train a Faster RCNN. As a result, when compared to YOLO, the computational time is lengthy (Alganci et al., 2020). In this video, Faster RCNN outperforms YOLO compares the accuracy and speed of the Faster RCNN, SSD, YOLO, and NVIDIA (Dixit et al., 2019). Faster-RCNN is the most accurate algorithm with the slowest speed, whereas YOLO is super-fast with low accuracy. In this study, the number of frame errors in detection is less than 1% for four videos of varying lengths. The object that caused the error is a banner.

Table 3
Comparison of high-quality videos in Faster RCNN

CCTV time	6:30:15 p.m. – 6:30:34 p.m.	6:30:15 p.m. – 6:30:54 p.m.	6:30:15 p.m. – 6:31:14 p.m.	6:30:15 p.m. – 6:31:34 p.m.	
Length of video	20s	40s	60s	80s	
Computational time	747 seconds	1496 seconds	2272 seconds	3111 seconds	
Total number of frames	599	1200	1798	2399	
Error detection	Number of frames	3	3	3	
	Percentage	$3/599 \times 100\% = 0.5\%$	$3/1200 \times 100\% = 0.25\%$	$3/1798 \times 100\% = 0.17\%$	$3/2399 \times 100\% = 0.12\%$
	Object detected	Banner	Banner	Banner	Banner

Figure 14 captures the outcome of the Lim Chwee Leong Road vehicle detection. Fast RCNN detects only vehicles in the video and uses green colour bounding boxes to show the percentage similarity of objects in the dataset. When executing the Faster RCNN, the error detection is proven in the video.

Table 4 compares the lengths of the video Tun Dr Lim Chong Eu Expressway, covering the 20s, 40s, 60s, and 80s of low-quality videos in Faster RCNN. When the video's duration is longer, the amount of computing time required increases. Compared to an 80-second video, this study chose a 20-second video



Figure 14. Result Lim Chwee Leong Road vehicle detection

duration because the computational time for the vehicle detection using Faster RCNN is too long for the computational process. The total number of frames taken increases as the video length increases. Faster RCNN is trained using the video frames as a dataset. Therefore, compared to YOLO, the computational time is lengthy (Srivastava et al., 2021). For the four videos of varying lengths, the number of frame errors in detection is 100%. The roof, unknown objects, and the bus are error-detected objects. Due to the low-quality video (Aqqa et al., 2019) and small dataset (Cao et al., 2019) used to train Faster RCNN, it cannot detect vehicles as accurately as YOLO in Table 4.

Table 4
Comparison of low-quality videos in Faster RCNN

CCTV time	6:30:15 p.m. – 6:30:34 p.m.	6:30:15 p.m. – 6:30:54 p.m.	6:30:15 p.m. – 6:31:14 p.m.	6:30:15 p.m. – 6:31:34 p.m.
Length of video	20s	40s	60s	80s
Computational time	757 seconds	1642 seconds	2449 seconds	3218 seconds
Total number of frames	599	1200	1798	2398
Number of frames	599	1200	1798	2398
Percentage	$599/599 \times 100\%$ = 100%	$1200/1200 \times$ $100\% = \mathbf{100\%}$	$1798/1798 \times$ $100\% = \mathbf{100\%}$	$2398/2398 \times$ $100\% = \mathbf{100\%}$
Object detected	Roof, unknown object	Roof, unknown object, bus	Roof, unknown object, bus	Roof, unknown object, bus

According to Cao et al. (2019), the researchers proposed an improved Faster RCNN and used the TT100K (Tsinghua-Tencent 100K) dataset, which saves 100,000 images, including 30,000 traffic-sign occurrences. Then, compare it to other research papers that also use Faster RCNN. However, only 400 video frame samples were labelled and trained for this study. Faster RCNN requires more sample data to train the algorithm.

The KITTI dataset trains the Faster RCNN in Zhang et al.'s (2018) research. The test video was shot on a real-life road in 1280x720 resolution. The dataset for this study was derived from CCTV video, including the training and labelling images. Following that, the images from the experiment are used to train the model.

Figure 15 represents the test time in high-quality and low-quality videos. Low-quality videos require more computational time than high-quality videos.

Figure 16 shows the detection of errors in high-quality and low-quality videos. According to Aqqa et al. (2019), video quality is an important factor, often overlooked. The video is tested using Faster RCNN, SSD, YOLO, and RetinaNet for object detection at various video compression levels to investigate the quality distortion caused by compression artefacts during video capture. In this study, low-quality video is tested using YOLOv3 and Faster RCNN. There are numerous errors found in the results.

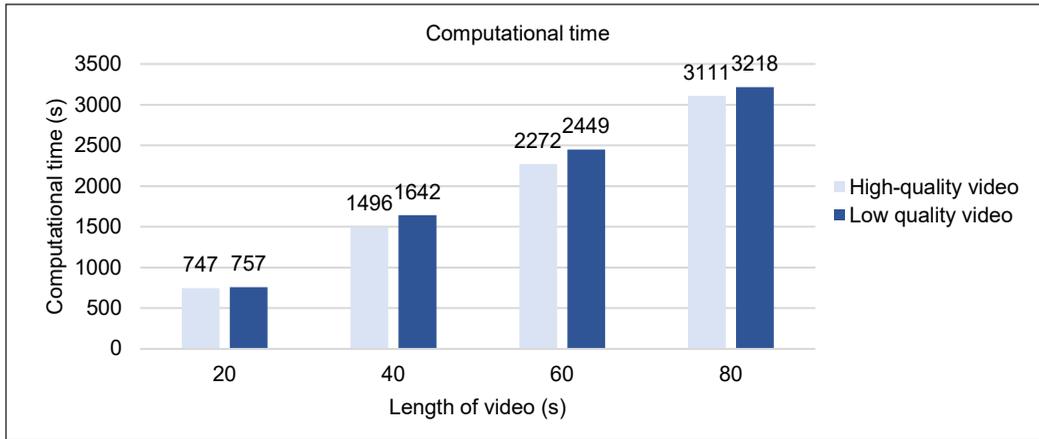


Figure 15. Comparison of computational time between high-quality videos and low-quality videos in Faster RCNN

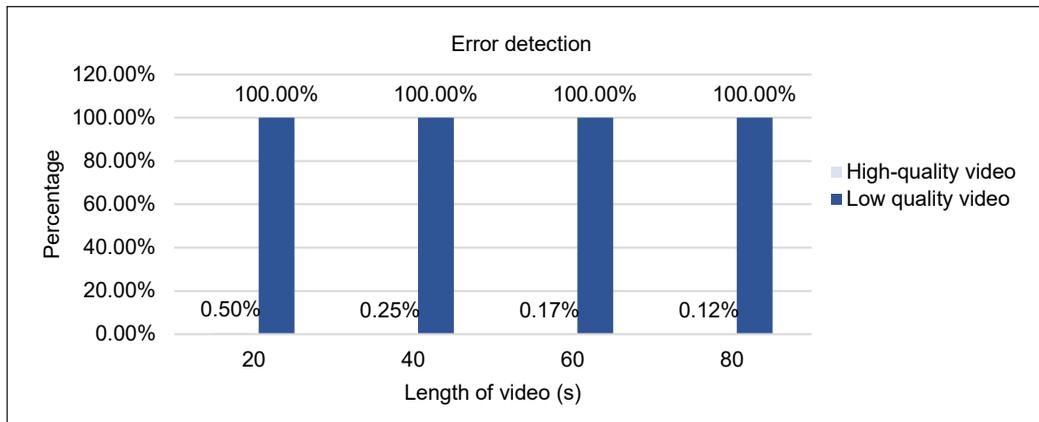


Figure 16. Comparison of error detection between high and low-quality videos in Faster RCNN

Comparison Between YOLOv3 and Faster RCNN

High-quality (Tun Dr Lim Chong Eu Expressway) and low-quality (Lim Chwee Leong Road) videos were tested in YOLOv3 and Faster RCNN. The similarities between YOLOv3 and Faster RCNN are that they use anchor boxes based on network structure, bounding boxes, and the same length of videos for the experiment. However, there are also differences between YOLOv3 and Faster RCNN. Table 5 compares YOLOv3 and Faster RCNN in high-quality videos. The following comparison is based on Tables 1 and 3.

The video provided by the Penang state government impacts the reliability of this data. However, the video contains reliable information that can be saved as historical data for future use. According to Calles et al. (2017), near miss cases for transportation provide a good opportunity to determine whether there are problems and intervene before the actual accident. Behaviour and driving experience are two factors that can explain the risk of near

Table 5
Comparison in high-quality videos

YOLOv3	Difference	Faster RCNN
Social Distancing Monitoring and Bird's Eye View	Method	Vehicle detection
Fast	Speed	Slow
Higher than Faster RCNN	Error detection	Low
Very high	Near Miss detection	null
Low	Accident detection	null

Table 6
Comparison in low-quality videos

YOLOv3	Difference	Faster RCNN
Social Distancing Monitoring and Bird's Eye View	Method	Vehicle detection
Fast	Speed	Slow
Low	Error detection	Very high
High	Near Miss detection	null
Very low	Accident detection	null

miss accidents among young drivers. Collaboration with hospitals and insurance companies should be considered in future studies to obtain more complete data and reduce data flaws.

Table 6 shows a comparison in low-quality videos of YOLOv3 and Faster RCNN. The following comparison is based on Tables 2 and 4.

According to Alganci et al. (2020), the researchers concluded that YOLOv3 has a shorter processing time. In this study, YOLOv3 takes less computational time than Faster RCNN in high-quality and low-quality videos.

Tables 5 and 6 show that the Faster RCNN lacks near miss and accident detection data. In addition, Social Distancing Monitoring and Bird's Eye View techniques are not available at Faster RCNN.

Finally, it is possible to compare YOLOv3 and Faster RCNN, where Faster RCNN requires more computational time than YOLOv3 in the high-quality and low-quality videos. Since YOLOv3 is a simpler architecture, Faster RCNN is trained to perform classification and bounding box regression simultaneously.

In the percentage of error detection comparison, Faster RCNN exhibited more accurate data than YOLOv3 in the videos of Tun Dr Lim Chong Eu Expressway but not in the videos of Lim Chwee Leong Road. Faster RCNN required more dataset samples based on the videos to train the algorithm, whereas YOLOv3 did not require any training. After all, its dataset already trained it. In vehicle detection, both algorithms show detection errors. YOLOv3 detected motorcycles, trucks, and lorries in Tun Dr Lim Chong Eu Expressway videos. The videos of Lim Chwee Leong Road detected motorcycles, tricycles, and buses as vehicles. In Tun Dr Lim Chong Eu Expressway videos, faster RCNN detected banners

as vehicles, while in Lim Chwee Leong Road, it detected roofs, unknown objects, and a bus as vehicles.

CONCLUSION

This research aims to investigate near miss cases in Pulau Pinang to reduce the number of accidents and carbon emissions in the city. This study has the potential to achieve the goals and SDGs outlined in the Penang 2030 mission. It is necessary to identify the black spot area in Tun Dr Lim Chong Eu Expressway and select the peak hours (18/12/2018, 6:30:15p.m. to 18/12/2018, 6:31:34p.m.) with a high-quality video and Lim Chwee Leong Road and the peak hours (5/2/2019, 6:30:15p.m. to 5/2/2019, 6:31:34p.m.) with low-quality video to conduct the vehicle detection by using YOLOv3 (Social Distancing Monitoring and Bird's Eye View method) and Faster RCNN. Faster RCNN takes longer than YOLOv3 computational time due to vehicle detection. In the high-quality video, Faster RCNN produced more accurate data than YOLOv3, while YOLOv3 produced more accurate data in the low-quality video. In YOLOv3, a near miss is likely higher on Tun Dr Lim Chong Eu Expressway than on Lim Chwee Leong Road. Image or video quality needs to be improved in future work. An important criterion is the camera angle. If the calculation can be converted into an autonomous calculation, the duration of the videos can be extended. Near miss is predicted to count automatically. It is expected that the current event will be used to forecast future events such as seasonal changes and the achievement of the Penang 2030 mission.

ACKNOWLEDGEMENT

We would like to thank the MBPP CCTV Team that assisted by providing necessary information and School of Mathematical Sciences, Universiti Science Malaysia for the funding.

REFERENCES

- Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy*, 19 (242), 1-20. <https://doi.org/10.3390/e19060242>
- Aldred, R. (2016). Cycling near misses: Their frequency, impact and prevention. *Transport Research Part A*, 90(1), 69-83. <https://doi.org/10.1016/j.tra.2016.04.016>
- Aldred, R., & Crosweller, S. (2015). Investigating the rates and impacts of near misses and related incidents among UK cyclists. *Journal of Transport & Health*, 2(3), 379-393. <https://doi.org/10.1016/j.jth.2015.05.006>
- Alganci, U., Soydas, M., & Sertel, E. (2020). Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. *Remote sensing*, 12(3), Article No. 458. <https://doi.org/10.3390/rs12030458>

- Aqqa, M., Mantini, P., & Shah, S. K. (2019, February). Understanding how video quality affects object detection algorithms. In *VISIGRAPP (5: VISAPP)* (pp. 96-104). Science and Technology Publications. <https://doi.org/10.5220/0007401600960104>
- Arinaldi, A., Pradana, J., A., & Gurusniaga, A., A. (2018). Detection and classification of vehicles for traffic video analytics. *Procedia Computer Science*, *144*, 259-268. <https://doi.org/10.1016/j.procs.2018.10.527>
- Behl, A., Bhatia, A., & Puri, A. (2014). Convolution and applications of convolution. *International Journal of Innovative Research in Technology (IJIRT)*, *1*(6), 2123-2126.
- Bull, C., B., Hagen, L., A., V., Lubin, A., Shivaraman, G., & Chibbaro, D. (2017). *Predictable is preventable: Tracking pedestrian near-miss incidents*. New Jersey Safe Routes to School Resource Center & Alan M. Voorhees Transportation Center. <https://www.njcrossingguards.org/wp-content/uploads/2017/11/Ped-Near-Miss-report-5.25.17.pdf>
- Calles, M., B., Nelson, T., & Winters, M. (2017). Comparing crowdsourced near-miss and collision cycling data and official bike safety reporting. *Transportation Research Record: Journal of the Transportation Research Board*, *2662*(1), 1-11. <https://doi.org/10.3141/2662-01>
- Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y., & Wu, Z. (2019). An improved faster R-CNN for small object detection. *IEEE Access*, *7*, 106838-106846. <https://doi.org/10.1109/ACCESS.2019.2932731>
- Cepni, S., Atik, M. E., & Druan, Z. (2020). Vehicle detection using different deep learning algorithms from image sequence. *Baltic Journal of Modern Computing*, *8*(2), 347-358. <https://doi.org/10.22364/bjmc.2020.8.2.10>
- Ciberlin, J., Grbic, R., Teslic, N., & Pilipovic, M. (2019). Object detection and object tracking in front of the vehicle using front view camera. In *2019 Zooming Innovation in Consumer Technologies Conference (ZINC)* (pp. 27-32). IEEE Publishing. <https://doi.org/10.1109/ZINC.2019.8769367>
- Ding, X., & Yang, R. (2019). Vehicle and parking space detection based on improved YOLO network model. *Journal of Physics: Conference Series*, *1325*, Article 012084. <https://doi.org/10.1088/1742-6596/1325/1/012084>
- Dixit, K. S., Chadaga, M. G., Savalgimath, S. S., Rakshith, G. R., & Kumar, M. N. (2019). Evaluation and evolution of object detection techniques YOLO and R-CNN. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(2S3), 824-829. <https://doi.org/10.35940/ijrte.B1154070782S319>
- Gad, A. F. (2020). *Faster R-CNN explained for object detection tasks*. PaperspaceBlog. <https://blog.paperspace.com/faster-r-cnn-explained-object-detection/>.
- Giroto, E., Andrade, S., M. D., & Gonzalez, A. D. (2016). Professional experience and traffic accidents/ near-miss accidents among truck drivers. *ELSEVIER: Accidents Analysis and Prevention*, *95*(Pt A), 299-304. <https://doi.org/10.1016/j.aap.2016.07.004>
- Huang, Y. Q., Zheng, J. C., Sun, S. D., Yang, C. F., & Liu, J. (2020). Optimized YOLOv3 algorithm and its application in traffic flow detections. *Applied Sciences*, *10*(9), Article 3079. <https://doi.org/10.3390/app10093079>
- Johnson, K. D., Patel, S. R., Baur, D. M., Edens, E. Sherry, P. Malhotra, A., & Kales, S., N. (2014). Association of sleep habits with accidents and near misses in United States transportation operators.

Journal of Occupational & Environmental Medicine, 56(5), 510-515. <https://doi.org/10.1097/JOM.0000000000000132>

- Kataoka, H., Suzuki, T., Oikawa, S., Matsui, Y., & Satoh, Y. (2018). Drive video analysis for the detection of traffic near-miss incidents. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3421-3428). IEEE Publishing. <https://doi.org/10.1109/ICRA.2018.8460812>.
- Ke, R., Lutin, J., Spears, J., & Wang, Y. (2017). A cost-effective framework for automated vehicle-pedestrian near-miss detection through onboard monocular vision. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 25-32). IEEE Publishing.
- Mahdi, N. N. R. N., Bachok, N., Mohamed, N., & Shafei, M. N. (2014). Risk factors for near miss incident among long distance bus drivers in Malaysia. *Iranian Journal of Public Health*, 43(3), 117-124.
- Makizako, H., Shimada, H., Hotta, R., Doi, T., Tsutsumimoto, K., Nakakubo, S., & Makino, K. (2018). Associations of Near-Miss traffic incidents with attention and executive function among older Japanese drivers. *Gerontology*, 64, 495-502. <https://doi.org/10.1159/000486547>
- Matsui, Y., Hitosugi, M., Doi, T., Oikawa, S., Takahashi, K., & Ando, K. (2013). Features of pedestrian behavior in car-to-pedestrian contact situations in Near-Miss incidents in Japan. *Traffic Injury Prevention*, 14(1), 58-63. <https://doi.org/10.1080/15389588.2013.796372>
- Matsui, Y., Takahashi, K., Imaizumi, R., & Ando, K. (2011). Car-to-pedestrian contact situations in near-miss incidents and real-world accidents in Japan. In *22nd International Technical Conference on the Enhanced Safety of Vehicles* (No. 110164). National Traffic Safety and Environment Laboratory.
- Nadai, S. D., Parodi, F., & Pizzorni, D. (2012). A system of systems to Near Miss accidents in dangerous goods road transportation. In *IEEE 2012 7th International Conference on System of Systems Engineering (SoSE)* (pp. 219-222). IEEE Publishing. <https://doi.org/10.1109/SYSoSE.2012.6384198>.
- Nostikasari, D., & Shelton, K. (2017). *Learning from close calls: A glimpse into near-miss experiences*. Rice University Kinder Institute of Urban Research. <https://doi.org/10.25611/e51k-cpxo>
- Ong, Y. (2020). *Near miss vehicle collisions estimation using YOLO* (Master's dissertation). Universiti Sains Malaysia, Malaysia.
- Poulos, R., G., Hatfield, J., Rissel, C., Flack, L., K., Shaw, L., Grzebieta, R., & McIntosh, A., S. (2017). Near miss experiences of transport and recreational cyclists in New South Wales, Australia. Findings from a prospective cohort study. *Accidents Analysis and Prevention*, 101, 143-153. <https://doi.org/10.1016/j.aap.2017.01.020>
- Qin, Z., Yu, F., Liu, C., & Chen, X. (2018). How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. *Mathematical Foundations Computing*, 1(2), 149-180. <https://doi.org/10.48550/arXiv.1804.11191>
- Rahman, A., Salam, A., Islam, M., & Sarker, P. (2012). An image based approach to compute object distance. *International Journal of Computational Intelligence Systems*, 1(4), 304-312. <https://doi.org/10.1080/18756891.2008.9727627>
- Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9), 2352-2449. https://doi.org/10.1162/neco_a_00990

- Rome, L. D., Brown, J., Baldock, M., & Fitzharris, M. (2018). Near-miss crashes and other predictors of motorcycle crashes: Findings from a population-based survey. *Traffic Injury Prevention, 19*(2), S20-S26. <https://doi.org/10.1080/15389588.2018.1536822>
- Sanders, R., L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accidents Analysis and Prevention, 75*, 26-34. <https://doi.org/10.1016/j.aap.2014.11.004>
- Silva Consultants. (2016). *Using near miss reporting in security*. Silva Consultants. <https://www.silvaconsultants.com/new-security-tips/using-near-miss-reporting-in-security>
- Siregar, M. L., Agah, H. R., & Hidayatullah, F. (2018). Near-miss accident analysis for traffic safety improvement at a “channelized” junction with U-turn. *International Journal of Safety and Security Engineering, 8*(1), 31-38. <https://doi.org/10.2495/SAFE-V8-N1-31-38>
- Sonnleitner, E., Barth, O., Palmanshofer, A., & Kurz, M. (2020). Traffic measurement and congestion detection based on real-time highway video data. *Applied Sciences, 10*(18), Article 6270. <https://doi.org/10.3390/app10186270>
- Srivastava, S., Divekar, A., V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data volume, 8*(66), 1-31.
- Storgard, J., Erdogan, I., Lappalainen, J., & Tapanien, U. (2012). Developing incident and near miss reporting in the maritime industry - A case study on the Baltic Sea. *Procedia Social and Behavioral Sciences, 48*, 1010-1021. <https://doi.org/10.1016/j.sbspro.2012.06.1078>
- Uchida, N., Kawakoshi, M., Tagawa, T., & Mochida, T. (2010). An investigation of factors contributing to major crash types in Japan based on naturalistic driving data. *International Association of Traffic and Safety Sciences (IATSS), 34*(1), 22-30. <https://doi.org/10.1016/j.iatssr.2010.07.002>
- Vinitha, V., & Velantina, V. (2020). Social distancing detection system with artificial intelligence using computer vision and deep learning. *International Research Journal of Engineering and Technology (IRJET), 7*(8), 4049-4053.
- Wang, C., Dai, Y., Zhou, W., & Geng, Y. (2020). A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition. *Hindawi, Journal of Advanced Transportation, 2020*, Article 9194028. <https://doi.org/10.1155/2020/9194028>
- WHO. (2015). *World Health Statistics 2015*. World Health Organization <https://www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/world-health-statistics-2015.pdf>
- Zhang, Z., Trivedi, C., & Liu, X. (2018). Automated detection of grade-crossing-trespassing near misses based on computer vision analysis of surveillance video data. *Safety Science, 110*(Part B), 276-285. <https://doi.org/10.1016/j.ssci.2017.11.023>
- Zohra, A., F., Kamilia, S., & Souad, S. (2018). Detection and classification of vehicles using deep learning. *International Journal of Computer Science Trends and Technology (IJCTST), 6*(3), 23-29.

Appendix A
 Supplementary Table
Previous research comparison of near miss

No	Researchers	Type of paper		Application data			Type of method				Type of accidents				Remarks
		Review	Research	Image process	Empirical data	Survey / Questionnaire	Model	Software	Others	Vehicles	Pedestrians	Motorcycles/ Bicycles			
1	Uchida et al., (2010)	✓	✓	✓		✓	✓		✓			✓			<ul style="list-style-type: none"> • Driving data acquisition system • Japan
2	Nadai et al., (2012)	✓	✓		✓		✓								<ul style="list-style-type: none"> • System of Systems architecture (SoSE) • Biometric data • Can-Bus data • Italy
3	Storgard et al., (2012)	✓	✓		✓		✓								<ul style="list-style-type: none"> • Logistics model • Wald Test • Baltic Sea
4	Matsui et al., (2013)	✓	✓		✓				✓			✓			<ul style="list-style-type: none"> • Time to collision (TTC) • Japan
5	Matsui et al., (2011)	✓	✓		✓		✓		✓			✓			<ul style="list-style-type: none"> • Pedestrian time-to-vehicle • vehicle time-to-collision • Japan
6	Johnson et al., (2014)	✓	✓		✓		✓		✓						<ul style="list-style-type: none"> • Regression analyses • USA
7	Mahdi et al., (2014)	✓	✓		✓		✓								<ul style="list-style-type: none"> • Multiple logistic regression • Malaysia
8	Aldred & Crossweller, (2015)	✓	✓		✓			✓				✓			<ul style="list-style-type: none"> • KeySurvey software • UK
9	Sanders, (2015)	✓	✓		✓							✓			<ul style="list-style-type: none"> • Internet survey • USA

Supplementary Table (continue)

No	Researchers	Type of paper		Application data		Type of method					Type of accidents				• Remarks
		Review	Research	Image process	Empirical data	Survey / Questionnaire	Model	Software	Others	Vehicles	Pedestrians	Motorcycles/ Bicycles			
10	Aldred, (2016)	✓			✓	✓									• Semi-structured interview • UK
11	Giroto et al., (2016)		✓		✓	✓			✓						• Self-applied questionnaire • Brazil
12	Bull et al., (2017)		✓		✓		✓		✓			✓			• Swedish Traffic Conflict Technique (TCT) • USA • TTC • USA
13	Calles et al., (2017)		✓		✓		✓		✓			✓			• Histograms of Oriented Gradients (HOG) pedestrian detector • TTC • USA
14	Ke et al., (2017)		✓		✓		✓								
15	Nostikasari & Shelton, (2017)		✓		✓		✓					✓			• Questionnaire form • USA
16	Poulos et al., (2017)		✓		✓		✓					✓			• Questionnaire form • Australia
17	Kataoka et al., (2018)		✓		✓				✓			✓			• Near-Miss Incident Database (NIDB) • Japan
18	Makizako et al., (2018)		✓		✓		✓				✓				• Interview method • Japan
19	Rome et al., (2018)		✓		✓		✓					✓			• Survey logistic regression • Australia

Supplementary Table (continue)

No	Researchers	Type of paper		Application data		Type of method			Type of accidents				• Remarks
		Review	Research	Image process	Empirical data	Survey / Questionnaire	Model	Software	Others	Vehicles	Pedestrians	Motorcycles/ Bicycles	
20	Siregar et al., (2018)	✓	✓	✓	✓			✓		✓			<ul style="list-style-type: none"> • Fervor application • Indonesia
21	Zhang et al., (2018)	✓	✓	✓			✓			✓			<ul style="list-style-type: none"> • Region of interest (ROI) • Grade-crossing-trespassing near miss
22	Wang et al., (2020)	✓	✓	✓			✓	✓		✓	✓	✓	<ul style="list-style-type: none"> • USA • Retinex Image Enhancement Algorithms • YOLOv3 • China



Effect of Process Conditions on Catalytic Hydrothermal Oxidation of *p*-Xylene to Terephthalic Acid

Mohamad Zarqani Yeop¹, Kamariah Noor Ismail² and Ahmad Rafizan Mohamad Daud^{2*}

¹*School of Chemical Engineering, Universiti Teknologi MARA, 81750 UiTM, Masai, Johor, Malaysia*

²*School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

ABSTRACT

This study investigates the influence of hydrothermal process conditions on the yield of terephthalic acid (TPA). Deionised water was employed as a green reaction medium substitute for acetic acid solvent widely used in the Amoco oxidation process for TPA production. Utilising the unique properties of water at elevated temperature and pressure, TPA was synthesised from *p*-xylene under subcritical (250 °C, 300 °C and 350 °C) and supercritical (400 °C) water conditions in a 10 mL micro-bomb batch reactor. Process conditions, including hydrogen peroxide (H₂O₂) oxidant concentrations, manganese bromide (MnBr₂) catalyst and water loadings, were varied at a fixed reaction time of 60 minutes. The *p*-xylene conversion and TPA yield were determined using high-performance liquid chromatography (HPLC). In addition, the presence of chemical functional groups and chemical compositions of the reaction products were examined using Fourier transform infrared spectroscopy (FTIR) and gas chromatography-mass spectrometer (GC-MS), respectively. It was found that an optimum TPA yield of 94.56% was observed at 350°C with hydrogen peroxide, deionised water and manganese bromide catalyst set at 1.5 mL, 2.5 mL, and 2 mL, respectively. Other major reaction products identified were *p*-tolualdehyde and 1,4-hydroxymethyl benzaldehyde.

Keywords: Hydrothermal, oxidation, sub-and supercritical water, terephthalic acid

ARTICLE INFO

Article history:

Received: 04 October 2021

Accepted: 19 January 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.16>

E-mail addresses:

zarqani@uitm.edu.my (Mohamad Zarqani Yeop)

knoor@uitm.edu.my (Kamariah Noor Ismail)

ahmad2057@uitm.edu.my (Ahmad Rafizan Mohamad Daud)

* Corresponding author

INTRODUCTION

Terephthalic acid (TPA) is an important petrochemical intermediate used in manufacturing polyester (Lee et al., 2021). Almost 70% of the global terephthalate

feedstocks are industrially synthesised from *p*-xylene via an Amoco oxidation process (Tomás et al., 2013; Walt, 2020). In this process, *p*-xylene is oxidised with air at around 175-200 °C and 15-30 bar using soluble cobalt (Co)/manganese (Mn) catalyst, a bromide (Br) promoter and acetic acid (AcOH) solvent (Tomás, 2013). Although the process is highly efficient, applications of AcOH as a solvent to facilitate the reaction has led to health, safety, and environmental concerns. The AcOH reacted with the bromide catalysts to produce hazardous methyl bromide (CH₃Br) gas. Moreover, the formation of undesired bromide associated liquid products also reduces the concentration of bromine catalysts in the reaction system and causes serious corrosion to the process equipment.

Over the years, attempts have been made to develop a safer and greener synthesis route for TPA, including the elimination of purification steps via a new spray concept (Li et al., 2013), development and use of new types of homogeneous catalysts involving bromide salts of manganese, cobalt and copper (Holiday et al., 1998; Pérez et al., 2011), *N*-hydroxyimide (Falcon et al., 2010), heterogeneous cobalt-based catalysts (Xu et al., 2020; Vakros, 2021) and application of high-temperature liquid water to replace AcOH solvent (Dunn & Savage, 2002; Fraga-Dubreuil & Poliakoff, 2006).

The high-temperature, high-pressure water reaction conditions (>100 °C, >1 atm) known as hydrothermal have received increasing attention because of the ability to fine-tune the water physicochemical properties by pressure and temperature adjustments (Byrappa & Yoshimura, 2012; Carr et al., 2011; Cocero, 2018; Savage, 2009;). Around the critical conditions (T_c : 374 °C, P_c : 22.1 MPa, ρ_c : 320 kg.m⁻³), water characteristics such as diffusivity, viscosity, dielectric constant, and solvation properties change dramatically with small changes in pressure due to large compressibility of critical fluids. Consequently, favourable reaction kinetics and mechanisms (Brunner, 2014; Kwak et al., 2009), as well as an increase in the dissolution of oxygen and functionalised aromatic compounds (Kruse & Dinjus, 2007a) can be obtained to support various types of organic and aromatic chemical reactions (Eckert & Chandler, 1998; Daud et al., 2021; Kruse & Dinjus, 2007b) through free radical and ionic reactions (Kruse & Dinjus, 2007b; Jiang et al., 2020). From a chemical process perspective, hydrothermal is regarded as safe, non-toxic, environmentally benign, and easy to handle (Cocero, 2018; Dunn & Savage, 2005).

Holiday et al. (1998) demonstrated that selective *p*-xylene conversion to the TPA in a good yield of around 60% could be achieved in subcritical water. Thus, following this work, catalytic hydrothermal oxidation (Dunn & Savage, 2002 & 2005; Dunn et al., 2003; Fraga-Dubreuil & Poliakoff, 2006; Osada & Savage, 2009a; Osada & Savage, 2009b; Pérez et al., 2011) and non-catalytic hydrothermal oxidation (Kim et al., 2002) of *p*-xylene under hydrothermal conditions have been investigated. Generally, the *p*-xylene hydrothermal oxidation pathway mimics the AcOH solvent-based Amoco process pathway. From the outset, efforts have been directed at bypassing the rate-limiting step of *p*-toluic acid oxidation to 4-carboxybenzaldehyde (4-CBA). These studies suggested that TPA

synthesis is sensitive to changes in hydrothermal oxidation conditions, particularly reaction temperature, oxidant concentration and types of catalysts used. A summary of previous studies on hydrothermal oxidation of *p*-xylene is presented in Table 1.

This research examines the effects of hydrothermal reaction conditions involving reaction temperature (sub- and supercritical water, 250-400 °C), oxidant concentrations, manganese bromide catalyst loadings and water loadings during *p*-xylene conversion and yield to the TPA. The reaction temperature is expected to influence the *p*-xylene conversion reaction significantly. Changes in reaction temperature directly affect the dielectric constant of water responsible for the dissolution of organics in high-temperature water. Moreover, a higher reaction temperature accelerates the degradation of hydrogen peroxide to form hydroxyl free radicals and increases the overall oxidation reaction rate. Therefore, investigations were conducted in conditions encompassing the sub- and supercritical water range.

Table 1
Previous studies on TPA synthesis from p-xylene under subcritical and supercritical water conditions

Conditions	Oxidant	Catalyst	Findings	Ref.
Subcritical water, 333 °C	O ₂	MnBr ₂	Yield, 64%	Holiday et al., (1998)
Subcritical & Supercritical water, 240-400 °C & 220-300 bar	H ₂ O ₂	No Catalyst	Conversion: SubCW ~ 89.02% SCW ~ over 99%	Kim et al., (2002)
Subcritical water, 300 °C	H ₂ O ₂	MnBr ₂	Yield, 49±8%	Dunn and Savage (2002)
Subcritical water, 300°C	O ₂	MnBr ₂	Yield >80 %	Dunn and Savage (2005)
Subcritical water, 300 °C	O ₂	MnBr ₂	Yield > 70 % Selectivity, 90%	Osada & Savage (2009b)
Subcritical water, 300 °C	O ₂	MnBr ₂	Yield > 80 % Selectivity > 90%	Osada & Savage (2009a)
Subcritical water, 330 °C		Cu/Co/NH ₄ /Br	Yield, 70.5%	Perez et. al. (2011)

METHODOLOGY

Materials

p-xylene 8.1M (reagent grade, 99%), TPA powder (>98%) hydrogen peroxide solution 9.7M (H₂O₂, 30% wt./wt.) and manganese (II) bromide catalyst (MnBr₂) 0.014M were purchased from Merck Sdn. Bhd. Organic solvents dimethyl sulfoxide (DMSO) and dichloromethane (DCM) were obtained from Chemolab Supplies Sdn Bhd. Both solvents are reagent grade (>99.9%). All hydrothermal oxidation experiments were conducted using deionised water as a reaction medium. The experiments were accomplished using a 10mL micro-bomb reactor fashioned from Swagelok's high pressure-high temperature tubing and fittings. An electrical tube furnace was used to provide heat to the reactor.

Experimental

P-xylene catalytic hydrothermal oxidation experiments were conducted in a 10mL micro-bomb batch reactor. The reactor was assembled using a ½ inch Swagelok stainless steel tubing and end caps (Figure 1). At first, reactants *p*-xylene and H₂O₂ and MnBr₂ catalyst were carefully charged into the micro-bomb reactor. Next, deionised water was added according to the amount specified in Table 2. Then, the reactor end cap was properly tightened using a wrench to ensure leak-free experiments. Heating is provided by a programmable tube furnace whereby heating rate and temperature are controlled using a proportional integral derivative (PID) controller. Prior to the introduction of the reactor, the furnace temperature was set to the required reaction temperature at 250 °C, 300 °C, 350 °C and 400 °C. The reaction temperature was controlled within ±5°C, and reaction time was taken immediately after the reactor was placed and secured. For all experiments, the reaction time was set at 60 min. At the end of the reaction, the micro-bomb reactor was quickly withdrawn from the furnace and quenched in cold water to room temperature. Subsequently, the micro-bomb reactor was carefully opened to avoid sudden release of residual pressure and losses of liquid sample. The reactor was not equipped with a gas collection port, and, therefore, gas products were not recovered.

The liquid sample was transferred to a sample vial, and the reactor was rinsed with 5.0 mL of DMSO solvent to remove any remaining products. Gas chromatography-mass spectrometer (GC-MS), Fourier transforms infrared spectroscopy (FTIR), and high-performance liquid chromatography (HPLC) was used to analyse the liquid products. An experimental error was estimated from four experiments conducted under identical conditions (H₂O₂ 1.5mL, MnBr₂ 2mL, deionised water 1.5mL, 300 °C and 60 min). It was found that the standard deviation for this experiment was 5 for the TPA yield and the standard experimental error percentage was ±4%.

Table 2
Range of experimental reaction parameters studied

Reaction parameters	Range	Units
Temperature	250–400	°C
<i>p</i> -xylene	0.1	mL
Water	1.5–2.5	mL
MnBr ₂	1.5–2.5	mL
H ₂ O ₂	0.5–2.0	mL

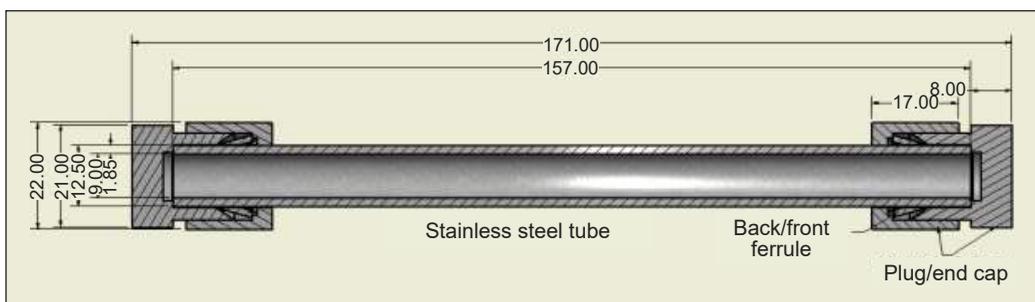


Figure 1. Construction of the hydrothermal micro-bomb reactor

Analysis

The liquid sample was analysed using a Perkin Elmer Spectrum One FTIR to identify the functional groups present in the mixture. All spectra were generated using 4 cm^{-1} and 4 scans spectral resolution at a wavelength range of 4000 to 515 cm^{-1} .

The desired TPA product was analysed by a Perkin Elmer reverse phase-high performance liquid chromatography (RP-HPLC). A mixture of acetonitrile (5%) and a diluted buffer solution of sodium acetate/acetic acid (95%) was used as the mobile phase, and the flow was adjusted to 0.7 mL/min . Standards and liquid product samples were manually injected at $6\text{ }\mu\text{L}$, and separation was accomplished within 15 min run time using a LichroCART column maintained at ambient temperature. The peak of the TPA was detected at 254 nm wavelength using a UV detector. The TPA product concentration in the sample was determined from a calibration curve plotted using the peak area for the TPA standard solutions prepared at five different concentrations (Figure 2).

The liquid by-products obtained from the experiments were identified using a Varian 450 gas chromatography/Varian 240 mass spectrometry (GC/MS). Prior to injection, the liquid sample was diluted with DCM solvent. The GC-MS was fitted with a BP5MS column (30m, 0.25 mm diameter and $0.25\text{ }\mu\text{m}$ film thickness). The initial column temperature was set at $60\text{ }^\circ\text{C}$ (held for 0.5 min). The temperature was then ramped to a final temperature of $180\text{ }^\circ\text{C}$ at $5\text{ }^\circ\text{C min}^{-1}$ and held for 0.5 min . The temperature of the injector was kept at $200\text{ }^\circ\text{C}$. Chromatogram acquisition was made in full scan mode between 50 – 500 m/z . Helium was used as carrier gas at 1 mL/min flowrate. The reaction by-products were identified by matching their mass spectra to the NIST mass spectrum library.

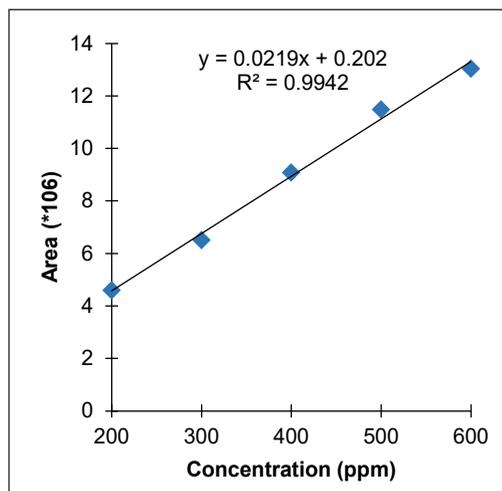


Figure 2. HPLC calibration curve for TPA product quantification prepared using TPA standard solutions

RESULTS AND DISCUSSION

Effect of Reaction Temperature

The effect of reaction temperature was investigated at a fixed catalyst loading of 2.0 mL , deionised water of 2.5 mL and hydrogen peroxide of 1.5 mL . The temperature varied between subcritical ($250\text{ }^\circ\text{C}$ – $350\text{ }^\circ\text{C}$) and supercritical ($400\text{ }^\circ\text{C}$). As shown in Figure 3, the TPA yield increases in the subcritical water region, with a maximum TPA yield of 94.6% obtained at $350\text{ }^\circ\text{C}$. As the reaction temperature approaches near-critical water

condition (T_c 374 °C), the increase in *p*-xylene conversion to TPA was likely assisted by the availability of reactive oxygen species derived from hydrogen peroxide degradation. It was widely accepted that higher reactivity and yield observed in hydrogen peroxide assisted oxidation reaction system was closely related to the presence of OH and HO₂ free radicals (Croiset et al., 1997; Jiang et al., 2020). These short-lived radicals were responsible for activating the cracking of C-C and C=C bonds through a series of oxidation reactions leading to the TPA formation. In addition, higher reaction temperature also decreases the water dielectric constant (Chaudhary et al., 2021), which leads to a higher dissolution of non-polar compounds such as *p*-xylene in a high-temperature water medium.

Although the formation of the TPA was favoured in the subcritical water region, the yield was decreased under supercritical water temperature of 400 °C. Dunn & Savage (2002 & 2003) reported that the TPA was unstable hydrothermally at above 300 °C and was likely degraded to other by-products such as benzoic acid through decarboxylation reaction. Similar to the trend observed in this work, Kim et al. (2002) observed a maximum TPA yield at a subcritical water condition of 300 °C during *p*-xylene oxidation investigation in sub- and supercritical water conditions.

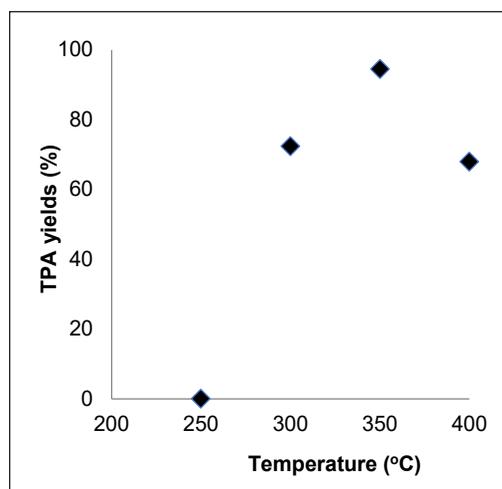


Figure 3. Effect of temperature on TPA yield. Reaction conditions: MnBr₂ catalyst loading 2.0 mL, water loading 2.5 mL, hydrogen peroxide loading 1.5 mL and 60 min reaction time.

Effect of Catalyst Loading

Figure 4 shows two distinctive trends for the TPA yields obtained at different MnBr₂ catalyst loadings and reaction temperatures. Firstly, the TPA yield obtained was maximum in sub- and supercritical water conditions using 2.0 mL MnBr₂ catalyst loading. Secondly, the trend showed an increase in the TPA yield with increasing temperature under the subcritical water region, reaching a maximum of 350 °C. It is followed by a decrease in yields at the supercritical water temperature of 400 °C for all catalyst loadings investigated, mimicking the trend observed in Figure 3. An optimum condition was observed at 2.0 mL MnBr₂ catalyst loading and 350 °C reaction temperature.

During hydrothermal oxidation, *p*-xylene was transformed through a series of oxidation reactions starting with the production of *p*-tolualdehyde. It is then further oxidised to *p*-toluic acid and 4-carboxybenzaldehyde before being converted to the TPA. This pathway is limited by forming an oxidation-resistant *p*-toluic acid intermediate, which can be

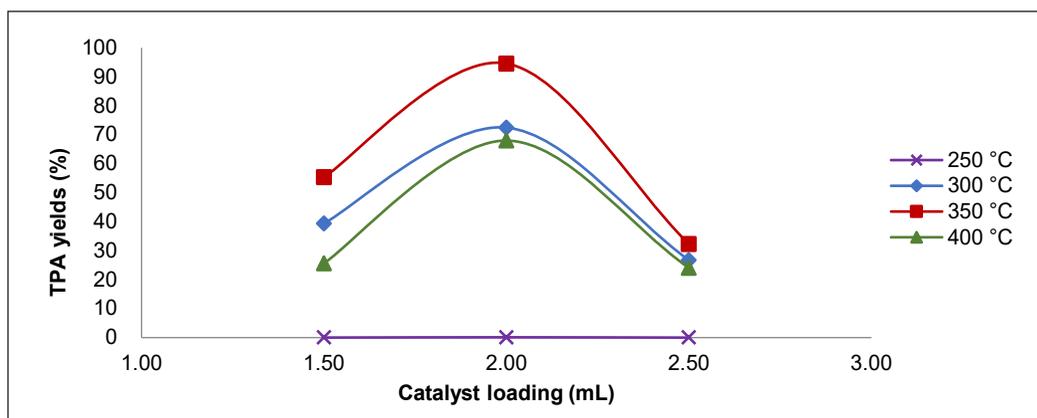


Figure 4. Effect of MnBr_2 catalyst loadings on TPA yields at different reaction temperatures. Water and hydrogen peroxide loadings were fixed at 2.5 mL and 1.5 mL, respectively. Reaction time 60 min.

overcome using catalysts such as MnBr_2 . In the presence of a bromine initiator, the oxidation rate was enhanced, increasing the TPA yield (Holiday et al., 1998; Tomás et al., 2013). The trend observed in Figure 4 is similar to the works of Li and Li (2008), which found that an optimum condition for the TPA formation existed during catalytic hydrothermal oxidation of the *p*-xylene. The authors found that increasing the catalyst loading beyond the optimum condition will negatively affect the TPA yield. Dibromide radicals were formed in excess at high catalyst concentration during fast chain reactions leading to undesired benzylic bromides. This reaction reduces the concentration of the catalytic sites needed for the oxidation reactions.

Effect of an Oxidant Loading

The effect of hydrogen peroxide oxidant loadings was assessed at different reaction temperatures, as illustrated in Figure 5. The TPA yield obtained at the lower end of the reaction temperature investigated (250 °C) was below 1%, indicating *p*-xylene resistance to hydrothermal oxidation at low temperature. It is consistent with the previous results presented in Figures 3 and 4. By contrast, the TPA yields increase at a higher hydrothermal temperature of 350 °C and oxidant loading of 1.5 mL. The increase in *p*-xylene oxidation reactivity coincides with the degradation of hydrogen peroxide occurring at between 280–300 °C, which releases the OH and HO_2 radicals. It is thought that these radicals had a role in activating and increasing the reactivity of the reaction system. After the initial free radical activated reactions stage, the higher oxygen concentration was likely available in the hydrothermal system, allowing *p*-xylene and its subsequent intermediates to be oxidised to the TPA.

Similar to the trends observed in Figure 4, Figure 5 suggests an optimum condition for the TPA yields were obtained at 350 °C and an oxidant loading of 1.5 mL. A significant

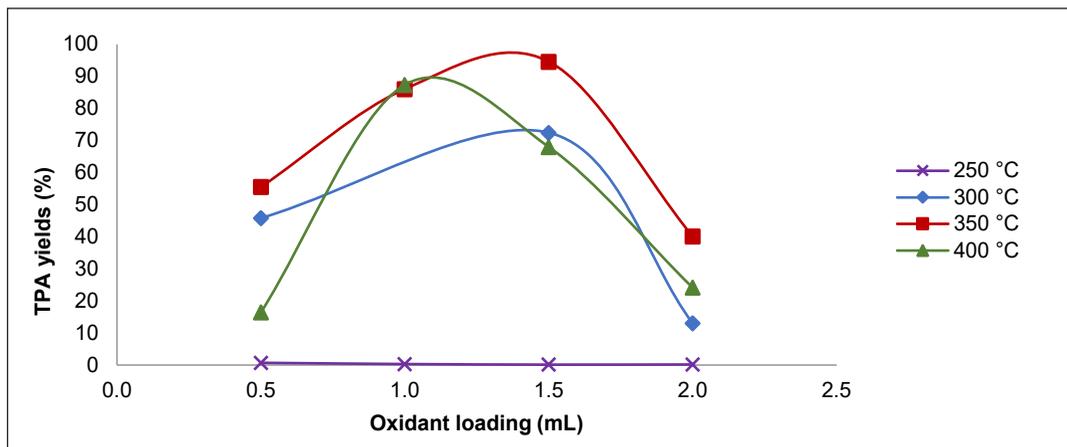


Figure 5. Effect of oxidant loadings on TPA yields at different reaction temperatures. Water and MnBr_2 catalyst loadings were fixed at 2.5 mL and 2.0 mL, respectively. Reaction time 60 min.

drop in the TPA yield can be observed at higher oxygen loading above the optimum condition. The trends show that the oxidative conversion of *p*-xylene requires a moderate reaction temperature of 300-350 °C due to the strong oxidant agent and the exothermic nature of the reaction. When a high reaction temperature of 400 °C was used together with a high oxidant concentration of above 1 mL, the combination may have induced multiple cleavages of *p*-xylene structure through radical intermediates leading to the transformation of *p*-xylene into unwanted liquid products and gases. In this regard, the *p*-xylene was either converted to other intermediates than the target TPA through oligomerisation reaction (Osada & Savage, 2009b) or experienced burning and coking (Dunn & Savage, 2005) when oxygen was present in excess at a high enough temperature.

Effect of Water Loading

Water loadings' effect was evaluated and conducted at the optimum temperature of 350 °C as previously determined. The water loading as a solvent was varied at 1.5 mL and 2.5 mL. As shown in Figure 6, a satisfactory TPA yield of around 94% was obtained using 2.5 mL water loading. The TPA yield was slightly reduced to 88% at lower water loading. The change in the TPA yields observed could be due to the water density effect. At higher water density, the

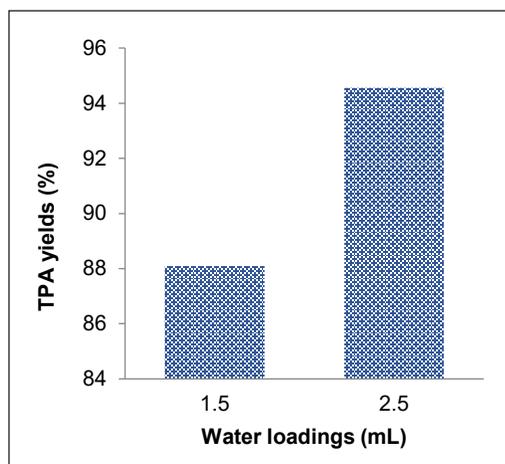


Figure 6. Effect of water loadings on TPA yields. Reaction conditions fixed at 1.5 mL hydrogen peroxide, 2.0 mL MnBr_2 catalyst, 350 °C and 60 min reaction time.

water solvent surrounds the reactant and thus facilitates the dispersion of the reactant for catalytic oxidation reaction to occur. The results indicate the viability of the hydrothermal medium as a replacement for AcOH solvent to produce the TPA.

p-Xylene Hydrothermal Oxidation Products Characterization

Figure 7 shows the FTIR spectra accumulated for products of *p*-xylene hydrothermal oxidation obtained at different temperatures. All spectra showed similar patterns with several identifiable peaks that characterised the *p*-xylene oxidation products at absorption bands between 1700 cm^{-1} and 1600 cm^{-1} . The appearance of the absorption band signal at 1738 cm^{-1} corresponds to the C=O stretching for carboxylic acid, while absorption bands 1740 cm^{-1} –1690 cm^{-1} match those of C=O stretching in ketones and aldehydes. The peak at 2970 cm^{-1} (between 3000 cm^{-1} –2500 cm^{-1}) corresponds to O-H stretching from carboxylic acid constituents. Other peaks ascribing to the vibrations of C-H bending (originating from benzene ring), C=C aromatic bending, and aromatic fingerprint region were detected at 3016 cm^{-1} , 1654.31 cm^{-1} and 707 cm^{-1} wavelengths, respectively. As seen in all spectra obtained, the reaction products were characterised by chemical functionalities associated with the TPA and other by-products. The spectra for the 250 °C sample showed lower peak intensities than higher temperature samples indicating a lower concentration of the respective functionalities. The low TPA confirms it yields at 250 °C reaction temperature.

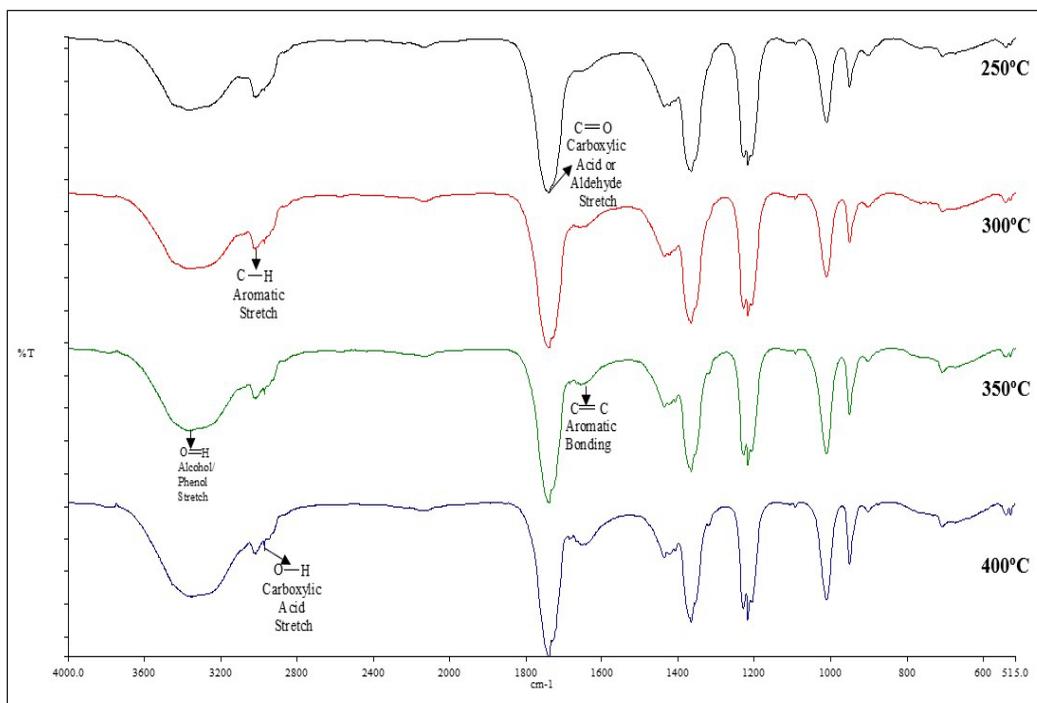


Figure 7. FTIR spectra of *p*-xylene hydrothermal oxidation products obtained between 250–400 °C.

Figures 8 to 10 show the GC-MS chromatograms for *p*-xylene hydrothermal oxidation reaction liquid products at 250-350 °C. The chromatograms tracked the evolution of reaction products (other than TPA, which HPLC analysed) as the reaction condition changed from subcritical to supercritical water temperatures. At 250 °C, *p*-tolualdehyde was the sole reaction product identified alongside the unreacted *p*-xylene indicating minimum *p*-xylene conversion had occurred at the lower end of the subcritical water temperature investigated. As the reaction temperature was raised to near-critical water conditions of 300 and 350 °C (Figures 9 and 10), a higher degree of *p*-xylene oxidation and degradation were observed. At these temperatures, the product mixture consisted of high proportions of oxygenated compounds such as *p*-tolualdehyde and 1,4-benzene dicarboxaldehyde, having a higher mass to charge ratio (*m/z*) than the starting *p*-xylene. Other constituents detected at lower intensities were benzene, 1,4-methylbenzyl methanol, and 1,4-hydroxymethyl benzaldehyde. The occurrence of compounds with carboxylic, hydroxyl, carbonyl and aldehyde chemical characteristics complimented the FTIR results. Their presence in the product mix suggested that the pathway for TPA synthesis using high-temperature water as a reaction medium resembles the pathway of the conventional AcOH solvent process (Tomas et al., 2013). The conventional pathway involves consecutive partial oxidation of methyl groups to form *p*-tolualdehyde, *p*-toluic acid and 4-carboxybenzaldehyde compounds as major intermediates.

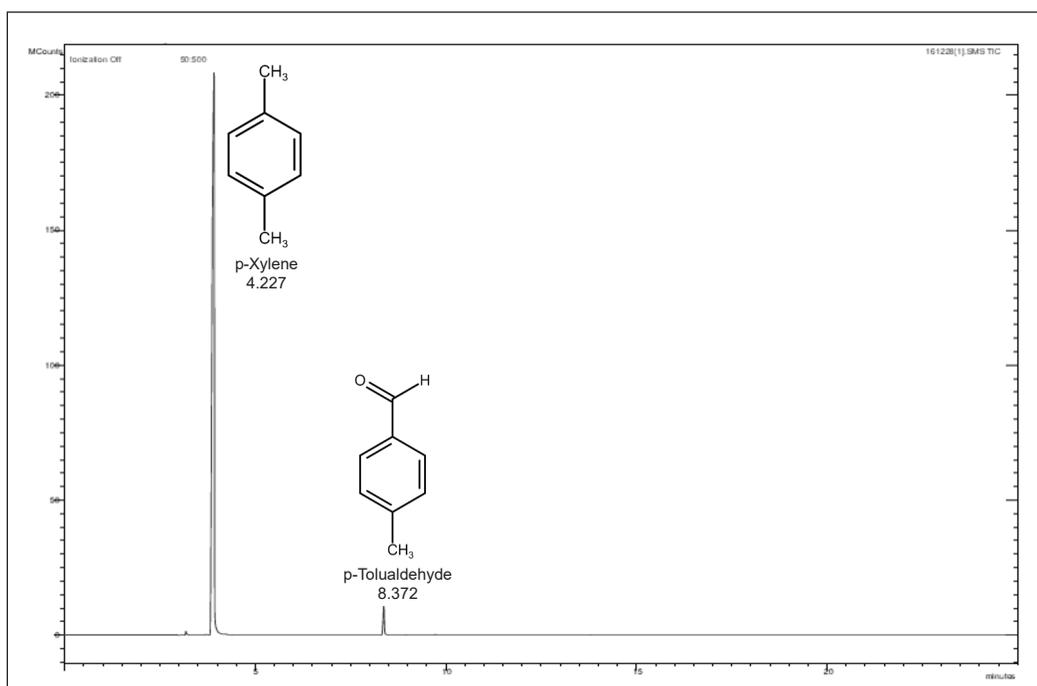


Figure 8. Products of *p*-xylene hydrothermal oxidation identified at 250 °C. Reaction conditions: MnBr₂ catalyst loading 2.0 mL, water loading 2.5 mL, hydrogen peroxide loading 1.5 mL and 60 min reaction time.

Catalytic Hydrothermal Oxidation of *p*-Xylene to TPA

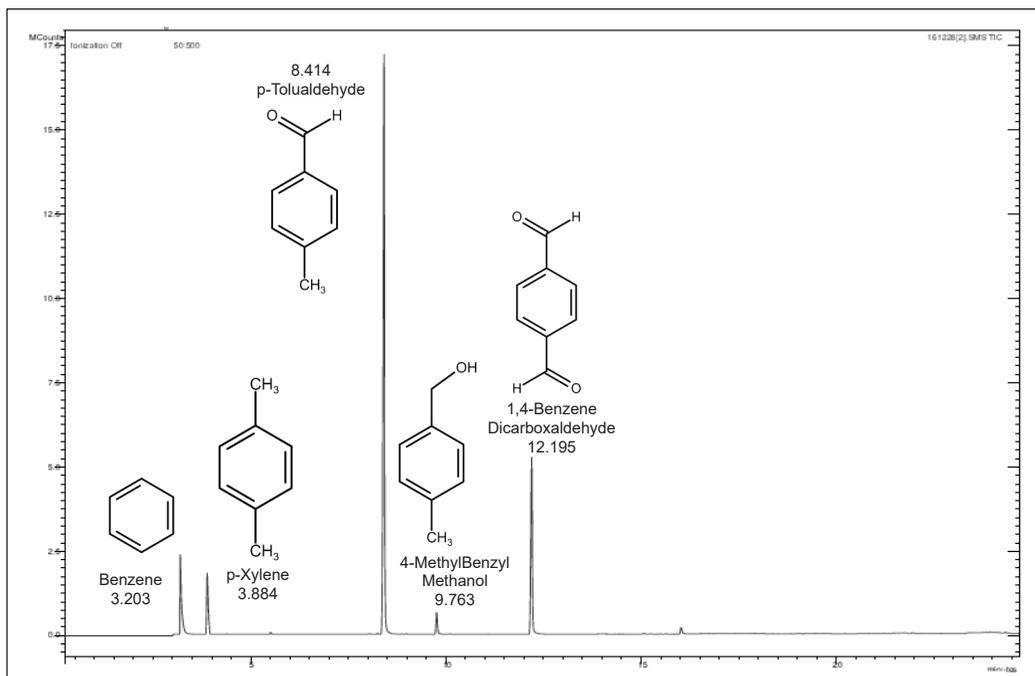


Figure 9. Products of *p*-xylene hydrothermal oxidation identified at 300 °C. Reaction conditions: MnBr₂ catalyst loading 2.0 mL, water loading 2.5 mL, hydrogen peroxide loading 1.5 mL and 60 min reaction time.

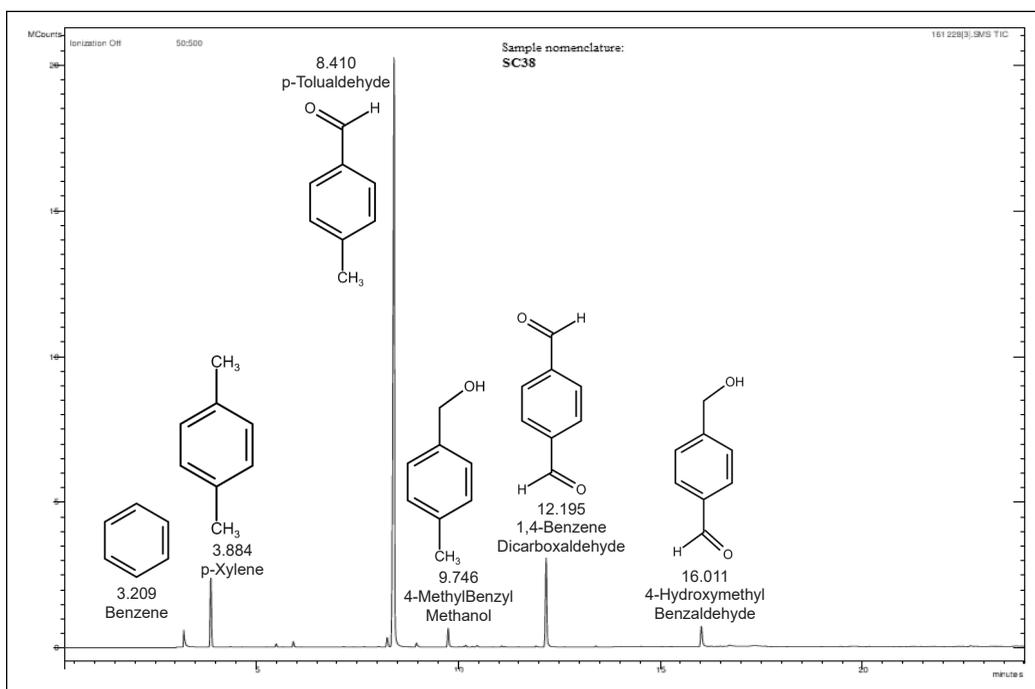


Figure 10. Products of *p*-xylene hydrothermal oxidation identified at 350 °C. Reaction conditions: MnBr₂ catalyst loading 2.0 mL, water loading 2.5 mL, hydrogen peroxide loading 1.5 mL and 60 min reaction time.

CONCLUSION

This study successfully evaluated the influence of hydrothermal process conditions (reaction temperature, catalyst loading, oxidant loading and water loading) on *p*-xylene conversion to TPA. The catalytic *p*-xylene conversion was promoted under subcritical water conditions with an optimum TPA yield of 94.6% obtained at 350 °C. Apparently, under catalytic subcritical water conditions, sufficient reactive oxygen species were available to initiate and support the *p*-xylene oxidation reactions leading to satisfactory yields of the TPA obtained. The results suggested the potential of the hydrothermal method as a promising alternative route for *p*-xylene oxidation to TPA as the main product. Replacing acetic acid with water as a solvent and reaction medium for the TPA synthesis has tremendous potential. Not only water is a green solvent, but hydrothermal could also present an economic TPA synthesis technology by eliminating the need for expensive acetic acid-water separation steps commonly formed during the reaction. Although the effectiveness of water as a medium for TPA synthesis was demonstrated in this work, further assessments on the diffusion of reactive species in high-temperature water and solvent cage effects which may cause the catalytic activity to drop, must be explored.

ACKNOWLEDGMENT

The authors thanked the School of Chemical Engineering, Universiti Teknologi MARA, and the Ministry of Higher Education Malaysia for FRGS research grants (600-RMI/FRGS 5/3 (25/2013)).

REFERENCES

- Byrappa, K., & Yoshimura, M. (2012). *Handbook of Hydrothermal Technology* (2nd Ed.). William Andrew.
- Brunner, G. (2014). Reactions in hydrothermal and supercritical water. *Supercritical Fluid Science and Technology*, 5, 265-322. <https://doi.org/10.1016/B978-0-444-59413-6.00005-4>
- Carr, A. G., Mammucari, R., & Foster, N. R. (2011). A review of subcritical water as a solvent and its utilisation for the processing of hydrophobic organic compounds. *Chemical Engineering Journal*, 172(1), 1-17. <https://doi.org/10.1016/j.cej.2011.06.007>
- Chaudhary, A., Dwivedi, A., & Upadhyayula, S. (2021). Supercritical fluids as green solvents. In *Handbook of Greener Synthesis of Nanomaterials and Compounds* (pp. 891-916). Elsevier. <https://doi.org/10.1016/B978-0-12-821938-6.00028-1>
- Cocero, M. J. (2018). Supercritical water processes: Future prospects. *The Journal of Supercritical Fluids*, 134, 124-132. <https://doi.org/10.1016/j.supflu.2017.11.018>
- Croiset, E., Rice, S. F., & Hanush, R. G. (1997). Hydrogen peroxide decomposition in supercritical water. *AIChE Journal*, 43(9), 2343-2352. <https://doi.org/10.1002/aic.690430919>

- Daud, A. R. M., Berruoco, C., Hellgardt, K., Millan, M., & Kandiyoti, R. (2021). Oxidative cracking of three to five-member ring polycyclic aromatic hydrocarbons in subcritical and supercritical water. *The Journal of Supercritical Fluids*, *167*, Article 105050. <https://doi.org/10.1016/j.supflu.2020.105050>
- Dunn, J. B., & Savage, P. E. (2002). Terephthalic acid synthesis in high-temperature liquid water. *Industrial & Engineering Chemistry Research*, *41*(18), 4460-4465. <https://doi.org/10.1021/ie0107789>
- Dunn, J. B., Burns, M. L., Hunter, S. E., & Savage, P. E. (2003). Hydrothermal stability of aromatic carboxylic acids. *The Journal of Supercritical Fluids*, *27*(3), 263-274. [https://doi.org/10.1016/S0896-8446\(02\)00241-3](https://doi.org/10.1016/S0896-8446(02)00241-3)
- Dunn, J. B., & Savage, P. E. (2005). High-temperature liquid water: A viable medium for terephthalic acid synthesis. *Environmental Science & Technology*, *39*(14), 5427-5435. <https://doi.org/10.1021/es048575+>
- Eckert, C. A., & Chandler, K. (1998). Tuning fluid solvents for chemical reactions. *The Journal of Supercritical Fluids*, *13*(1-3), 187-195. [https://doi.org/10.1016/S0896-8446\(98\)00051-5](https://doi.org/10.1016/S0896-8446(98)00051-5)
- Falcon, H., Campos-Martin, J. M., Al-Zahrani, S. M., & Fierro, J. L. G. (2010). Liquid-phase oxidation of p-xylene using N-hydroxyimides. *Catalysis Communications*, *12*(1), 5-8. <https://doi.org/10.1016/j.catcom.2010.07.010>
- Fraga-Dubreuil, J., & Poliakoff, M. (2006). Organic reactions in high-temperature and supercritical water. *Pure and Applied Chemistry*, *78*(11), 1971-1982. <https://doi.org/10.1351/pac200678111971>
- Holliday, R. L., Jong, B. Y., & Kolis, J. W. (1998). Organic synthesis in subcritical water: Oxidation of alkyl aromatics. *The Journal of Supercritical Fluids*, *12*(3), 255-260. [https://doi.org/10.1016/S0896-8446\(98\)00084-9](https://doi.org/10.1016/S0896-8446(98)00084-9)
- Jiang, Z., Li, Y., Wang, S., Cui, C., Yang, C., & Li, J. (2020). Review on mechanisms and kinetics for supercritical water oxidation processes. *Applied Sciences*, *10*(14), Article 4937. <https://doi.org/10.3390/app10144937>
- Kim, Y. L., Kim, J. D., Lim, J. S., Lee, Y. W., & Yi, S. C. (2002). Reaction pathway and kinetics for uncatalyzed partial oxidation of p-xylene in sub- and supercritical water. *Industrial & Engineering Chemistry Research*, *41*(23), 5576-5583. <https://doi.org/10.1021/ie010952t>
- Kruse, A., & Dinjus, E. (2007a). Hot compressed water as reaction medium and reactant: Properties and synthesis reactions. *The Journal of Supercritical Fluids*, *39*(3), 362-380. <https://doi.org/10.1016/j.supflu.2006.03.016>
- Kruse, A., & Dinjus, E. (2007b). Hot compressed water as reaction medium and reactant: 2. Degradation reactions. *The Journal of Supercritical Fluids*, *41*(3), 361-379. <https://doi.org/10.1016/j.supflu.2006.12.006>
- Kwak, J. W., Lee, J. S., & Lee, K. H. (2009). Co-oxidation of p-xylene and p-toluic acid to terephthalic acid in water solvent: Kinetics and additive effects. *Applied Catalysis A: General*, *358*(1), 54-58. <https://doi.org/10.1016/j.apcata.2009.01.037>
- Lee, H. L., Chiu, C. W., & Lee, T. (2021). Engineering terephthalic acid product from recycling of PET bottles waste for downstream operations. *Chemical Engineering Journal Advances*, *5*, Article 100079. <https://doi.org/10.1016/j.ceja.2020.100079>

- Li, K. T., & Li, S. W. (2008). CoBr₂-MnBr₂ containing catalysts for catalytic oxidation of p-xylene to terephthalic acid. *Applied Catalysis A: General*, 340(2), 271-277. <https://doi.org/10.1016/j.apcata.2008.02.025>
- Li, M., Niu, F., Zuo, X., Metelski, P. D., Busch, D. H., & Subramaniam, B. (2013). A spray reactor concept for catalytic oxidation of p-xylene to produce high-purity terephthalic acid. *Chemical Engineering Science*, 104, 93-102. <https://doi.org/10.1016/j.ces.2013.09.004>
- Osada, M., & Savage, P. E. (2009a). Terephthalic acid synthesis at higher concentrations in high-temperature liquid water. 1. Effect of oxygen feed method. *AIChE Journal*, 55(3), 710-716. <https://doi.org/10.1002/aic.11718>
- Osada, M., & Savage, P. E. (2009b). Terephthalic acid synthesis at higher concentrations in high-temperature liquid water. 2. Eliminating undesired byproducts. *AIChE Journal*, 55(6), 1530-1537. <https://doi.org/10.1002/aic.11761>
- Pérez, E., Fraga-Dubreuil, J., García-Verdugo, E., Hamley, P. A., Thomas, M. L., Yan, C., & Poliakoff, M. (2011). Selective aerobic oxidation of para-xylene in sub-and supercritical water. Part 2. The discovery of better catalysts. *Green Chemistry*, 13(9), 2397-2407. <https://doi.org/10.1039/C1GC15138J>
- Savage, P. E. (2009). A perspective on catalysis in sub-and supercritical water. *The Journal of Supercritical Fluids*, 47(3), 407-414. <https://doi.org/10.1016/j.supflu.2008.09.007>
- Tomás, R. A., Bordado, J. C., & Gomes, J. F. (2013). p-xylene oxidation to terephthalic acid: A literature review oriented toward process optimization and development. *Chemical Reviews*, 113(10), 7421-7469. <https://doi.org/10.1021/cr300298j>
- Vakros, J. (2021). Recent advances in cobalt and related catalysts: From catalyst preparation to catalytic performance. *Catalysts*, 11(4), Article 420. <https://doi.org/10.3390/catal11040420>
- Walt, P. (2020). A chemical model for the Amoco "MC" oxygenation process to produce terephthalic acid. In D. W. Blackburn (Ed.), *Catalysis of Organic Reactions* (pp. 321-346). CRC Press. <https://doi.org/10.1201/9781003066446>
- Xu, L., Chen, D., Jiang, H., & Yuan, X. (2020). Efficient oxidation of p-xylene to terephthalic acid by using N, N-dihydroxypyromellitimide in conjunction with Co-benzenetricarboxylate. *Applied Catalysis A: General*, 599, Article 117569. <https://doi.org/10.1016/j.apcata.2020.117569>

Hole Quality Assessment of Drilled Carbon Fiber Reinforced Polymer (CFRP) Panel Using Various Custom Twist Drill Geometries

Muhammad Faris Shauqi Kamaruzaman, Muhammad Hafiz Hassan* and Muhammad Fauzinizam Razali

School of Mechanical Engineering, Universiti Sains Malaysia, 14300 USM, Nibong Tebal, Pulau Pinang, Malaysia

ABSTRACT

Excellent hole quality is necessary for the aerospace industry's highly abrasive carbon fiber reinforced polymer (CFRP) drilling process. This work considered three different twist drill designs, tapered web, burnishing, and subland drill reamer for the drilling process. The drill bits were made of tungsten carbide and machined with custom helix angle, primary clearance, point angle, and chisel angle. The primary objective of this research is to determine the thrust force signature for each custom drill bit design and the delamination factor for the hole drilling at 3000 rev/min and 0.05 mm/rev. The finding indicates that the tapered web gave the best design by improving the maximum thrust force by 14.6% in drilling CFRP panels. Additionally, the tapered web design led to a low delamination factor on both entrance (1.0186) and exit sides (1.0475). The thrust force is directly proportional to the delamination factor when drilling a CFRP material in a single shot operation. The subland drill reamer produces higher thrust force, and delamination proved that the combination of drill and reamer design was unsuitable for high-speed drilling.

Keywords: CFRP, delamination, single shot, thrust force

ARTICLE INFO

Article history:

Received: 06 October 2021

Accepted: 28 January 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.17>

E-mail addresses:

muhdfarisshauqi@gmail.com (Muhammad Faris Shauqi Kamaruzaman)

mhafizhassan@usm.my (Muhammad Hafiz Hassan)

mefauzinizam@usm.my (Muhammad Fauzinizam Razali)

* Corresponding author

INTRODUCTION

Carbon fiber reinforced polymer (CFRP) is extensively utilized in the automobile and aerospace sectors (Jaafar et al., 2019). They have outstanding properties such as high density, high compressive and tensile tension, high operating temperature, and lightness, which reduce a vehicle's weight and fuel usage.

Drilling is one of the most common operations needed for post-processing components made from CFRP materials. Screws and rivets are often used in the assembly process to assemble the CFRP stack with other materials (Jia et al., 2020; John et al., 2021). Although most CFRP components were manufactured in the near-net form to minimize work, additional manufacturing, such as drilling, is still needed for the post-processing work. However, due to the heterogeneity and anisotropy of the CFRP material, the drilling phase can result in a different types of damage, such as fiber pull out and fiber-matrix debonding (Hou et al., 2020). Delamination is the primary damage found after the drilling of composite material. It happens because of the binder surrounding the drilled hole is damaged. This type of damage impairs installation tolerance, compromises structural stability, and is likely to degrade the CFRP product's long-term efficiency.

Assuring the hole accuracy on both sides of the CFRP composite at the entry and exit is challenging. When drilling CFRP, the right selection of drill bit's features, such as helix angle, point angle, cutting angle, and the drill bit's material composition is vital. It is reported that the drill bit with a smaller point angle of 90° successfully reduces the exit damage of the CFRP composite (John et al., 2021). Additionally, utilizing a drill bit with a small point angle results in a lower thrust force while drilling than a drill bit with a wide point angle. The comparative performance of standard twist drill design and special drill (namely "dagger drill") design in the drilling of a high-strength carbon-fiber-reinforced composite process is reported by Feito et al. (2019) and An et al. (2014 & 2013). The dagger drill has been found to promote a better surface finish, i.e., less burr defect and less damage to the delamination than the twist drill design due to smaller point angle and helix angles. However, due to the poor chip evacuation capacity compared to the small helix angle, the "dagger drill" is not preferable for metallic part drilling.

Wika et al. (2011) performed several drilling trials of CFRP-Ti stack-up materials using four types of twist drills with different flute and helix angles. Results revealed that with the large flute volume for chip evacuation and heat dissipation, the two-flute drill bit with a higher helix angle produced the smallest cutting force and the lowest cutting temperature compared to the drill with the three flutes. SenthilKumar et al. (2013) used 118° and 130° point angle drills to examine the effects of point angle on tool performance when drilling the composite/Ti stack. It was determined that the higher point angle (130°) drills outperformed those with lower point angle (118°) in terms of the tool wear and chip evacuation analysis.

Delamination occurs through two distinct mechanisms: entry delamination (peel-up) and exit delamination (push-out). Peel-up delamination occurs due to the drill's advancement; the upper layers of material appear to draw up the drill's cutting face rather than being removed (Higuchi et al., 2020). On the other side, push-out delamination occurs

due to the drilling chisel tip indenting the uncut layers of the CFRP laminate. According to work reported by Liu et al. (2018), delamination initiates when the thrust force of the drill reaches the toughness of the interlaminar layers. This study considers three custom designs of twist drill types have been to achieve the lowest possible thrust force and the least amount of delamination on drilled CFRP panels. In addition, the significant effect of the drill geometry of the design like helix angle and chisel edge angle has been discovered in this research study.

METHODOLOGY

Work Piece Material

The CFRP material was made from 26 plies of unidirectional prepregs carbon composite manufactured by Hexcel Composite Sdn. Bhd. The CFRP panel is 3.25 mm thick, with 0.125 mm thick carbon fiber material in each ply. The stacking of the layers is symmetrical, following the sequences of $[45/135/90_2/0/90/0/90/0/135/45_2/135]_s$. At the top and bottom of the CFRP laminate, a 0.08 mm thin layer of glass/epoxy woven fabrics was used to avoid delamination at the hole's entry and exit throughout the drilling operation. It results in a cumulative thickness of 3.587mm for the whole CFRP panel, including the paint application. The CFRP was compacted using a vacuum pump in a controlled atmosphere throughout the curing process. A mold for the laminate was prepared and placed inside the autoclave. The cure cycle consisted of raising the temperature to 180 °C at 3 °C/min and maintaining it for 120 minutes. Then the temperature was brought down to room temperature at 3 °C/min. The whole cycle was carried out at the pressure of 700 kPa in an autoclave and placed in a vacuum bagging which was evacuated to 70 kPa. Hence, the nominal fiber volume fraction is 60% by applying that curing recipe.

Drill Bit Geometries

Figure 1 depicts three distinct designs of twist drill types: tapered web, burnishing, and subland drill reamer. Gandtrack Asia Sdn. Bhd. manufactures this drill bit with a diameter of 6.35mm and a tolerance of h8 (+0, -20 μ m). The drill bit was made of tungsten carbide due to its unique properties, such as tolerance to extreme temperatures

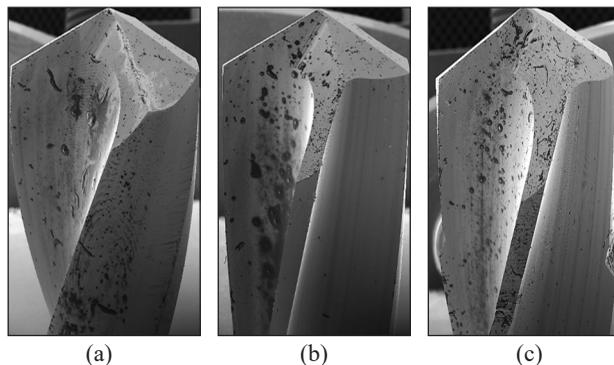


Figure 1. Tungsten carbide drill bit: (a) tapered web; (b) burnishing; and (c) subland reamer design

and retaining a strong cutting edge even after drilling multiple holes (Katiyar et al., 2016). The Tungsten Carbide (WC) rod composition consists of WC~93.36 wt % and Cobalt~6.64 wt %. It has a density of 14.35 g/cm³ and a hardness value of 1625 HV, significantly higher than the workpiece material. Therefore, the drilling tools can easily shear the surface of the workpiece material without causing the breakage of the tool itself. Table 1 summarizes the detailed range of the drill geometry used in this study.

Table 1
Drill bit geometry configurations

	Tapered web	Burnishing	Subland reamer
Helix angle	25°	11°	25°
Primary clearance	6°	6°	8°
Point angle	120°	120°	120°
Chisel edge	30°	30°	45°

Thrust Force Measurement

As shown in Figure 2, the drilling process was carried out on a CNC High-Speed Milling unit, model Alpha T2liFB. The thrust force signal was recorded during the drilling operation using a dynamometer (Kistler four-component dynamometer model 9272) mounted at the bottom of the jig. The data acquisition system, connected to the dynamometer, consists of a multichannel charge amplifier (type 5070) and Kistler DynoWare software. The thrust force and torque signature were generated when the four-component dynamometer transferred the charge signal to the multichannel charge amplifier. The multichannel charge amplifier converts the resulting charge signal to voltage, proportional to the applied force. The resulting signals were converted to force by the calibrated data displayed in the software. Drilling conditions were similar for the three drill bit designs, with a 3000 rev/min speed

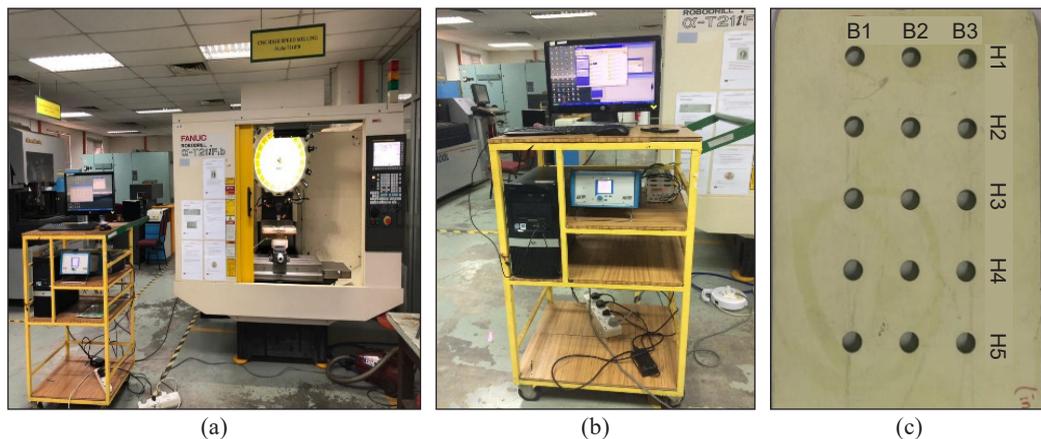


Figure 2. Setup for CFRP drilling: (a) Positioning of work-piece in the CNC; (b) data acquisition system for thrust force measurement; and (c) labeling of hole location

and a 0.05 mm/rev feed rate. Five holes were drilled (Figure 2c) to obtain the average thrust force for each drill bit. As force is sensed during the drilling process, data is sent to the data acquisition system for force signature against cutting time.

Delamination Observation

Delamination was measured at the entry and exit of drilled holes using a 5×/0.15 magnification Alicona Infinite Focus Optical microscope (IFM G4 System). As shown in Figure 3, the CFRP panel was positioned next to the lens to facilitate scanning and capturing the image of the hole. The delamination of the CFRP panel was evaluated on both the entry and exit sides to distinguish between the two types of delamination.

The scanned hole images were post-processed in the *Image J* program (Figure 3b) to measure the delamination area on the CFRP panel's entrance and exit sides. The following Equation 1 is used to calculate the value of the delamination factor,

$$F_d = \frac{A_{\max}}{A_{\text{nom}}} \quad [1]$$

Where F_d is the delamination factor, A_{\max} is the hole damaged area, and A_{nom} is the nominal area.

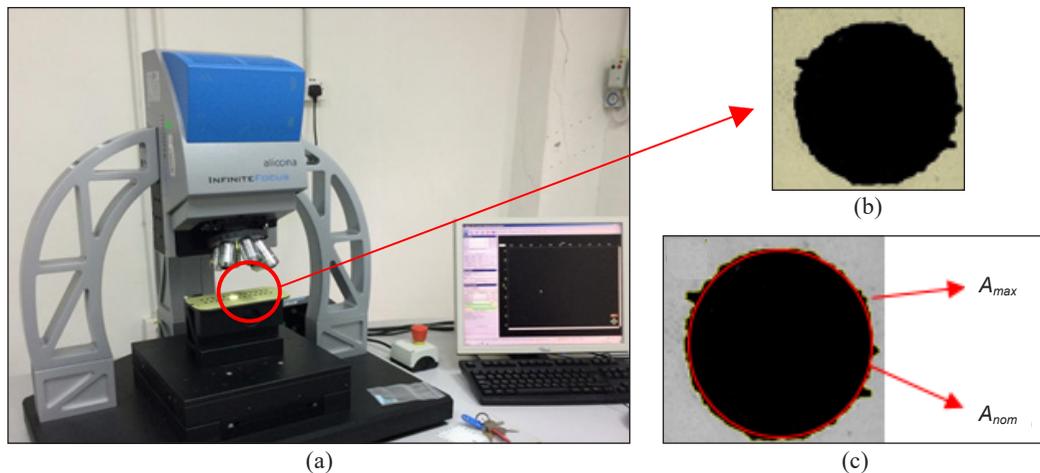


Figure 3. (a) Hole delamination observation; (b) sample delamination image recorded at the hole entrance location; and (c) processing image used for delamination factor measurement

RESULTS AND DISCUSSION

Thrust Force Analysis

Figure 4 depicts the thrust force signature and average maximum thrust force of tapered web, burnishing, and subland drill reamer. The thrust force signature for tapered web and

burnishing drills can be separated into three regions: the initial stage after the drill bit contacts the CFRP panel along the cutting lips, the drilling phase, the CFRP thickness, and the final stage after the drill bit has penetrated the panels. However, for the subland drill reamer configuration, the thrust force signature consists of five regions, where the two additional regions are associated with the reamer phase. The drilling operation for the subland drill reamer configuration began with a 6.2 mm drill diameter and progressed to a 6.35 mm drill diameter with four cutting edges that act as reamers.

The maximum thrust force, F_{tmax} , was introduced in this study to determine the optimal thrust force output of a drill bit configuration. The F_{tmax} is the maximum value of force measured from the force signature during the drilling process. Figure 4(b) shows that the tapered web produced a minimum F_{tmax} value of 193.25 N. It is believed that the fast helix angle of the tapered web was improved by 14.6% from burnishing design which is a slow helix type. The fast helix type would help evacuate the chips from the hole to the surrounding, hindering the hot dust chip from accumulating at the cutting surface of the drill bit. It is agreed with Ashrafi et al. (2013), which also drilled a CFRP panel at the fast helix geometry and achieved minimum delamination at the hole exit.

Subland drill reamer contributed the largest thrust force signature, with a F_{tmax} value of 248.99 N. At the drilling stage of the subland drill reamer design, the drill bit was hard to penetrate the CFRP panel due to the higher chisel edge angle, which is 45° . It is worth noting that raising the chisel edge angle of the drill bit from 30° to 45° increases the F_{tmax} by 18.59%. When the chisel edge angle increases, the shortened cutting lips result in a less effective cutting operation.

Similarly, Figure 5 shows the torque measurement ranging from 0.4 Nm to 0.5 Nm for drilling CFRP materials. Again, the highest torque values are produced by subland drill

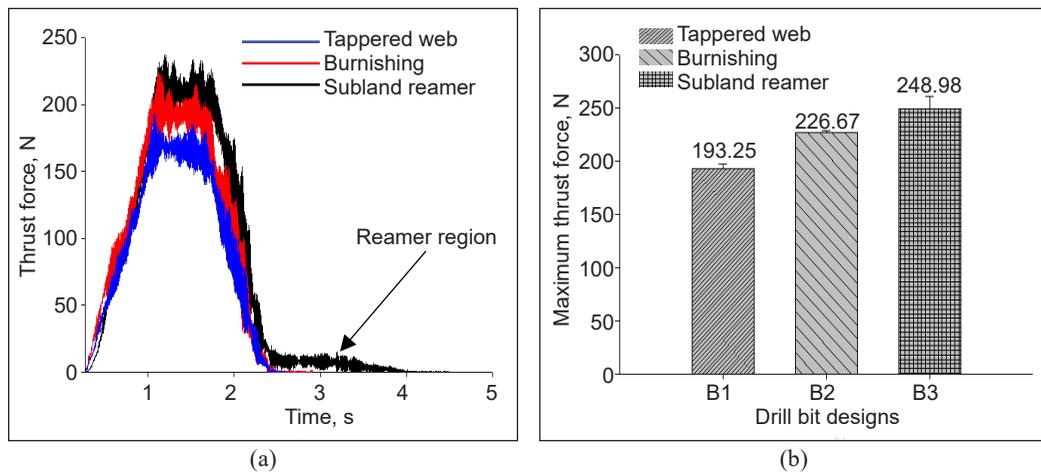


Figure 4. (a) Thrust force distribution for all drill bit design; and (b) maximum thrust force generated during the drilling process

reamer design. The lowest torque values are produced by tapered web design. Again, the thrust force and torque trend are the same, influenced by the tool geometry.

Delamination Analysis

Figure 6 illustrates the results for entry delamination (peel-up) and exit delamination (push-out). The formation of entry delamination at the hole entrance is considered a minimum for all drill bit designs. The various bit designs only result in substantial damage and uncut fiber at the drilled hole's exit. The subland drill reamer's delamination was more severe than the tapered web and burnishing drill bit. This severe exit delamination issue happens as the drill bit acts like an extruder rather than a cutter. Uncut fiber is also visible on the exit tapered web, which happens because of push-out delamination in the fiber's final layer. The work of Yang et al. (2019) has claimed that while using a reamer drill bit, the pace of the cut must be reduced to prevent these two issues while drilling. That was stated that using the reamer drill bit must be cut at a lower speed.

As shown in Figure 7, the composite matrix will be affected if the thrust force exceeds the bonding strength between laminate during the drilling process. It will lead to delamination damage, especially at the exit part of the panels. Besides, when drilling CFRP materials at a higher point angle, the cutting edge will easily produce downward bending deformation; then, the fibers cannot be cut by the cutting edge, which leads to the uncut fiber in the hole exit. According to the drilling process in Figure 7, the point angle of the drill was ϕ ; the force of the drill on the hole-exit CFRP can be simplified as the force F , which can be illustrated as radial cutting force F_x and thrust force F_z . As the resin bonded to the fiber, the intensity of the bonding force was P_b , and the fiber plastic/

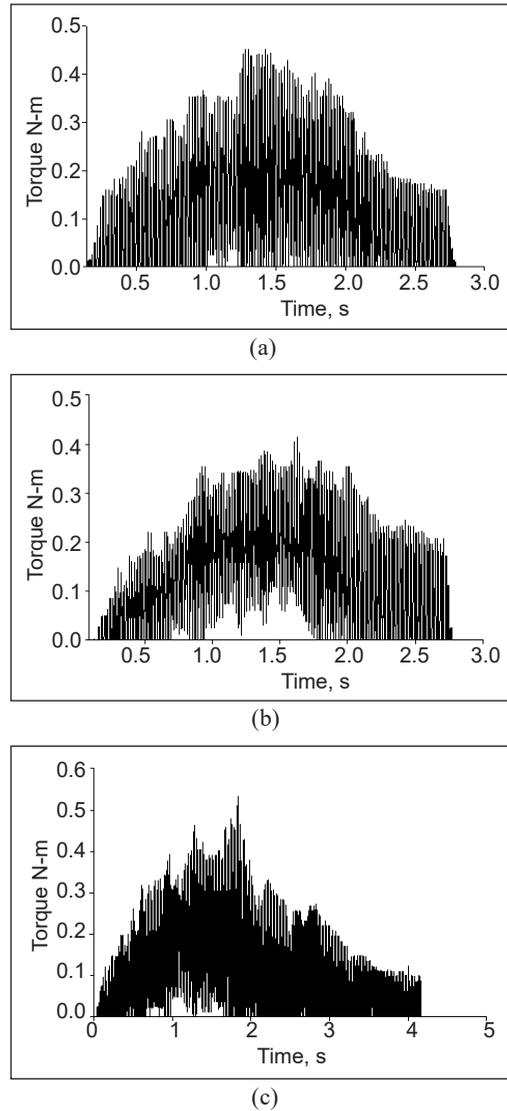


Figure 5. Measurement of torque: (a) tapered web design; (b) burnishing design; and (c) subland drill reamer design

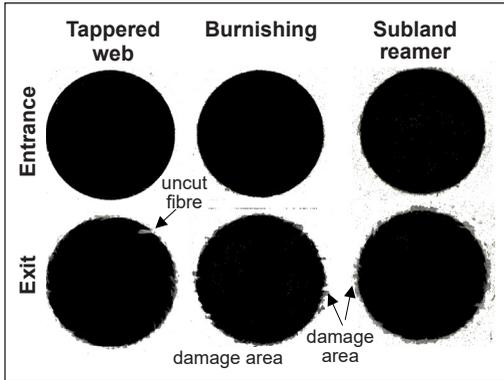


Figure 6. Observation of delamination at the entrance and exit sides of the hole drilled by each drill bit design

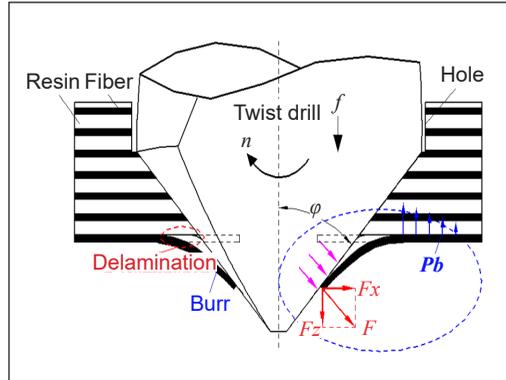


Figure 7. Delamination formation at different point angle

polymer bonding strength was σ_b . When drilling a CFRP at a constant point angle (120°) and higher primary clearance (8°), it would provide a higher thrust force and torque hence contributing to the disbond of the laminate at the hole-exit.

The F_d 's delamination factor was introduced, as illustrated in Figure 8, to quantify the value of delamination at the entrance and exit of the drilled hole. The $F_{d-entrance}$ and F_{d-exit} values were 1.0186–1.0221 and 1.0475–1.0759, respectively. The tapered web drill bit type recorded the lowest delamination factor on the entrance and exit sides because the value is closer to the nominal value of 1.0. This nominal value signifies the tapered web's best cutting efficiency than burnishing and subland reamer drill bits. On the other hand, the subland reamer drill bit demonstrated the largest delamination factor, 1.0221 and 1.0759, respectively, with an increment of 2.21 % and 7.59 % from the nominal line. It may be because the drilling parameters, 3000 rev/min and 0.05 mm/rev, are incompatible with the subland reamer drill bit type used to drill the CFRP panel.

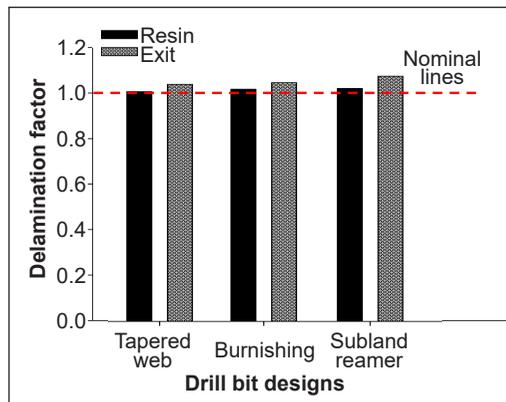


Figure 8. Magnitude of delamination factor measured at the entrance and exit sides of the hole drilled by each drill bit design

CONCLUSION

This paper investigated custom twist drill designs' thrust force magnitude and delamination factor while drilling CFRP panels. The experimental results lead to the following conclusions.

The best design of the twist drill type is a tapered web since it generates the least thrust force during the CFRP drilling operation.

When the helix angle design was increased from the fast helix to the slow helix, the effect of the helix angle would improve the F_{tmax} by 14.6% when drilling a CFRP panel. However, for the chisel edge angle, when increasing the chisel edge angle from 30° to 45°, the F_{tmax} would increase to 18.59%.

The delamination factor of the CFRP panel is directly proportional to the F_{tmax} obtained from the drill bit design. Therefore, the higher F_{tmax} would be contributed to the higher delamination factor.

ACKNOWLEDGEMENT

The authors want to thank the Spirit Aero System Sdn Bhd and Gandtrack Asia Sdn Bhd for providing the raw materials for this work. Short Term Grant (304/PMEKANIK/6315557) funded this research. In addition, the authors greatly appreciate the School of Mechanical Engineering USM's facility support.

REFERENCES

- An, Q. L., Xu, J. Y., Cai, X. J., & Chen, M. (2013). Experimental investigation on drilling force and hole quality when drilling of T800S/250F CFRP laminate. *Advanced Material Research*, 797, 155-160. <https://doi.org/10.4028/www.scientific.net/amr.797.155>
- An, Q., Cai, X., Xu, J., & Chen, M. (2014). Experimental investigation on drilling of high strength T800s/250f CFRP with twist and dagger drill bits. *International Journal of Abrasive Technology*, 16(3), 183-196.
- Ashrafi, S. A., Sharif, S., Farid, A. A., & Yahya, M. Y. (2013). Performance evaluation of carbide tools in drilling CFRP-Al stacks. *Journal of Composite Materials*, 48(17), 2071-2084. <https://doi.org/10.1177%2F0021998313494429>
- Feito, N., Muñoz-Sánchez, A., Díaz-Álvarez, A., & Miguelez, M. H. (2019). Multi-objective optimization analysis of cutting parameters when drilling composite materials with special geometry drills. *Composite Structures*, 225, Article 111187. <https://doi.org/10.1016/j.compstruct.2019.111187>
- Higuchi, R., Warabi, S., Ishibashi, W., & Okabe, T. (2020). Experimental and numerical investigations on push-out delamination in drilling of composite laminates. *Composites Science and Technology*, 198, Article 108238. <https://doi.org/10.1016/j.compscitech.2020.108238>
- Hou, G., Zhang, K., Fan, X., Luo, B., Cheng, H., Yan, X., & Li, Y. (2020). Analysis of exit-ply temperature characteristics and their effects on occurrence of exit-ply damages during UD CFRP drilling. *Composite Structures*, 231, Article 111456. <https://doi.org/10.1016/j.compstruct.2019.111456>
- Jaafar, M. F., Salleh, M. S., Izamshah, R., Hassan, M. H., Sundi, S. A., Hafiz, M. S. A., & Kasim, M. S. (2019). Influence on thrust force and delamination for one shot drilling of carbon fibre reinforced plastic (CFRP). *International Journal of Mechanical & Mechatronics Engineering*, 19(1), 43-56.

- Jia, Z. Y., Zhang, C., Wang, F. J., Fu, R., & Chen, C. (2020). An investigation of the effects of step drill geometry on drilling induced delamination and burr of Ti/CFRP stacks. *Composite Structures*, 235(1), Article 111786. <https://doi.org/10.1016/j.compstruct.2019.111786>
- John, K. M., Kumaran, S. T., Kurniawan, R., & Ahmed, F. (2021). Evaluation of thrust force and delamination in drilling of CFRP by using active spring restoring backup force. *Materials Today: Proceedings*, 49, 269-274. <https://doi.org/10.1016/j.matpr.2021.02.033>
- Katiyar, P. K., Singh, P. K., Singh, R., & Kumar, A. L. (2016). Modes of failure of cemented tungsten carbide tool bits (WC/Co): A study of wear parts. *Journal of Refractory Metals and Hard Materials*, 54, 27-38. <https://doi.org/10.1016/j.jrmhm.2015.06.018>
- Liu, L., Qi, C., Wu, F., Zhang, X., & Zhu, X. (2018). Analysis of thrust force and delamination in drilling GFRP composites with candle stick drills. *The International Journal of Advanced Manufacturing Technology*, 95(5), 2585-2600. <https://doi.org/10.1007/s00170-017-1369-8>
- Senthilkumar, M., Prabukarthi, A., & Krishnaraj, V. (2013). Study on tool wear and chip formation during drilling carbon fiber reinforced polymer (CFRP)/titanium alloy (Ti6AL4V) stacks. *Procedia Engineering*, 64, 582-592. <https://doi.org/10.1016/j.proeng.2013.09.133>
- Wika, K. K., Sharman, A. R. C., Goulbourne, D., & Ridgway, K. (2011). *Impact of number of flutes and helix angle on tool performance and hole quality in drilling composite/titanium stacks* (No. 2011-01-2744). SAE Technical Paper.
- Yang, Y., Yang, Y., Liu, X., Huang, K., & Ren, H. (2019). Optimized design and application of a directional reaming-while-drilling polycrystalline diamond compact bit. *Engineering Failure Analysis*, 105, 699-707. <https://doi.org/10.1016/j.engfailanal.2019.07.023>

ZnO Multilayer Thin Films as the Seed Layer for ZnO Nanorods: Morphology, Structural and Optical properties

Rohanieza Abdul Rahman^{1,2}, Muhammad AlHadi Zulkefle^{1,2}, Sukreen Hana Herman^{1,3*} and Rosalena Irma Alip^{1,4}

¹Integrated Sensors Research Group, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

²NANO-Electronic Centre, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

³Microwave Research Institute (MRI), Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

⁴Terahertz Sensing and Research Team, RIKEN Center for Advanced Photonics, RIKEN, Sendai, Japan

ABSTRACT

The effect of zinc oxide (ZnO) multilayer thin film thicknesses, deposited via the sol-gel spin coating technique, on the morphology, structural and optical properties of ZnO nanorods (ZNR) grown on the ZnO thin films were explored in this investigation. The ZNR was grown using the chemical bath deposition method on the ZnO thin film seed layer (SL). We found that ZnO thin film SL morphology changes according to the number of layers based on the results. Eventually, these changes also influence the structures of ZNR. ZNR structures improved when the thickness of the seed layer increased. Besides the surface roughness, better crystalline quality films were obtained when more layers were deposited. This crystalline quality then influenced the optical characteristics of both ZnO and ZNR thin films. The optical properties from UV-Vis showed transmittance in the visible region, showing that the ZnO films produced were suitable to be applied to solar cells. ZNR-based solar cells have become one of the promising materials to be studied further due to the environment-friendly, low-cost, and well-abundant material for solar cell applications.

ARTICLE INFO

Article history:

Received: 07 October 2021

Accepted: 02 March 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.18>

E-mail addresses:

rohanieza.abdrahman@gmail.com (Rohanieza Abdul Rahman)

alhadizulkefle@gmail.com (Muhammad AlHadi Zulkefle)

hana1617@uitm.edu.my (Sukreen Hana Herman)

rosalena@uitm.edu.my (Rosalena Irma Alip)

* Corresponding author

Keywords: Chemical bath deposition, crystallinity, nanorods, seed layer, solar cells, sol-gel, spin coating

INTRODUCTION

Zinc oxide (ZnO) is classified as an n-type semiconductor that could be found abundantly. ZnO is classified in II-VI semiconductors and has become favorable

among researchers because it owns a wide band gap (3.2–3.4 eV) and is also low-cost and non-toxic. These characteristics have left it within the detectable region that possesses high optical transmission, sufficient for the basic range sun detection (Rwenyagila et al., 2014). Therefore, it has wide applications in optoelectronic devices due to its electrical, structural, and optical properties (Khan et al., 2021; Sharma et al., 2014; Djuriscic et al., 2010). Besides, ZnO also possesses numerous families of nanostructures. Numerous studies show that ZnO could be synthesized to produce nanoring, nanobelts, nanowires, and nanorods (Alenezi et al., 2018; Chen et al., 2021; Mohammadzadeh et al., 2020).

Nanorods are one of the ZnO nanostructures that have been greatly explored by many researchers (Huey et al., 2021; Kannan et al., 2020; Zhou et al., 2019). ZnO nanorods (ZNR) are 1-dimensional (1D) nanostructures with fascinating physical properties. Moreover, 1D structures like ZNR are believed to have more surface-to-volume ratio than 2-dimensional (2D) structures; thus, this type of nanostructures offers an enhancement to increase the sensitivity in many applications (Zhang et al., 2012). Various methods could be utilized to grow ZNR, such as vapor-liquid-solid (VLS), physical vapor deposition (PVD), pulse laser deposition (PLD), chemical vapor deposition (CVD), and the most favorable approach among researchers is chemical bath deposition (CBD) (Kumar et al., 2018; Jimenez-Cadena et al., 2010; Hajezi et al., 2008; Rodriguez-Martinez et al., 2020; Mosalagae et al., 2020). CBD is a favorable method to synthesize ZNR owing to the ability to produce good crystallinity quality films, aside from the low cost and low-temperature process. Many researchers are actively studied in controlling the morphology and structures of ZNR using this CBD method (Roy et al., 2013; Khranovsky et al., 2012; Abdulrahman et al., 2020). Many reported studies to show that vital parameters could control the growth of ZNR, such as growth temperature and duration, type of SL, precursor concentration, pH of the aqueous solution, and thickness of SL (Abdulrahman et al., 2020).

Considering all these essential considerations, the SL becomes an important element in controlling the growth of ZNR. Hence, the SL's thickness is influential in producing a well-aligned ZNR. A few approaches controlled this thickness. One of the approaches considered is depositing a multilayer of SL during the deposition process.

Multilayer or layer by layer ZnO can be produced or deposited by many methods. The method of deposition includes RF magnetron sputtering, pulse laser deposition (PLD), spray pyrolysis, dip-coating, and sol-gel spin coating techniques (Hasabeldaim et al., 2020; Al Farsi et al., 2021; Regmi & Velumani, 2021; Abdel-Galil et al., 2021; Lokesh et al., 2020). Advantages such as simple, practical in terms of cost, and large area deposition technique make sol-gel spin coating favorable among researchers. According to the previous research and findings, it is important to determine a suitable thickness for the SL since it can be an essential factor in producing well-aligned ZNR (Banari et al., 2021). Well-aligned ZNR offers more effective performance in many applications. For example, Gunes et al. (2007)

stated that aligned ZnO nanorods are employed in solar cell applications to raise the persistence of the cell. In addition, the movement of the charge carrier was also regulated by the tidy rods that were constructed by well-aligned nanorods. Based on this, the suitable thickness of the ZnO SL needs to be controlled to produce a well-aligned ZNR. In this investigation, multilayer ZnO thin films were deposited layer by layer via the sol-gel spin coating technique to obtain a suitable thickness for ZnO thin film that functions as the SL for ZNR. One to five layers of ZnO were deposited, and these prepared ZnO thin films were applied as the SL to grow ZNR by the CBD method. All ZnO and ZNR films were characterized morphologically, structurally, and optically to study the impact of the SL thickness on the ZNR properties.

METHODOLOGY

This study can be separated into two different parts. The first part is the deposition of ZnO thin films with different layers, followed by the growth of ZNR on five different numbers SL. Then, after the fabrication process of ZnO and ZNR thin films was performed, the characterization process for physical, structural, and optical properties was carried out. All these processes are explained in the next subsections. Finally, the summary of the overall methodology is illustrated in Figure 1.

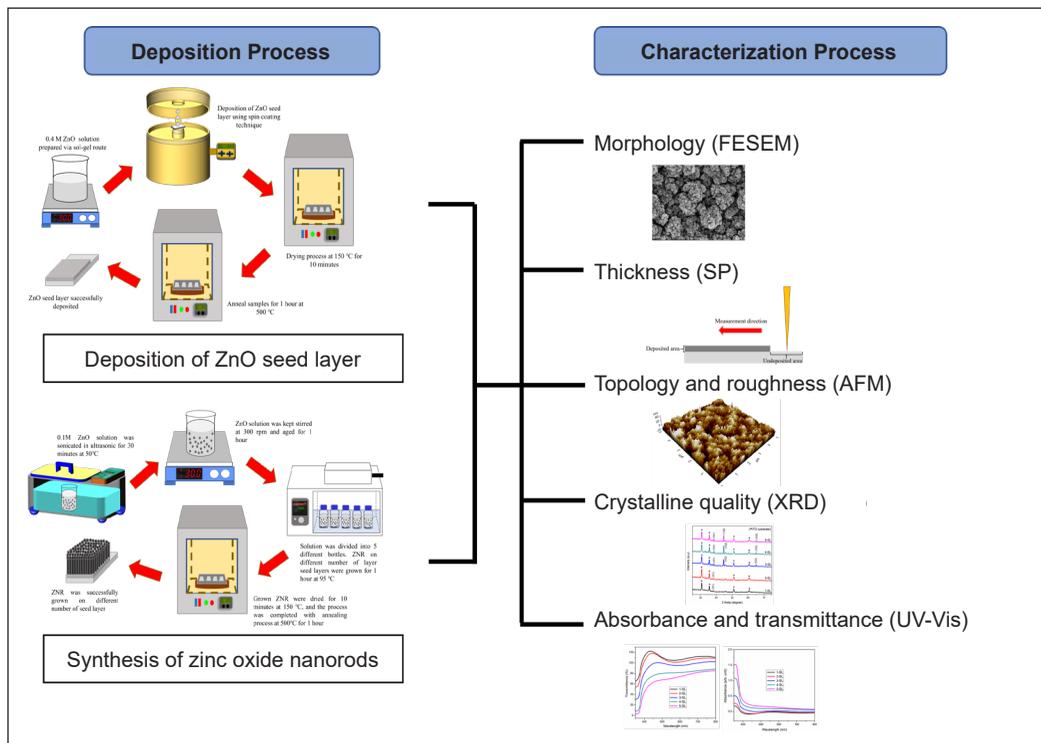


Figure 1. Summary of deposition and characterization process

Deposition of ZnO Thin Films by Sol-Gel Spin Coating Technique

0.4M ZnO solution was produced via the sol-gel method by mixing the following chemicals: zinc acetate anhydrate, monoethanolamine (MEA), and 2-methoxyethanol. For the precursor, zinc acetate anhydrate was used, while MEA acted as the stabilizer, and for the solvent, 2-methoxyethanol was employed. At first, zinc acetate was weighed and put in a Schott bottle. 2.4 mL of MEA was poured into the bottle before 2-methoxyethanol was added until the total solution was 100 mL. This mixture was stirred on a hot plate at 300 rpm with 80 °C heat. The heat was switched off after 3 hours, and the stirring process continued as the aging process for 24 hours to produce a homogenous and transparent (clear) 0.4M ZnO solution.

The deposition process was performed via the spin-coating technique when the aging process was completed. ZnO thin film had been deposited with different numbers of SL. This investigation employed indium tin oxide (ITO) as a substrate. Beforehand, ITO substrates were cleaned using a standard cleaning process, using ethanol and deionized water (DI). The spin-coating process started by placing the ITO substrate on the stage of the spin-coater (center), then ZnO solution (10 drops) was dropped onto the spinning substrate. This spinning process was set up at 3000 rpm for 60 seconds. After that, the deposited sample was let dry in a furnace at 150°C for ten minutes. These processes were repeated to obtain ZnO thin films with 1, 2, 3, 4, and 5 layers. Each layer was dried after being deposited with ZnO solution. Finally, the annealing process was carried out for one hour, at 500°C, after the last layer was deposited to complete the deposition process. Figure 2 shows the experimental diagram for ZnO thin film deposition process, while Table 1 describes the classification of the samples produced in this investigation.

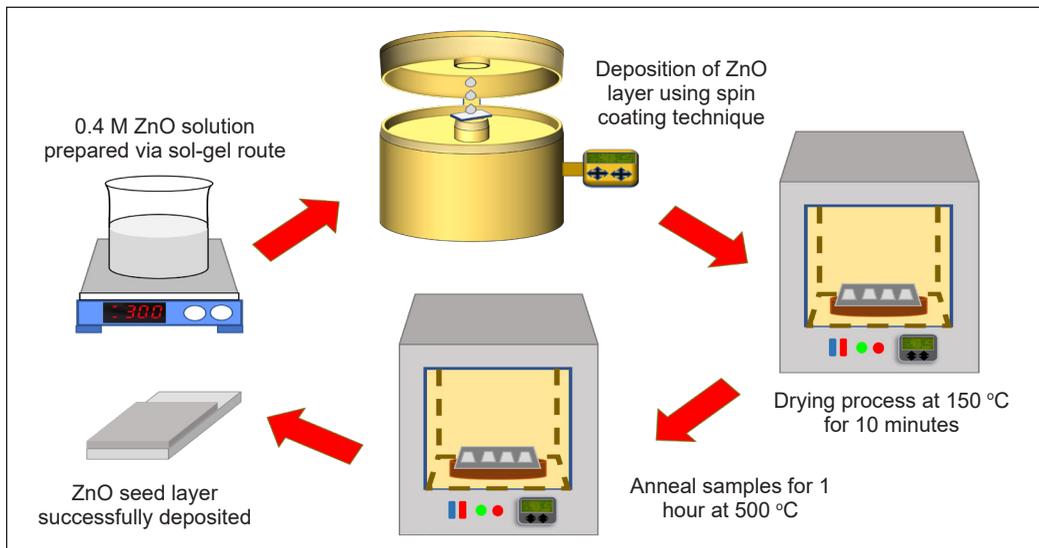


Figure 2. Sol-gel spin coating process for ZnO thin films samples with different number of layers

Table 1
 Classification for ZnO samples produced

Number of deposition layer	1	2	3	4	5
Classification	1-SL	2-SL	3-SL	4-SL	5-SL

Growth of Zinc Oxide Nanorods via Chemical Bath Deposition Method

A 0.1M ZnO solution was first prepared by dissolving zinc nitrate hexahydrate ($\text{Zn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, 99%) and hexamethylenetetramine ($(\text{CH}_2)_6\text{N}_4$, 99.5%) in deionized water (DI) to grow ZNR. Then, the molar ratio of $\text{Zn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ and $(\text{CH}_2)_6\text{N}_4$ was fixed to 1:1, and DI water was poured into the used beaker to get 500 mL of ZnO solution. The sonication process was the next step using an ultrasonic bath (Hwasin Technology PowerSonic 405, 40 kHz). ZnO solution was sonicated at 50 °C for 30 minutes. After that, the solution was kept stirred on a hot plate at 300 rpm for 1 hour at room temperature. When the aging process was completed, the prepared ZnO solution was divided into five different Schott bottles, with 100 mL volume each. ZnO thin films with the different number of layers deposited earlier were immersed upside down in each Schott bottle.

Then, all these bottles were put into a hot bath (Mettmert) that contained preheated DI water. The hot bath temperature was set to 95 °C, and this growth process was fixed to one hour. After one hour, all the samples were taken out of the hot bath. These samples were first rinsed with DI water to eliminate the precipitate powder before drying at 150 °C for 10 minutes. Finally, the annealing process was performed for one hour at 500 °C to complete the growing process of ZNR. In summary, these processes are illustrated in Figure 3. The chemical in Equations 1 to 5 described the growth mechanism of ZnO nanorods.

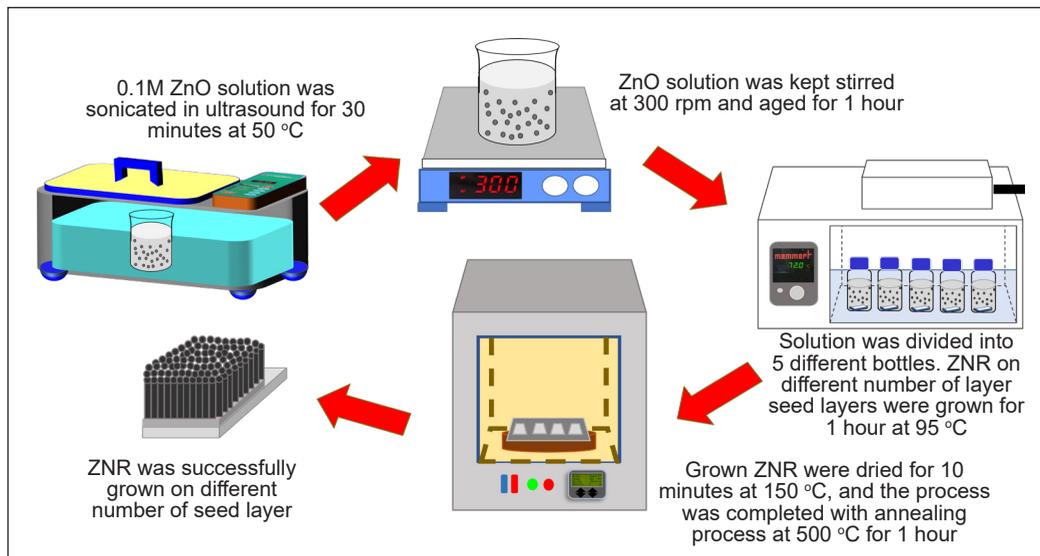
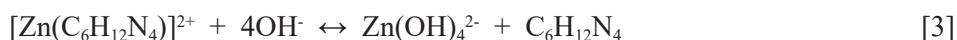
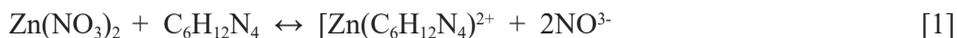


Figure 3. The growing process of ZNR by chemical bath deposition



Characterization of ZnO and ZNR Thin Film

All the deposited ZnO and ZNR thin films were characterized to examine their physical, optical, and structural properties. Morphological of ZnO and ZNR were explored by field emission scanning electron microscope (FESEM, Hitachi SU-030), while the optical properties were explored with an ultraviolet-visible spectrometer (UV-Vis, Jasco/V-670 EX). A surface profilometer (SP, KLA-Tencor P-6 Stylus Profiler) was chosen to measure the thickness of all samples in this study. An atomic force microscope (AFM, Park Systems XE-100) characterized the surface topology and the roughness of thin films. The characterization was performed using X-ray diffraction (XRD, PANalytical X'Pert PRO) to determine all samples' crystallinity quality (structural properties). The results obtained from these characterizations are present and discussed in the next section.

RESULTS AND DISCUSSION

Morphology of ZnO thin films deposited with the different number of layers was observed using FESEM. Figure 4 presents the images obtained for ZnO thin film deposited with 1 and 5 layers. A contrast change could be identified in Figures 4(a) and (b), in which the structures of ZnO thin films become denser and uniform when further layers are deposited. Even though 1-layer ZnO thin film shows a uniform structure, small cracks could be observed. These cracks improved the increased number of deposition layers. Crack-free and clear agglomeration structure was shown in Figure 4(b), proving that several layers produce a better structure of ZnO thin films. The difference between these structures might be from the spin-coating process. The uniformity of a thin film could be affected by the spin-coating process.

In contrast, in a 1-layer ZnO thin film, the distribution of ZnO nanoparticles was not evenly spread, thus producing less uniform ZnO structures. Increasing the deposition layer could overcome this issue. The non-uniform ZnO nanoparticles were then covered by the subsequent layers of ZnO nanoparticles, which resulted in a more uniform structure of ZnO thin film. The uncovered area was covered by the next layer of ZnO deposition during the first layer, resulting in better and uniform structures. This statement was supported

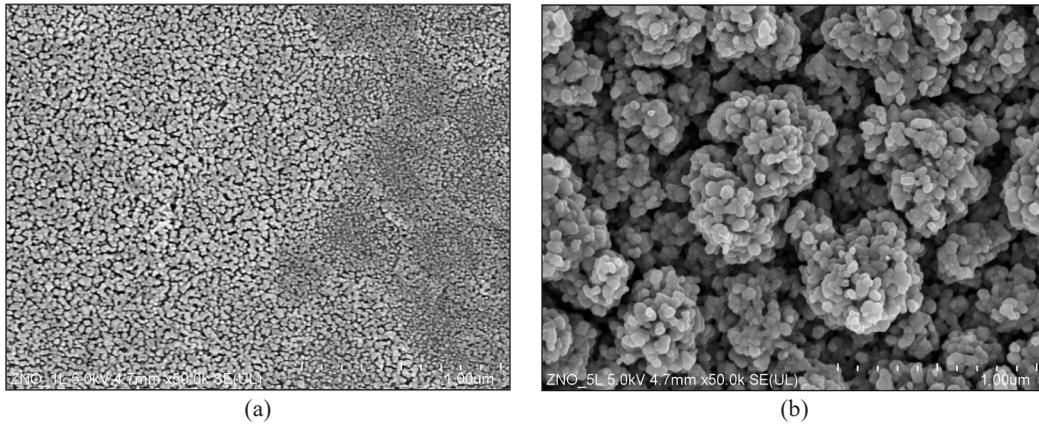


Figure 4. Comparison of ZnO thin films deposited with (a) 1-layer and (b) 5-layers

by the FESEM images in Figure 4(b), in which an improvement of ZnO nanostructures was achieved when the number of layers was increased. This finding indicates that good quality ZnO thin films (dense and crack-free) could be produced with the increment of deposition layers.

In a study reported by Kumar et al. (2013), the same parameter was varied. However, the thickness for each layer was constantly deposited, which is ~ 19 nm. According to their findings, the morphology and structure of their ZnO thin films were improved as the number of depositions was increased (Kumar et al., 2013). It is due to the thickness increment of the thin films when further layers are deposited. Based on their study, Kumar et al. (2013) also stated that thicker film produces better morphology of ZnO thin films. Similar to this study, particles in the 5-layer ZnO thin film show a higher thickness, correlated with the increase in the number of layers, compared with the 1-layer ZnO thin film. This statement is proven by the surface profilometer (SP) measurement. The thickness of the samples significantly changes with the number of layers. Table 2 describes the thickness value for all samples deposited with different layers.

Table 2
The value of thickness ZnO thin films deposited with variation layers

Sample	1-SL	2-SL	3-SL	4-SL	5-SL
Film thickness (nm)	58.52	123.23	195.03	236.85	284.20

Table 2 shows the thickness values for each ZnO thin film deposited with a different number of layers. Based on the results, the thickness of the samples shows a proportional relationship with the number of deposition layers. Significant changes could be seen in each sample. As shown in Table 2, an almost constant value was found when each layer was deposited on the existing ZnO layer. An average value of ~ 56 nm was determined

based on the measured data. This finding proved that the addition of each layer contributes to the increasing thickness trend of ZnO thin films with a different number of layers. This occurrence might come from the amount of ZnO nanoparticles that rise accordingly with the deposition layer. A higher number of layers provide many ZnO nanoparticles settling on the substrate or existing layer of ZnO (Khan et al., 2017). Hence, a higher value of thickness is produced. This attained result can be related to the finding reported by Shariffudin et al. (2012) in which ZnO thin films in their study had higher thickness when several layers were added. They varied the number of layers during the spin-coating process, the thickness of the films changed, and the grain size for ZnO nanoparticles (Shariffudin et al., 2012).

Other than that, since the total thickness of ZnO thin film deposited with 5-layers is almost five times that of a 1-layer sample, we attribute this linear increment of the thickness to the drying process that we conducted at the end of each layer during the deposition process. The drying process improved and allowed the particles in one layer to be stable and dense before the subsequent layer was deposited. Regarding this drying process, Addamo et al. (2008) stated that when each layer was dried, the layer would grow over irregular and preformed the crystalline surface or structure, resulting in higher thickness for the films. It was based on their finding that they acquired thicker films when every layer was dried before the next layer was deposited. Besides, the increment of the thin film's thickness had influenced the surface topology and roughness of the ZnO thin films. It was proven by the results obtained from AFM measurement, as shown in Figure 5.

Figures 5(a) to 5(e) presented the 3-dimensional (3-D) images of ZnO thin films deposited for this study. The images indicate that the addition of the layer on top of the previous layer had affected the topology of the sample and the surface roughness. It could be observed that the roughness of the ZnO thin films increased with the addition of the number of layers, as the last sample with the highest thickness possesses the roughest surface. According to the AFM analysis, the surface roughness for all samples was recorded in the range of 4.77~9.20 nm, respectively. The rising surface roughness might be attributed to the formation of the grains that becomes larger significantly with the changes in thickness (Kumar et al., 2013). Besides, the correlation between the thickness of thin films and the film's roughness was also investigated by Kaiyong et al. (2005) and Daniel et al. (2014). Their studies concluded that roughness varies with thickness, where when thin films become thick, the film roughness increases. It corresponds with the findings in this study. This finding could be related to the agglomeration of the samples, which influenced the surface roughness of the thin films. Since the number of layers was increased, the total amount of available ZnO particles also increased. The ZnO particles then agglomerate, making the surface rougher.

The crystalline quality of the samples was influenced by the difference in the thickness value in this study. Figure 6 revealed the XRD spectra for all samples with different numbers of the deposition layer.

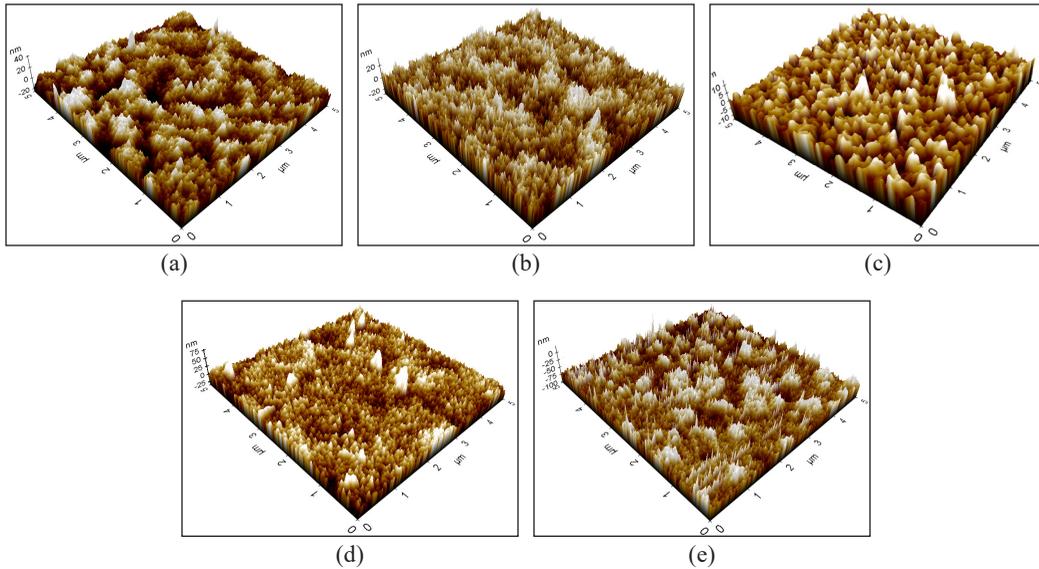


Figure 5. Surface topology of ZnO thin films with different number of deposition layers

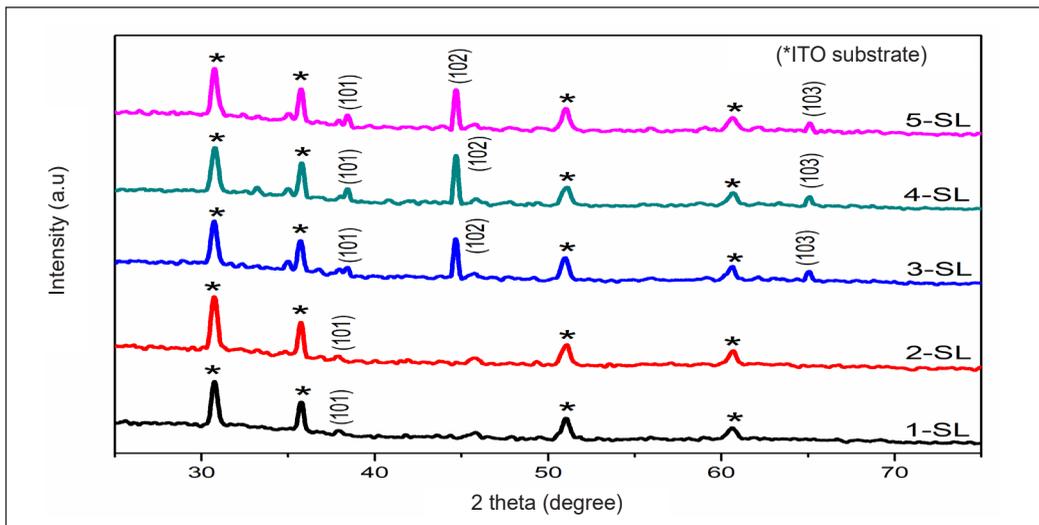


Figure 6. XRD analysis for all thin films deposited with different number of layers

The XRD patterns for all ZnO deposited with different layers are shown in Figure 6. Referring to the XRD pattern obtained, it is clearly shown that all deposited films are in good concurrence with the standard ZnO diffraction pattern of JCPDS no 36-1451. Referring to the XRD spectra, the diffraction peaks 2θ of $\sim 37\text{-}38^\circ$, $\sim 44.69^\circ$, and 65.11° , which correspond to (101), (102), and (103) crystalline peaks of ZnO, were observed. It could be seen that (101) peaks appear in all samples (1–5 layers). However, the intensity increment can be detected when further layers are deposited. The intensity increased in

the 3-SL sample and became higher at 4 and 5-SL. This (101) also shifted to the right due to improved crystalline quality. Meanwhile, (102) and (103) peaks are only present in 3, 4, and 5-SL ZnO thin films. Both peaks could not be observed in 1 and 2-SL samples but started to appear when ZnO thin film was deposited with three layers.

Among all peaks, (102) becomes dominant, and the highest intensity can be seen in 5-SL ZnO thin films. This result might be attributed to the thickness of the thin films that increase proportionally with the number of layers. The thicker film has more atomic layer to diffract X-rays (Goncalves et al., 2017). For the changes in the crystal orientation, we assume that as the layer increases, more atoms somehow tend to align in the (102) direction and pass the crystallographic information to the subsequent layers. The drying process conducted at each layer also might influence the crystalline quality of ZnO thin films. During the first layer deposition, the ZnO thin films were randomly oriented (Shariffudin et al., 2012). After the layer was dried, the stabilizer and solution vaporized. Then, after the deposition of the second layer, the first layer becomes a shape or template for the next layer, thus supporting the orientation and enhancing the crystalline properties of the deposited samples. With sufficient thickness and the number of layers, the peak and orientation become dominant compared to other peaks. On the other hand, the annealing process at the end of the deposition process also plays a crucial part in enhancing the crystallinity quality of the ZnO thin films.

Details of crystal properties for each of the peaks appearing in all samples are summarized in Table 3. The full width at half maximum (FWHM), crystallite size, d , dislocation line density, δ and interplanar spacing, d , values are tabulated based on the peaks obtained. Crystallite sizes were calculated using Scherer's formula, according to Equation 6.

$$d = \frac{0.94\lambda}{\beta \cos \theta} \tag{6}$$

d is the crystallite size for the samples, obtained according to the crystallographic peaks in XRD spectra, λ is the X-ray wavelength (1.542 Å), β is the diffraction line broadening/expanding at half of the maximum intensity (FWHM) of the peak, and θ is the Bragg's angle of diffraction in radian (obtained from 2θ). From the determination of d , dislocation line density could be calculated using Equation 7, and interplanar spacing, d , for all peaks was determined using Equation 8

$$\delta = \frac{1}{d^2} \tag{7}$$

$$d = \frac{n\lambda}{2 \sin \theta} \tag{8}$$

According to Equation 7, dislocation line density was obtained by dividing 1 with the value of crystallite size, calculated using Equation 6. d in Equation 7 referring to the value of calculated crystallite size. For interplanar spacing, d , $n=1$ is the order value of diffraction, while λ is the same value used in Equation 6, which is the X-ray wavelength (1.542 Å), and θ is the Bragg's angle of diffraction, which is the same value used for d calculation. An inversely proportional relationship between FWHM for all existence peaks with several layers could be observed in Table 3. Unlike FWHM, as described in Table 3, the crystallite size for ZnO thin films directly correlates with the number of layers. The value of crystallite size increase with the increment of the number of deposition layer. These findings are similar to the studies reported by Khan et al. (2017) and Kamalianfar et al. (2013). Both researchers also obtained the same finding, which the crystallite size for their ZnO thin films increased when they increased the number of layers throughout the deposition of the thin films. According to this, it is proven that the thickness had influenced the deposited samples' crystalline quality.

Table 3
Detail crystal properties for ZnO thin films deposited with 1, 2, 3, 4 and 5-layers

Sample	Orientation, hkl	2 θ (°)	FWHM	Crystallite size, D (nm)	Dislocation line density, δ ($\delta \times 10^{-3}$ line/nm ²)	Interplanar spacing, d (Åm)
1-SL	101	37.96	0.5284	27.96	1.279	2.3705
2-SL	101	37.96	0.4874	30.32	1.088	2.3705
	101	38.47	0.3654	40.44	0.612	2.3403
3-SL	102	44.68	0.3113	47.47	0.438	2.0284
	103	65.11	0.3504	42.17	0.562	1.4329
4-SL	101	38.48	0.3638	40.62	0.606	2.3397
	102	44.68	0.3087	47.86	0.437	2.0284
	103	65.11	0.3449	42.84	0.545	1.4328
5-SL	101	38.48	0.3552	41.60	0.578	2.3397
	102	44.68	0.3066	48.19	0.431	2.0284
	103	65.23	0.3059	48.31	0.428	1.4303

Kamalianfar et al. (2013) also stated that the thickness or thicker film would give better crystallinity quality because thicker film tends to have a smaller value of dislocation line density, δ . The results obtained in this study agree with this statement, where the dislocation line density and δ value decline with the increment of the crystallite size and number of layers. Decreasing the value of δ means that the film has less dislocation of atom arrangement, thus improving the crystalline properties of the samples produced. As for the interplanar spacing, d , the value decreases as the peak appears at a higher degree (2 θ). It is due to the increase of hkl , which is (101), (102), and (103). The increase of the hkl value produced a smaller value of interplanar spacing for each peak present (Bindu et al., 2014). Other than that, it could be observed that the interplanar spacing decrease with

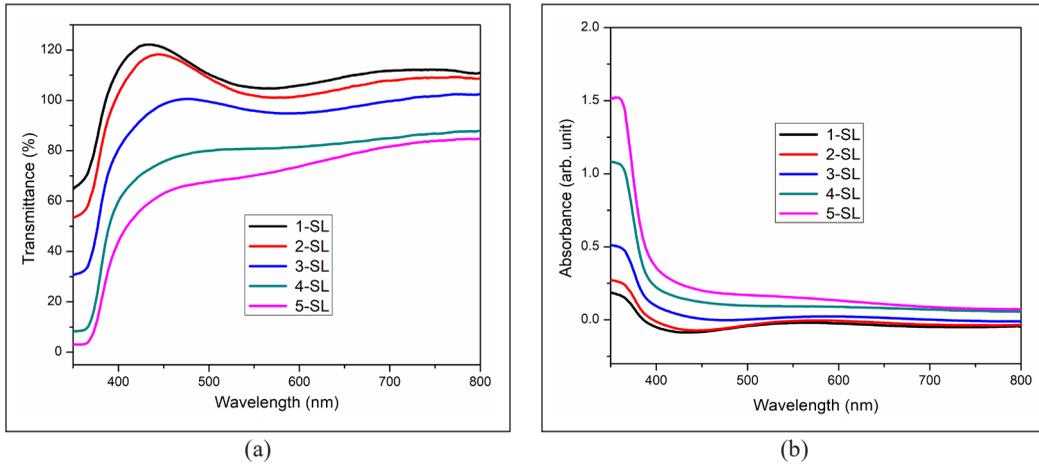


Figure 7. Optical properties (transmittance and absorbance) for 1-SL, 2-SL, 3-SL, 4-SL, and 5-SL ZnO thin films obtained from UV-Vis

the increment of deposition layers. This difference in d values suggests a different value of lattice parameters and atomic radii.

As the deposition layers or number of layers of ZnO thin films was increased, the thickness of the thin films was also increased, as stated earlier. Besides crystallinity, this thickness also affects the optical properties of the deposited ZnO thin films. As shown in Figure 7, the transmittance and absorbance for all samples were measured in the wavelength range between 350–800 nm. Generally, ZnO absorbs in the UV region of the electromagnetic spectrum (Shalu et al., 2020). Based on the transmittance spectra obtained, the transmittance values obtained decrease with the deposition layer of ZnO thin films. It could be related to the thickness of the thin films, which has a linear relationship with the number of layers of ZnO thin films. ZnO thin film with 1-Layer transmits the highest value of transmittance due to the film thickness, which has the lowest value, that allowed the incident light to penetrate easily. Due to this, the transparency of ZnO 1-SL gave the highest value as compared to the other samples. In contrast, ZnO thin film with five layers of deposition gave the lowest transmittance value, caused by the thicker film when further layers were deposited. This thicker layer prevents the incident light penetration for the film with higher thickness. By referring to the thickness value obtained from the surface profilometer, 5-Layer samples have the highest value, which means that there was more interference for incident light to penetrate; thus, the least incident light was transmitted and caused the reduction in transmittance value. This transmittance reduction has satisfied the Beer-Lambert law, given by Equation 9.

$$T = \exp(-\alpha d) \quad [9]$$

Where the intensity of transmitted light, T , decreases exponentially with the thickness of the sample, d , and absorption coefficient, α . As the incident light penetrates and shines through the sample, the total of the incident light reduces due to scattering from the films (Taha et al., 2015). This light scattering might cause the surface roughness and grain size of the films that could be increased from the thickness increment (Teh et al., 2017). The absorbance analysis exposed the light absorption behavior ranging between 350 and 800 nm. Figure 7(b) presents the absorption behavior of all ZnO thin films deposited with 1, 2, 3, 4, and 5 layers. Based on the absorption spectra, the absorbance of all ZnO thin films was proportionally increased with the film thickness. This behavior might be due to the enhancement of the crystallinity quality of the films, which was proved by the XRD analysis obtained. From the UV-Vis measurement, the optical energy band gap for deposited films was estimated and determined using the Tauc relation, given Equation 10.

$$(\alpha hv) = B(hv - E_g)^n \quad [10]$$

Where α is the absorption coefficient, hv is the photon energy, E_g is the optical bandgap, and B is an energy-independent constant (1×10^5 to $1 \times 10^6 \text{ cm}^{-1} \text{ eV}^{-1}$), which is dependent on the electron-hole mobility, and $n = \frac{1}{2}$ is used for directly allowed transitions. The calculated optical bandgap for all deposited samples was summarized in Table 4, and the extrapolating graphs are presented in Figures 8(a) to 8(e). The optical energy bandgap is deduced from the intercept of the x-axis at zero y-axes from the plot of αhv against hv through extrapolation of the linear part of the plot. By referring to the calculation results, the increment of optical energy bandgap value for ZnO thin films with a different number of layers could be observed. A slight increment of optical energy band gap was found for each layer that increased. This increasing trend can be related to the crystallinity quality of the thin films. The energy bandgap reduces with the quantum size effect, whereby the bigger the crystal size, the smaller the energy bandgap would obtain (Suzuki et al., 2005). This statement is supported by the crystal size calculated based on XRD spectra, in which the size decreased when a further layer was added. Thus, smaller crystallite size produced a higher value of E_g . Saravanan et al. (2015) also obtained the same trend of optical energy bandgap in their study. In their study, ZnO had been deposited at 20, 50, and 150 nm. They discovered that the optical energy bandgap for their ZnO films increased as the thickness increased, similar to the finding in this study.

Next is the investigation of the effect of the ZnO SL thickness on the ZNR growing

Table 4
The value of optical band gap for all ZnO thin films

Number of layers	Optical band gap value (eV)
1	3.15
2	3.19
3	3.21
4	3.24
5	3.25

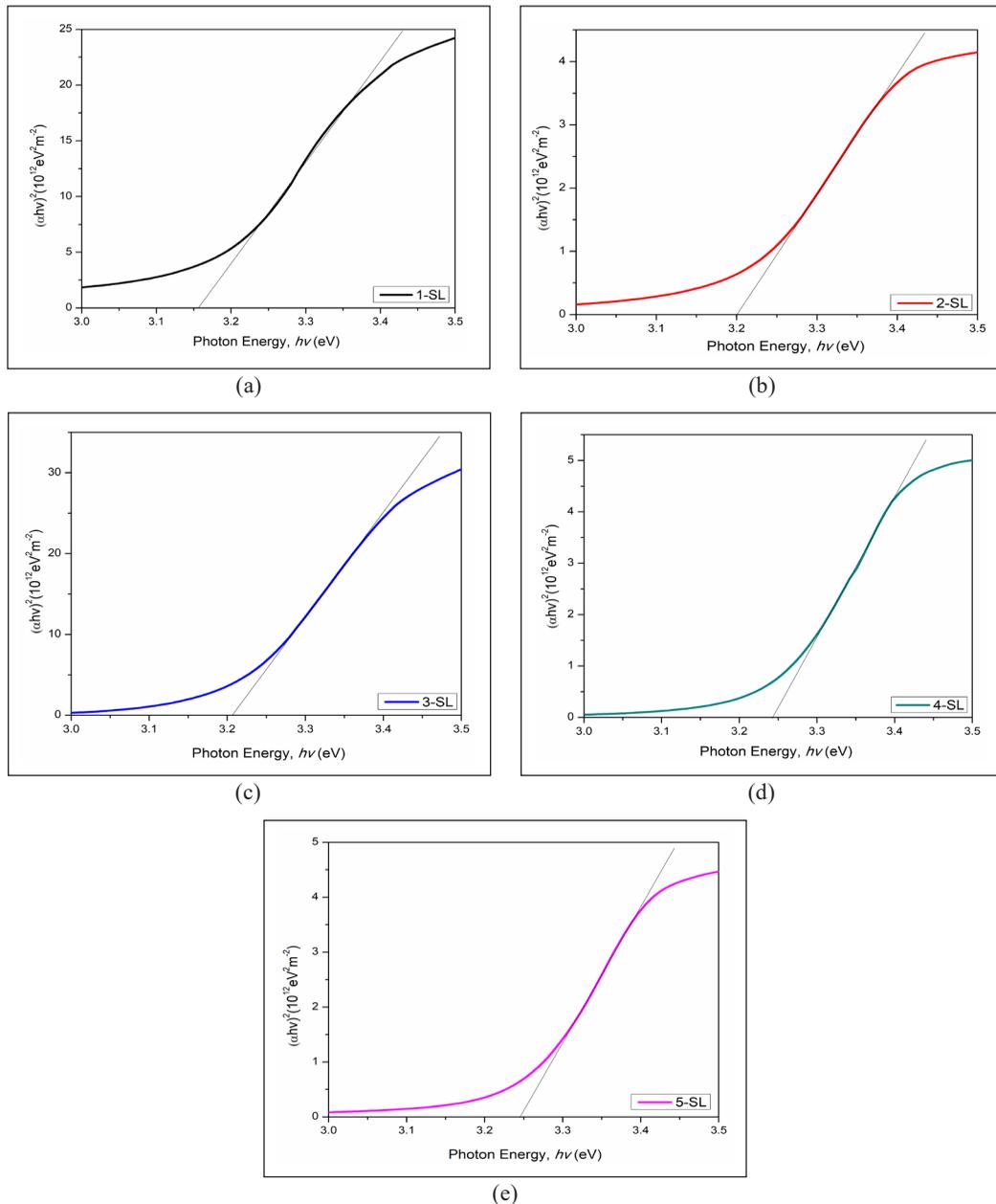


Figure 8. The TAUC plot for estimation of the optical energy band gap, E_g value for all samples (a) 1-SL, (b) 2-SL, (c) 3-SL, (d) 4-SL, and (e) 5-SL

process. As described in the methodology section, the growing process and conditions were constant for all samples. Figures 9(a) to 9(c) shows the morphology for ZNR grown on 1, 3, and 5-layers ZnO thin films observed with FESEM. From the FESEM images, the structures of ZNR improved as the number of layers increased. For 1-Layer SL, irregular

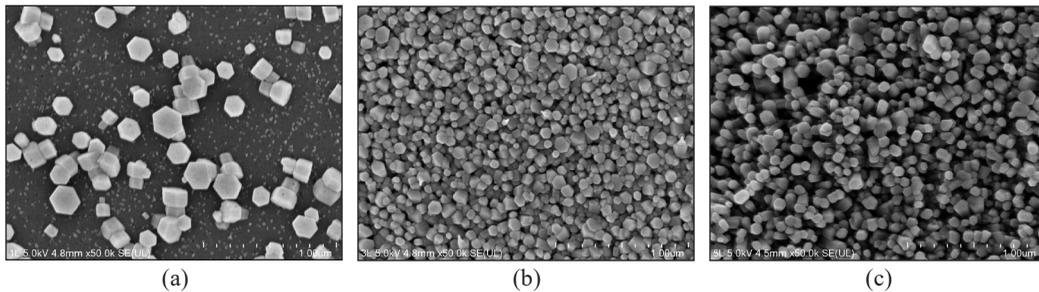


Figure 9. ZNR structures improved with the increasing number of layers, (a) 1-Layer ZnO SL, (b) 3-Layers SL, and (c) 5-Layers ZnO SL

nanorods were grown on the SL, as shown in Figure 9(a). Even though very sharp hexagonal nanorod structures were observed, the growth of the nanostructures was not uniform, and a large diameter could also observe. This uneven growth might be due to the particles of the ZnO, which were not evenly spread during the deposition process for the 1-Layer ZnO thin film. As explained earlier, the ZnO particles' distribution was improved when a further layer was deposited. The uniformity of the SL is important to produce aligned and dense ZNR because this SL will become the nuclei site for the ZNR to grow.

Significant changes in structures presented in Figures 9(b) and 9(c), 3, and 5 layers of ZnO SL were used to grow ZNR. Uniform and dense ZNR nanostructures grown on both 3 and 5-Layers SL. The enhancement of the growth of ZNR was attributed to the uniform ZnO SL produced when a further layer was added. For example, there might be an uneven distribution of ZnO particles during the first layer. Then another layer was added, and that area was covered; thus, the uniform surface of the nuclei site was produced. A uniform nuclei site would assist the growth of ZNR since the nanorods will grow based on the nuclei site. As a result, uniform and dense ZNR was produced. Similar phenomena were reported by Ikizler et al. (2014). They claimed that the surface seed of the parental seeds could serve as nuclei for the further growth of nanorods.

The surface topology and roughness of the ZNR samples were observed and compared. From the observation, the surface roughness for all samples has a proportional relationship with the thickness value, similar condition with the seed layer. Compared with the seed layer samples, ZNR exhibits a significant increment. The value of the surface roughness ranges from 6.53~70.82 nm. ZNR grown on 5-layers of seed layers gave the roughest surface, which is 70.82 nm. The gradually increasing trend of ZNR surface roughness might be due to the growing process has undergone several processes: nucleation, crystal, and grain growth. Since the ZNR was grown on the seed layer, the last sample (5-SL) produced the roughest surface due to the agglomeration, as shown from FESEM images. From this finding, the 5-SL sample could be a potential candidate for solar cell application due to the highest surface roughness among all prepared samples. The light-trapping process would be more effective on the rougher surface due to the different directions from the surface.

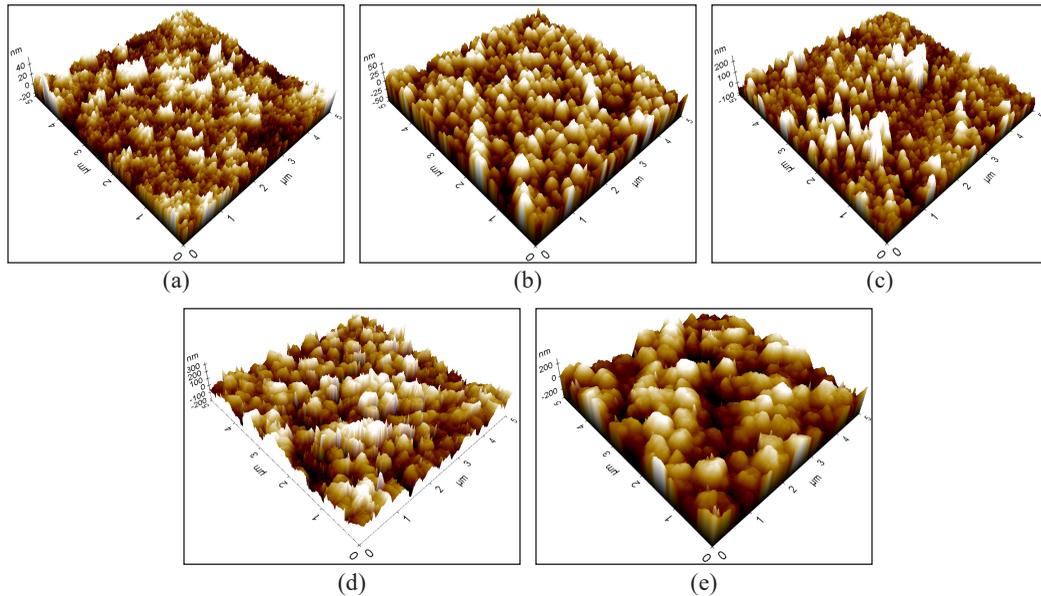


Figure 10. AFM images of the ZNR thin film with (a) 1-layer, (b) 2-layers, (c) 3-layers, (d) 4-layers, and (e) 5-layers seed layer

Scattered light beams will fall on the interfaces non-perpendicularly, which results in higher angular-dependent reflectance at the interfaces compared than in the case of the normal incidence (Lubomir et al., 2014).

The existences of the ZNR crystal structure grown on different numbers of SL were analyzed using XRD measurement, as shown in Figure 10. The variations of crystalline structure for all samples were evident in the XRD patterns, and three major diffraction peaks can be seen. These three diffraction peaks were identified as (100), (002), and (101), which indicates the good crystalline quality of ZNR films prepared. These corresponding peaks are commonly reported peaks for ZnO nanorods structures. Among these three peaks, the (002) peak represents the structural properties of ZnO in the c-direction, the nanorods' growth direction due to the plane's minimum energy. Therefore, the presence of the (002) peak in all samples indicates that the growth direction of ZNR is in perfect alignment (Hock et al., 2016).

It is also can be seen that the intensity of the (100), (002), and (101) plane increase as the thickness of the SL increase. The increasing intensity might be due to the SL's thickness, which affects crystalline quality. On the other hand, it could be related to the SL's thickness, which increases proportionally with the number of layers and improves the crystalline quality of ZnO SL. Due to this, the crystallinity of ZNR was also improved and led to higher peak intensity. Pokai et al. (2016) stated that the cumulative energy on the ZnO surface would increase when the ZnO SL is increased, thus will resulting in high crystallinity of ZnO SL. This finding also can be supported by the morphology obtained

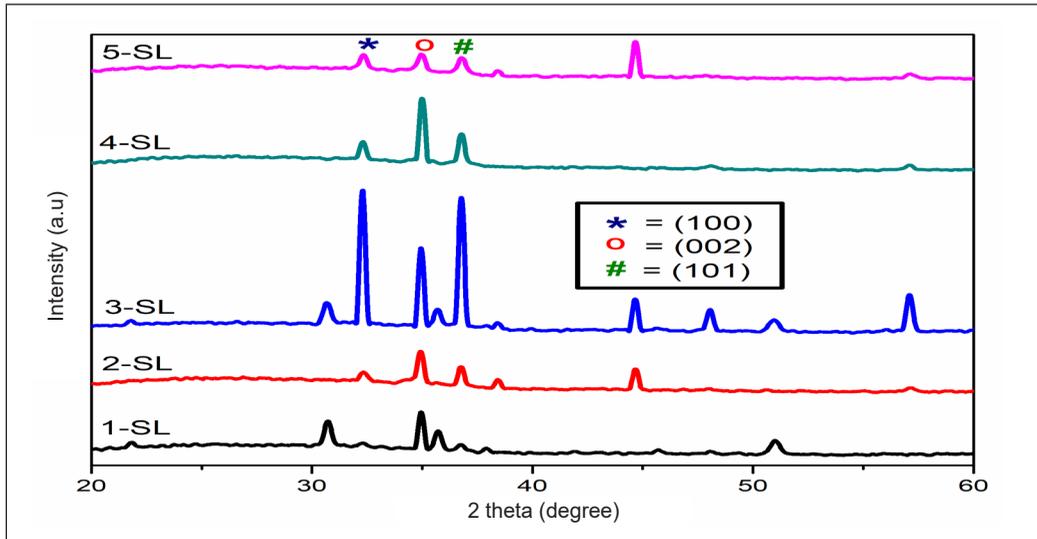


Figure 11. XRD patterns of ZNR grow on different numbers of SL (1, 2, 3, 4, and 5-layers)

by FESEM images. The distribution and growth of ZnO nanorods enhanced with the increase of SL number.

However, the increasing trend of peak intensity started to decrease when 4 and 5 layers were deposited for the SL. It means that ZNR grows at 3-Layers and has the highest intensity for (100), (002), and (101) planes. Notable decrement changes can be seen when the fourth layer was deposited and continue to decrease when the fifth layer was added. The occurrence may cause by the misalignment of the ZnO nanorods during the growth process. It could be explained by the seed/rod number ratio in forming a single rod. As correlated with the morphology of ZNR, misalignment could happen when there are multiple nanorods growing on a single seed. This growth process may lead to the misalignment of a rod, thus affecting the crystallinity of ZnO nanorods. This finding is supported by the study by Ikizler et al. (2014), which also obtained the same finding when the thickness of the SL increases.

From the XRD spectra, full width at half maximum (FWHM), the crystallite size (D), dislocation density (δ), and interplanar spacing (d) were calculated for (100), (002), and (101) planes, while lattice constant (C_{002}) and (a_{100}) were calculated by referring to (002) and (100) planes, respectively. Strain, ϵ_{zz} , and stress, σ were also calculated and tabulated in Table 5. D , δ , and d were calculated using Equations 6, 7, and 8, as explained earlier for ZnO thin film (seed layer). FWHM, D , δ , and d values for ZNR films decreased as layers increased. However, these values started to rise again for the 4 and 5-Layers samples. The differences in crystallite size may cause the broadening of XRD peaks (Madhavi et al., 2019). It could be proved from the XRD spectra in Figure 11 5-Layers sample has a broader peak compared with the 3 and 4-Layers samples. Other than that, there are reported studies that state that the variation in crystallite is due to the aggregation and

recrystallization during the crystal growth process, including heat structure (McGinty et al., 2020; Padmanabhan et al., 2020).

While for the lattice constant a and c , the values were obtained by Equations 11 and 12.

$$a = \frac{\lambda}{\sqrt{3} \sin \theta} \quad [11]$$

$$c = \frac{\lambda}{\sin \theta} \quad [12]$$

Here, λ is the constant value of wavelength for the X-ray radiation source (0.154 Å), and θ is the value obtained from 2θ in the spectra. In this study, both a and c values for all ZNR films are 2.918~3.202 for a and 5.112~5.127 for c , which approaches the theoretical value of a (3.249) and c (5.204). These values also prove that the ZNR films produced in this study have a hexagonal wurtzite structure. As for the strain, ϵ_{zz} , and stress, σ , the values were calculated based on Equations 13 and 14.

$$\epsilon_{zz} = \frac{c_{film} - c_{bulk}}{c_{bulk}} \times 100\% \quad [13]$$

$$\sigma = \frac{2C_{13}^2 - C_{33}(C_{11} + C_{12})}{2C_{13}} \epsilon_{zz} \quad [14]$$

For Equation 13, C_{film} signifies the lattice parameter of the ZnO nanorod films, and C_{bulk} is the unstrained lattice parameter for bulk ZnO (5.2066 Å). According to the Equation 14 which used to calculate the stress, C_{ij} is the elastic stiffness constant of bulk ZnO (C_{11} =208.8 GPa, C_{12} =119.7 GPa, C_{13} =104.2 GPa, and C_{33} =213.8 GPa) and ϵ_{zz} is the lattice strain obtained from Equation 13. The difference in film thickness had affected the crystal growth, including the formation of strain and stress in the lattice (Chason et al., 2018). The positive and negative strain values represent the tensile strain when the film is stretched and the compressive strain when compressed (Chason et al., 2018). In short, the positive value could be classified as a tensile strain; in contrast, it is considered a compressive strain. While for stress, if the value obtained is positive, it is considered tensile stress, and the negative value indicates the compressive stress [49–50]. Therefore, based on the calculation, the strain values for ZNR films in this study were obtained in a negative value, which is compressive strain. This phenomenon could be related to the film's crystallite size, decreasing as the thickness increases. Rezaie et al. (2018) reported that the increase in film thickness might be promoted by decreasing crystallite size because bigger crystallite size may induce lower-strained films. Moreover, the reduction of crystallite size and film strain enhancement was supported by the decreased d and c_{film} values obtained. In contrast with stress, the values were positive, representing the tensile stress.

The difference in strain and stress of ZNR films might be attributed to deficiencies in the crystal's lattice, such as stacking faults, grain boundaries, and dislocation density (Irvine et al., 2013). Usually, the grain boundary formation is believed to develop the strain and stress of the films. The interactions between the grains would lead to the tensile stress, while adding an adatom into the grain boundaries causes compressive stress (Yang et al., 2019). Meanwhile, the growth rate and atom surface mobility could be related to their strain and stress magnitudes (Magnfalt et al., 2015). Stresses might occur in the grown film that used the aqueous solution approach during the fabrication process, resulting in a lattice strain. Compressive tensile may occur during the growth process; thus, the strain and stress are determined by the growth parameters.

Table 5

2θ, FWHM, crystallite size, dislocation density, interplanar spacing, lattice constant, strain and stress of ZNR grow on different number of SL

Samples	Orientation, hkl	2θ (°)	FWHM	Crystallite size, D (nm)	Dislocation line density, δ ($\delta \times 10^{-3}$ line/nm ²)	Interplanar spacing, d (Ånm)	Lattice constant		Strain, ϵ_{zz} (%)	Stress, σ , (GPa)
							a ₁₀₀	c ₀₀₂		
1-SL	100	32.35	0.4497	32.86	0.926	2.7677	3.190	5.127	-1.5307	3.56
	002	34.97	0.3695	40.00	0.625	2.5668				
	101	36.83	0.4675	31.61	1.000	2.4404				
2-SL	100	32.35	0.4042	32.35	0.748	2.7674	2.918	5.126	-1.5480	3.60
	002	34.97	0.3413	34.97	0.534	2.5665				
	101	36.74	0.3348	36.74	0.513	2.4468				
3-SL	100	32.24	0.3001	32.25	0.414	2.7764	3.202	5.127	-1.5480	3.60
	002	34.97	0.1544	34.97	0.109	2.5665				
	101	36.74	0.3087	36.74	0.436	2.4468				
4-SL	100	32.36	0.3295	44.83	0.498	2.7670	3.191	5.112	-1.8131	4.22
	002	35.06	0.3062	48.26	0.429	2.5594				
	101	36.75	0.3499	42.22	0.561	2.4467				
5-SL	100	32.35	0.3745	39.43	0.643	2.7674	3.193	5.112	-1.8150	4.22
	002	35.06	0.4019	36.77	0.740	2.5594				
	101	36.83	0.4406	33.53	0.889	2.4403				

The variations in the crystalline quality and lattice properties of the ZNR films may contribute to the variation and changes in the optical properties of fabricated ZNR samples. Fig. 12 (a) and (b) represent all samples' transmittance and absorbance analyses, ranging between 350-800 nm, respectively. The transmittance of the ZnO nanorods grown shows a high percentage of transparency, in the range of 30~95% in the visible range. The range of the transparency value for ZNR films is slightly lower than for ZnO thin films. Fabricated films become thicker when ZNR is grown on the ZnO thin film, resulting in a

lower transparency value of ZNR films. As described in Figure 12(a), the transmittance of ZNR films shows a decrement trend with the increment of several layers due to the higher absorption capacity of thicker films. The least transparency value in the thicker film might be reasoning to the thick layer, which prevents the incident light from penetrating through the layer. Other than that, the percentage of transmittance would reduce in the UV region due to the onset of excitonic absorption (Ikizler & Peker, 2014).

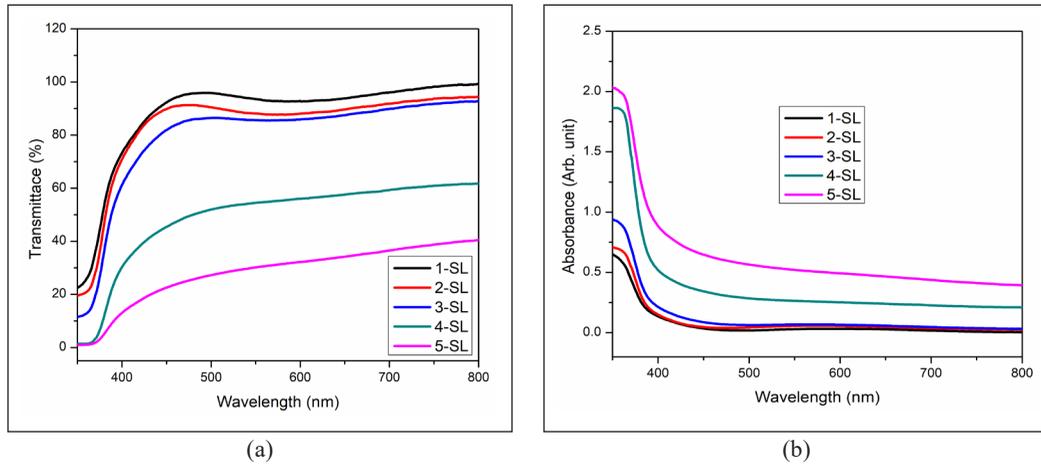


Figure 12. (a) Transmittance and (b) Absorbance spectra for ZNR films grown on different numbers of SL

As for the optical absorbance, the absorption properties of the ZNR films with different numbers of SL increase proportionally to the SL numbers in the UV region (below 400 nm) and lower in the visible region (400–800 nm). The result shows that ZnO nanorods grown on 5-Layers of SL show maximum UV absorbance value. This finding corresponds to the optical energy bandgap, calculated referring to eq. (10). The E_g values for all ZNR samples are presented in Table 6. The E_g values for all samples were estimated from the intercept of the x-axis, and the plot for all samples is shown in Figure 13. From Tauc’s plot, the E_g of ZNR films has a directly proportional relationship with the number of layers of SL. This increment occurs due to the film’s thickness that increases when the SL number increases. Therefore, the increasing number of layers would produce a thicker film, which influences the transmittance, absorbance, and optical energy band gap.

The increase in grain size and strain might cause the increment in band gap value (Ikizler & Peker, 2014). In addition, this grain size will increase if the thickness or chemical composition increases. These two

Table 6
Optical energy bandgap for ZnO nanorods sample grown on a various number of SL

Sample	Optical band gap value (eV)
1-SL	3.23
2-SL	3.27
3-SL	3.28
4-SL	3.29
5-SL	3.67

factors will directly change the material (thin film) band gap (El Zawawi et al., 2017). Therefore, even though the same material was used, the optical energy band gap would change significantly with thickness and chemical composition changes. Other than that, crystallinity, temperature, particle size, and a few other factors will also affect the value of the energy band gap of thin film (Mahato et al., 2017).

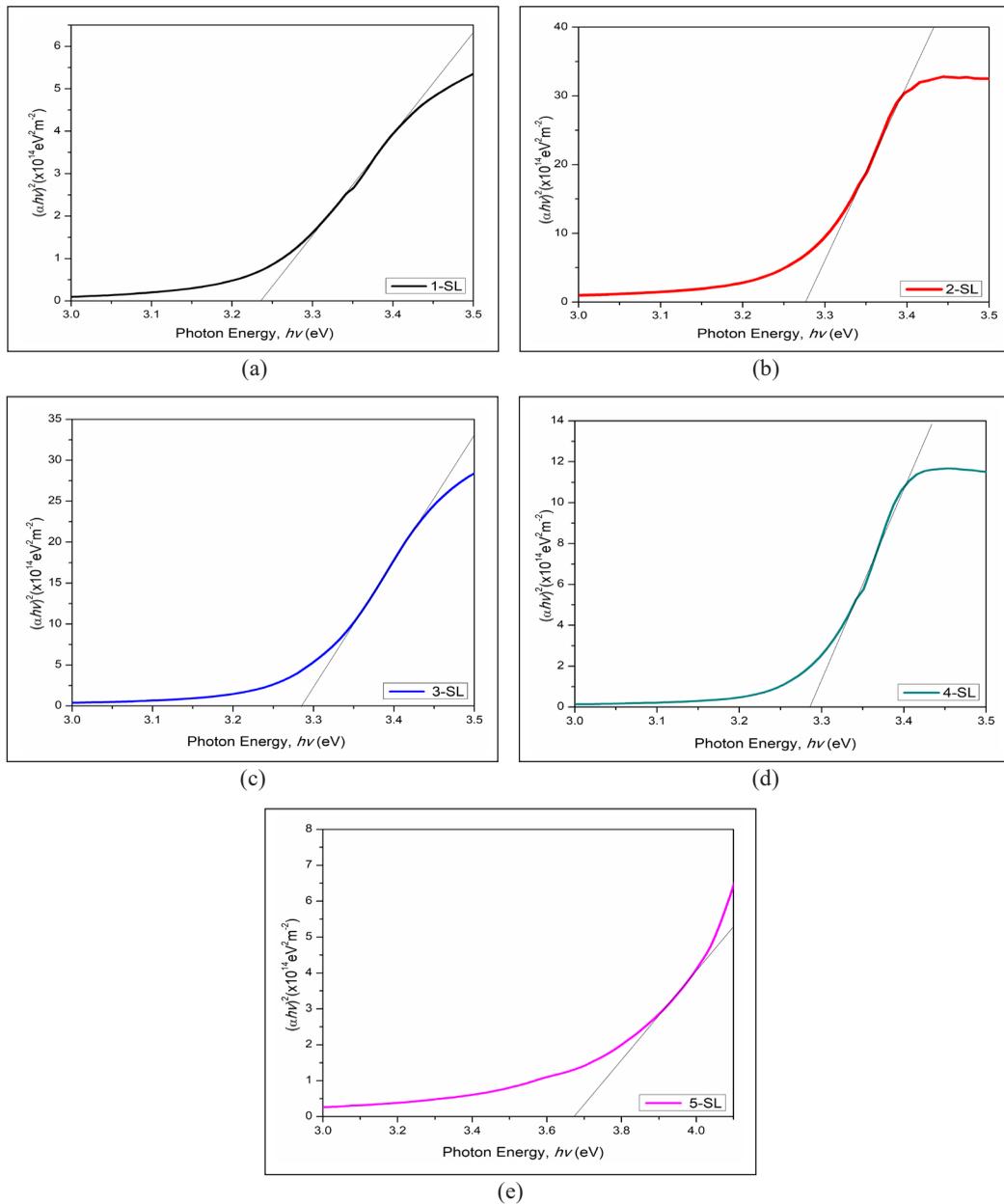


Figure 13. Estimation of E_g for ZNR grown on (a) 1-SL, (b) 2-SL, (c) 3-SL, (d) 4-SL, and (e) 5-SL of a seed layer

CONCLUSION

In conclusion, multilayer ZnO thin film and well-aligned ZNR films were successfully synthesized using the sol-gel spin coating technique and chemical bath deposition methods. From the results obtained, it was found that the thickness of the films could influence the characteristics and properties of the fabricated ZnO films. Increasing the thickness of ZnO thin film that functions as the SL for ZNR growth influenced the structure and orientation of ZNR. For consideration of morphology ZNR, the images obtained revealed that 5-layers ZnO SL produced well-aligned and dense-packed ZNR. It was evident that the crystalline quality and optical properties of ZnO and ZNR films were also improved by the increment of the thickness of the SL. XRD analysis shows that all ZNR samples exhibit (100), (002), and (101) peaks, which prove that the films have good crystalline quality. Also, AFM analysis revealed that higher thickness of the seed layer produced rougher surface ZNR, which is one of the characteristics of the good potential for a solar cell.

ACKNOWLEDGEMENT

The authors want to thank all members of Nano Electronic Center (NET), Integrated Sensors Research Group (DERIA), and Nano-Science Technology Center (NST, Universiti Teknologi MARA (UiTM) Shah Alam for all the research facilities. In addition, the Ministry of Higher Education Malaysia partially supports this work under the Fundamental Research Grant Scheme (FRGS/1/2017/TK04/UITM/02/16).

REFERENCES

- Abdel-Galil, A., Hussien, M. S., & Yahia, I. S. (2021). Synthesis & optical analysis of nanostructures F-doped ZnO thin films by spray pyrolysis: Transport electrode for photocatalytic applications. *Optical Materials*, 114, Article 110894. <https://doi.org/10.1016/j.optmat.2021.110894>
- Abdulrahman, A. F., Ahmed, S. M., Ahmed, N. M., & Almessiere, M. A. (2020). Enhancement of ZnO nanorods properties using modified chemical bath deposition method: Effect of precursor concentration. *Crystal*, 10(5), Article 386. <https://doi.org/10.3390/cryst10050386>
- Addamo, M., Augigliaro, V., Di Paola, A., Garcia-Lopez, E., Loddo, V., Marci, G., & Palmisano, L. (2008). Photocatalytic thin films of TiO₂ formed by sol-gel process using titanium tetraisopropoxide as the precursor. *Thin Solid Films*, 516(12), 3802-3807. <https://doi.org/10.1016/j.tsf.2007.06.139>
- Al Farsi, B., Souier, T. M., Al Marzouqi, F., Al Maashani, M., Bououdina, M., Widatallah, H. M., & Al Abri, M. (2021). Structural and optical properties of visible active photocatalytic Al doped ZnO nanostructured thin films prepared by dip coating. *Optical Materials*, 113, Article 110868. <https://doi.org/10.1016/j.optmat.2021.110868>
- Alenezi, M. R. (2018). Hierarchical zinc oxide nanorings with superior sensing properties. *Materials Science and Engineering: B*, 236(237), 132-138. <https://doi.org/10.1016/j.mseb.2018.11.011>

- Banari, M., Memarian, N., & Vomiero, A. (2021). Effect of the seed layer on the photodetection properties of ZnO nanorods. *Materials Science and Engineering: B*, 272, Article 115332. <https://doi.org/10.1016/j.mseb.2021.115332>
- Bindu, P., & Thomas, S. (2014). Estimation of lattice strain in ZnO nanoparticles: X-ray peak profile analysis. *Journal of Theory Applied Physics*, 8, 123-134. <https://doi.org/10.1007/s40094-014-0141-9>
- Chason, E., Keckes, J., Sebastian, M., Thompson, G. B., Barthel, E., Doll, G. L., Murray, C. E., Stoessel, C. H., & Martinu, L. (2018). Review articles: Stress in thin films and coatings: Current status, challenges, and prospects. *Journal of Vacuum Science and Technology A*, 36(2), 1-49. <https://doi.org/10.1116/1.5011790>
- Chen, X. X., Chen, L., Li, G., Cai, L. X., Miao, G., Guo, Z., & Meng, F. L. (2021). Selectively enhanced gas-sensing performance to n-butanol based on uniform CdO-decorated porous ZnO nanobelts. *Sensors and Actuators B: Chemical*, 334, Article 129667. <https://doi.org/10.1016/j.snb.2021.129667>
- Daniel, L., Falko, S., & Dietrich, R. Z. (2014). Thin films with high surface roughness: Thickness and dielectric function analysis using spectroscopic ellipsometry. *Methodology*, 3(82), 1-8. <https://doi.org/10.1186/2193-1801-3-82>
- Djurisic, A. B., Ng, A. M. C., & Chen, X. Y. (2010). ZnO nanostructures for optoelectronics: Material properties and device applications. *Progress in Quantum Electronics*, 34(4), 191-259. <https://doi.org/10.1016/j.pquantelec.2010.04.001>
- El Zawawi, I. K., Mahdv, M. A., & El-Sayad, E. A. (2017). Influence of film thickness and heat treatment on the physical properties of Mn doped Sb₂Se₃ nanocrystalline thin films. *Journal of Nanomaterials*, 2017, Article 7509098. <https://doi.org/10.1155/2017/7509098>
- Gonçalves, R. S., Barrozo, P., Brito, G. L., Viana, B. C., & Cunha, F. (2017). The effect thickness on optical, structural, and growth mechanism of ZnO thin film prepared by magnetron sputtering. *Thin Solid Films*, 661, 40-45. <https://doi.org/10.1016/j.tsf.2018.07.008>
- Gunes, S., Neugebauer, H., & Sariciftci, N. S. (2007). Conjugated polymer-based organic solar cells. *Chemical Reviews*, 107(4), 1324-1338. <https://doi.org/10.1021/cr050149z>
- Hajezi, S. R., Hosseini, H. M., & Ghamsari, M. S. (2008). The role of reactants and droplet interfaces on nucleation and growth of ZnO nanorods synthesized by vapor-liquid-solid (VLS) mechanism. *Journal of Alloys and Compounds*, 455(1-2), 353-357. <https://doi.org/10.1016/j.jallcom.2007.01.100>
- Hasabeldaim, E. H. H., Ntwaeborwa, O. M., Kroon, R. E., Coetsee, E., & Swart, H. C. (2020). Luminescence properties of Eu doped ZnO PLD thin films: The effect of oxygen partial pressure. *Superlattices and Microstructures*, 139, Article 106432. <https://doi.org/10.1016/j.spmi.2020.106432>
- Hock, B. L., Riski, T. G., Sin, T. T., Chun, H. T., Alshanableh, A., Oleiwi, H. F., Chi, C. Y., Hj Jumali, M. H., & Muhammad Yahaya. (2016). Controlled defect fluorine-incorporated ZnO nanorods for photovoltaic enhancement. *Scientific Reports*, 6, Article 32645. <https://doi.org/10.1038/srep32645>
- Huey, J. T., Zainal, Z., Talib, Z. A., Hong, N. L., Shafie, S., Sin, T. T., Kar, B. T., & Bahrudin, N. N. (2021). Synthesis of high quality hydrothermally grown ZnO nanorods for photoelectrochemical cell electrode. *Ceramics International*, 47(10, Part A), 14194-14207. <https://doi.org/10.1016/j.ceramint.2021.02.005>

- Ikizler, B., & Peker, S. M. (2014). Effect of the seed layer thickness on the stability of ZnO nanorod arrays. *Thin Solid Film*, 558, 149-159. <https://doi.org/10.1016/j.tsf.2014.03.019>
- Irvine, W. T. M., Hollingsworth, A. D., Grier, D. G., & Chaikin, P. M. (2013). Dislocation reactions, grain boundaries, and irreversibility in two-dimensional lattices using topological tweezers. *Applied Physical Sciences*, 110(39), 15544-15548. <https://doi.org/10.1073/pnas.1300787110>
- Jimenez-Cadena, G., Comini, E., Ferroni, M., Vomiero, A., & Sberveglieri, G. (2010). Synthesis of different ZnO nanostructures by modified PVD process and potential use for dye-sensitized solar cells. *Materials Chemistry and Physics*, 124(1), 694-698. <https://doi.org/10.1016/j.matchemphys.2010.07.035>
- Kaiyong, C., Michael, M., Korg, B., Annett, R., & Klaus, D. J. (2005) Surface structure and composition of flat titanium thin films as a function of film thickness and evaporation rate. *Applied Surface Science*, 250(2005), 252-267. <https://doi.org/10.1016/j.apsusc.2005.01.013>.
- Kamalianfar, A., Halim, S. A., Behzad, K., Naseri, M. G., Navasery, M., Din, F. U., Zahedi, J. A. M., Lim, K. P., Chen, S. K., & Sidek, H. A. A. (2013) Effect of thickness on structural, optical, and magnetic properties of Co doped ZnO thin film by pulsed laser deposition. *Journal of Optoelectronics and Advanced Materials*, 15(3), 239- 243.
- Kannan, S., Subiramaniyam, N. P., & Lavanisadevi, S. U. Controllable synthesis of ZnO nanorods at different temperatures for enhancement of dye-sensitized solar cell performance. *Material Letters*, 274, Article 127994. <https://doi.org/10.1016/j.matlet.2020.127994>
- Khan, M. I., Bhatti, K. A., Alonizan, N., & Althobaiti, H. S. (2017). Characterization of multilayer ZnO thin films deposited by sol-gel spin coating technique. *Results in Physics*, 7, 651-655. <https://doi.org/10.1016/j.rinp.2016.12.029>
- Khan, Z. R., Abdullah S. Alshammari., Bouzidi, M., Shkir, M., & Shukla, D. K. (2021). Improved optoelectronic performance of sol-gel derived ZnO nanostructured thin films. *Inorganic Chemistry Communications*, 132, Article 108812. <https://doi.org/10.1016/j.inoche.2021.108812>
- Khranovskyy, V., Yakimova, R., Karlsson, F., Abdul, S. S., Holtz, P., Urgessa, Z. N., Oluwafemi, O. S., & Botha, J. R. (2012). Comparative PL study of individual ZnO nanorods, grown by APMOCVD and CBD technique. *Physica B: Condensed Matter*, 407(10), 1538-1442. <https://doi.org/10.1016/j.physb.2011.09.080>
- Kumar, S., Share, P. D., & Kumar, S. (2018). Optimization of CVD parameters for ZnO nanorods growth: Its photoluminescence and field emission properties. *Materials Research Bulletin*, 105, 237-245. <https://doi.org/10.1016/j.materresbull.2018.05.002>
- Kumar, V., Singh, N., Mehra, R. M., Kapoor, A., Purohit, L. P., & Swart, H. C. (2013). Role of film thickness on the properties of ZnO thin films grown by sol-gel method. *Thin Solid Films*, 539. <https://doi.org/10.1016/j.tsf.2013.05.088>
- Lokesh, K. S., Kumar, J. R. N., Kannantha, V., Pinto, T., & Sampreeth, U. (2020). Experimental evaluation of substrate and annealing conditions on ZnO thin films prepared by sol-gel method. *Materialstoday: Proceedings*, 24(2), 201-208. <https://doi.org/10.1016/j.matpr.2020.04.268>
- Lubomir, S., Libor, L., & Jarmila, M. (2014). Influence of surface roughness on optical characteristics of multilayer solar cells. *Applied Physics*, 12(6), 631-64. 10.15598/aece.v12i6.1078

- Madhavi, J. (2019). Comparison of average crystallite size by X-ray peak broadening and Williamson-Hall and size-strain plots for VO²⁺ doped ZnS/CdS composite nanopowder. *SN Applied. Science*, 1, Article 1509. <https://doi.org/10.1007/s42452-019-1291-9>
- Magnfalt, D., Fillon, A., Boyd, R. D., Helmersson, U., Sarakinos, K., & Abadias, G. (2015). Compressive intrinsic stress originates in the grain boundaries of dense refractory polycrystalline thin films. *Journal of Applied Physics*, 119(5), Article 055305. <https://doi.org/10.1063/1.4941271>
- Mahato, S., & Kar, A. K. (2017). The effect of annealing on structural, optical, and photosensitive properties of electrodeposited cadmium selenide thin films. *Journal of Science: Advanced Materials and Devices*, 2(2), 165-171. <https://doi.org/10.1016/j.jsamd.2017.04.001>
- McGinty, J., Yazdanpanah, N., Price, C., Joop, H. T., & Sefcik, J. (2020). Nucleation and crystal growth in continuous crystallization. In N. Yazdanpanah & Z. K. Nagy (Eds.), *The Handbook of Continuous Crystallization* (pp. 1-50). Royal Society of Chemistry. <https://doi.org/10.1039/9781788013581-00001>
- Mohammadzadeh, A., Azadbeh, M., Shokriyan, B., & Abad, S. N. K. (2020). Synthesis of ZnO nanocombs and tetrapods by catalyst-free oxidation of alpha brass powders in air atmosphere. *Ceramics International*, 46(2), 2552-2557. <https://doi.org/10.1016/j.ceramint.2019.09.112>
- Mosalagae, K., Murape, D. M., & Lepodise, L. M. (2020). Effects of growth conditions on properties of CBD synthesized ZnO nanorods grown on ultrasonic spray pyrolysis deposited ZnO seed layers. *Heliyon*, 6(7), 1-10. <https://doi.org/10.1016/j.heliyon.2020.e04458>
- Padmanabhan, S. C., Collins, T. W., Pillai, S. C., McCormack, D. E., Kelly, J. M., Holmes, J. D., & Morris, M. A. (2020). A conceptual change in crystallisation mechanisms of oxide materials from solutions in closed systems. *Scientific Reports*, 10, Article 18414. <https://doi.org/10.1038/s41598-020-75241-z>
- Pokai, S., Lomnonthakul, P., Horprathum, M., Kalasung, S., Eiamchai, P., Limwichean, S., Nuntawong, N., Pattantsetakul, V., Tuscharoen, S., & Kaewkhao, J. (2016). Influence of growth conditions on morphology of ZnO nanorods by low-temperature hydrothermal method. *Key Engineering Materials*, 675-676, 53-56. <https://doi.org/10.4028/www.scientific.net/kem.675-676.53>
- Regmi, G., & Velumani, S. (2021). Impact of target power on the properties of sputtered intrinsic zinc oxide (i-ZnO) thin films and its thickness dependence performance on CISE solar cells. *Optical Materials*, 119, Article 111350. <https://doi.org/10.1016/j.optmat.2021.111350>
- Rezaie, M. N., Manavizadeh, N., Nayeri, F. D., Bidgoli, M. M., Nadimi, E., & Boroumand, F. A. (2018). Effect of seed layers on low-temperature, chemical bath deposited ZnO nanorods-based near UV-OLED performance. *Ceramics International*, 44(5), 4937-4945. <https://doi.org/10.1016/j.ceramint.2017.12.086>
- Rodriguez-Martinez, Y., Alba-Cabarnas, J., Cruzata, O., Bianco, S., Tresso, E., Rossi, F., & Vaillant-Roca, L. (2020). In-situ pulsed laser induced growth of CdS nanoparticles on ZnO nanorods surfaces. *Material Research Bulletin*, 125, Article 110790. <https://doi.org/10.1016/j.materresbull.2020.110790>
- Roy, S., Banerjee, N., Sarkar, C. K., & Bhattacharyya, P. (2013). Development of an ethanol sensor based grown ZnO nanorods. *Solid-State Electronics*, 87, 43-50. <https://doi.org/10.1016/j.sse.2013.05.003>
- Rwenyagila, E. R., Ayei-Tuffour, B., Kana, M. G. Z., Akin-Ojo, O., & Soboyejo, W. O. (2014). Optical properties of ZnO/Al/ZnO multilayer films for large area transparent electrodes. *Journal of Material Research*, 29, 2912-2920. <https://doi.org/10.1557/jmr.2014.298>

- Saravanan, K., Krishnan, R., Hsieh, S. H., Wang, H. T., Wang, Y. F., Pong, W. F., Asokan, K., Avasthi, D. K., & Kanjilal, D. (2015). Effect of defects and film thickness on the optical properties of ZnO-Au hybrid films. *Royal Society of Chemistry Advances*, 51(5), 40813-40820. <https://doi.org/10.1039/c5ra02144h>
- Scholtz, L., Ladanyi, L., & Mullerova, J. (2014). Influence of surface roughness on optical characteristics of multilayers solar cells. *Applied Physics*, 12(6), 631-640. <https://doi.org/10.15598/aeee.v12i6.1078>
- Shalu, G., Shukla, M., Tiwari, A., Agrawal, J., Bilgaiyan, A., & Singh, V. (2020). Role of solvent used to cast P3HT thin films on the performance of ZnO/P3HT hybrid photo detector. *Physica E: Low-dimensional Systems and Nanostructures*, 115, Article 113694. <https://doi.org/10.1016/j.physe.2019.113694>
- Shariffudin, S. S., Salina, M., Herman, S. H., & Rusop, M. (2012). Effect of film thickness on structural, electrical, and optical properties of sol-gel deposited layer-by-layer ZnO nanoparticles. *Transaction on Electrical and Electronic Materials*, 13(2), 102-105. <https://doi.org/10.4313/TEEM.2012.13.2.102>
- Sharma, S., Vyas, S., Periasamy, C., & Chakrabarti, P. (2014). Structural and optical characterization of ZnO thin films for optoelectronic device applications by RF sputtering technique. *Superlattices and Microstructures*, 75, 378-389. <https://doi.org/10.1016/j.spmi.2014.07.032>
- Suzuki, K., & Kijima, K. (2005). Optical band gap of barium titanate nanoparticles prepared by RF-plasma chemical vapor deposition. *Japanese Journal of Applied Physics*, 44(4R), 2081-2082. <https://doi.org/10.1143/JJAP.44.2081>
- Taha, K. K., M'hamed, M. O., & Idris, H. (2015). Mechanical fabrication and characterization of zinc oxide (ZnO) nanoparticles. *Journal of Ovonic Research*, 11(6), 271-276.
- Teh, Y. C., Ala'eddin, A. S., Jamal, Z. A. Z., & Poopalan, P. (2017). Correlation of film thickness to optical band gap of sol-gel derived Ba_{0.9}Gd_{0.1}TiO₃ thin films for optoelectronic applications. *EPJ Web of Conferences*, 162, Article 01042. <https://doi.org/10.1051/epjconf/201716201042>
- Yang, G., & Park, S. J. (2019). Deformation of single crystal, polycrystalline materials, and thin films: A review. *Materials*, 12(12), 1-18. <https://doi.org/10.3390/ma12122003>
- Zhang, Y., Ram, M. K., Stefanakos, E. K., & Goswami, D. Y. (2012). Synthesis Characterization, and application of ZnO nanowires. *Nanofiber Manufacture, Properties and Application*, 2012, Article 624520. <https://doi.org/10.1155/2012/624520>
- Zhou, L., Zeng, W., & Li, Y. (2019). A facile one-step hydrothermal synthesis of a novel NiO/ZnO nanorod composite and its enhanced ethanol sensing property. *Material Letters*, 254, 92-95. <https://doi.org/10.1016/j.matlet.2019.07.042>

Review Article

A Review on 3D Augmented Reality Design Technique and Inward Leakage Testing on Protective Face Mask

Nur Amirah Kamaluddin¹, Murizah Kassim^{1,2*} and Shahrani Shahbudin¹

¹*School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

²*Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

ABSTRACT

Protective face mask identification is essential today to users as it is a prominent protective wearable to shield from being infected by Covid-19 viruses. Protective face masks consist of layers of fibers that can capture large respiratory droplets and microscopic particles such as viruses or dust. Thus, mask filtration efficiency results depend on the materials used for each layer. Detail about mask description and efficiency are still anonymous to users, which is vital in this COVID-19. Therefore, this paper reviews designing 3D augmented reality for the protective mask with its detail parameter and mask sizing recommendation on android mobile. About 73 articles on the protective face mask, 3D augmented reality modeling, masks inward leakage testing, breathing resistance, and measuring faces have been reviewed. The result examines the existing protective face mask, inward leakage testing parameter, breathing resistance parameters, 3D modeling techniques, mobile applications, and the application used for measuring faces. The identified result shows six recent and

familiar masks with 8% of arithmetic mean for inward leakage testing. The best flow efficiency is determined a 0.3 Microns bigger than 95%. The result also shows a detailed parameter for inward leakage testing in terms of inhalation resistance and flow rate. The comparison for 3D AR parameters is identified for application type, evaluated parameter, technical support parameter, AR platform, and software. This research is significant for developing AR

ARTICLE INFO

Article history:

Received: 08 October 2021

Accepted: 25 January 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.19>

E-mail addresses:

nuramirahkamaluddin97@gmail.com (Nur Amirah Kamaluddin)

murizah@uitm.edu.my (Murizah Kassim)

shahrani@uitm.edu.my (Shahrani Shahbudin)

* Corresponding author

mobile applications that ease and transparency information to the community for safety and health issues in Malaysia.

Keywords: 3D augmented reality, inward leakage testing, marker detection, mask material, mobile apps, protective face mask

INTRODUCTION

The recent COVID-19 pandemic has made protective masks crucial for all. COVID-19 could infect people through respiratory droplets produced when the infected individual sneezes or coughs loudly causing protective face masks to come in several types, layers, and materials (WHO, 2020). They are all designed to protect against airborne pollutants ranging from pollen to chemical fumes to infections. The capacity of the filter, and the amount of protection against pollutants and viruses, are determined by the materials and the technical design of the protective face mask (Long et al., 2020; Steinle et al., 2018). Figure 1 shows the various size of airborne contamination and pathogens. Therefore, the higher the pore size material design on the protective mask, the lower the filtration efficiency, and these studies show how important it is to design a protective face mask based on the material (Chua et al., 2020).

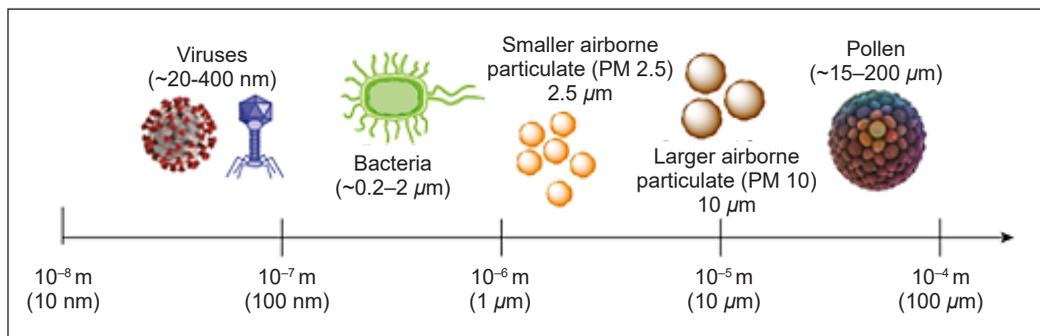


Figure 1. The size chart of common airborne contamination (Chua et al., 2020)

The filtration efficiencies of different materials for the mask are essential to be determined as different materials such as non-woven polypropylene were not threaded like cloth, but by spinning polypropylene to become a thread and to lay them in the form of a web to help provide the breathable condition, filtration, and water-resistant (Balamurali et al., 2021; Hao et al., 2020). Therefore, these criteria for face mask material are crucial to note as they would affect the filtering efficiency, breathable resistance, and the weave pattern of the material to capture the airborne particle. The result shows that the arrangement of the thread producing the front layer of the protective face mask is random. Thus, it has an electrostatic charge that can attract and capture particles of all sizes, showing that almost 95% of medium or small particles can filter. Figure 2 shows the layer of the N95 mask as

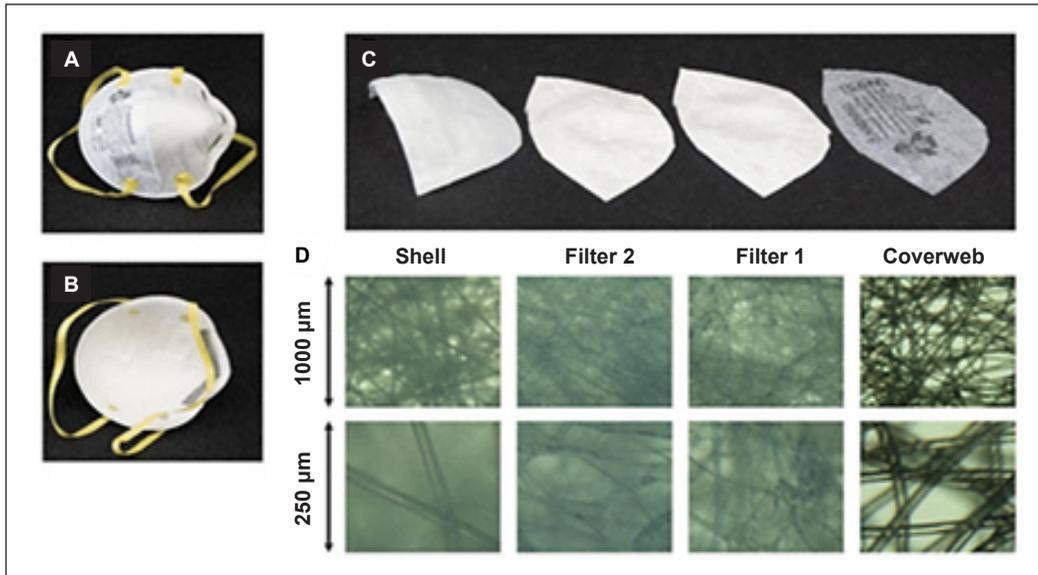


Figure 2. (A) outside layer mask, (B) the inside layer of the mask, (C) the inside layer is called a shell; the center layers are for Filter 1 and Filter 2, while the outside layer is called a cover web, and (D) the microscope images for the four layers showing the thread pattern (Huber et al., 2021).

an example of the space area between the thread, and this study is important in proving which material could give better protection.

Due to high demand during the early stage of covid-19, people buy protective face masks online as they are easier to find and sometimes cheaper-causing some sellers to sell counterfeit protective face masks that are not standard to NIOSH approval. Only by clarifying and understanding the important parameter while choosing a protective face mask will the user be able to understand the essential information shared using 3D augmented reality application. The examination of the inward leakage testing of how the test by doing several exercises would be done on a group of people while wearing different face masks and how the result on breathing resistance, face seal leakage, and filtering percentage will affect.

On the other hand, modeling and clarifying which design technique is used to design a 3D image of the protective face mask is important. Several studies show using a mobile application with AR technology that contains various medical training (Hossain et al., 2021; Leung et al., 2019; Sik-Lanyi, 2017). The app is connected to the database and runs simultaneously on the trainees' and trainers' mobile devices. This setting allows the trainer to watch the trainee's progress and switch the training scenario to judge the trainee's responses during training (Barratt et al., 2020). AR could be a technology that mixes real-world images and videos with computer-generated information and image processing, and students will have a clear view of learning and using the technique even though the virtual (Kassim & Bakar, 2021; Kassim & Zubir, 2019).

Other than that, every human in this world has different size faces causing some protective face masks that may not fit the faces. Based on the Centre for Control Disease (CDC), a suitable protective face mask is to choose a fitted face mask to prevent any airborne particle from passing through the gap. Thus, it is essential to studies on which software to use in measuring faces to recommend the size of protective face mask suitable for users.

This research reviews the AR modeling technique for protective masks and identifies the masks' inward leakage testing parameters to design a mobile AR for fitted faces for masks. This research helps healthcare users and society identify good protective masks for daily use. This research has reviewed 73 papers from index proceedings and journals to place the mask's details information. The 3D AR modeling will focus on the 3D images of the protective face mask and its data structure parameters to develop the mobile application. Comparing each activity done in the inward leakage testing will help users know which protective face mask to wear. In addition, there will be research gaps for the AR protective face mask and mobile applications in AR and VR. This information will help determine if there would be value in engaging augmented reality in sharing the information regarding the protective mask.

Protective Face Mask for Respirator

A respirator is a face mask that will protect the wearer from inhaling harmful air (Johnson, 2016). The United States Occupational Safety and Health Administration (OSHA) states that they should give respirators to any workers anytime they are exposed to danger. Therefore, choosing a suitable and adequate protective mask is essential. Generally, protective masks come in many styles and materials, and not all mask materials filter similarly (Hamid et al., 2016). In addition, each respirator needs to remove impurities like the previously stated particle settling. The existing research has shown diverse data regarding respirators, filtration performance, and test agents (Ghosh et al., 2020).

Nevertheless, different production has different versions. Therefore, collecting the data and inward leakage tests still needed to be researched. The challenges issues are modeling a 3D image of a protective mask and integrating it into a 3D augmented reality mobile application when the user scans the marker. The transition from the database stored to the face tracking system tracks the user's facial shape to determine the fitted protective mask. Total inward leakage is a sum of contaminated air leaking through a respirator from various sources, including the face seal, valves, and gaskets, as well as filtration (Baugh, 2015; Zhu et al., 2020).

For analysis, tests on people selected to represent the target group include gender, general facial traits, and facial measurements. Many people are wearing masks improperly, not fitting to the face or tying them properly, which impacts the seal to the face (Steinle

et al., 2018). Surgical and fabric masks showed improper wearing of face masks as the material used to produce shows different filtering percentages (Cherrie et al., 2018). The fit of sealed and loose-fitting surgical masks on human beings of different races and N95 filtering face-piece respirators has been researched (Karuppasamy & Obuchowski, 2021). Filtration efficiency for a protective face mask shows that the mask's design, materials used, and it is fit to the face are vital to point to the designer. Thus, choosing a fit protective mask for various environments is important (Ardon-Dryer et al., 2021; Bazaluk et al., 2021). The user could use this platform to select a suitable and adequate protective mask for different situations and environments. By wearing the proper protective face mask, people can also protect others from getting the infection (Santarsiero et al., 2021). This sense of curiosity will want to know how this protective mask could help us not get infected.

Moreover, extra features such as animation showing the layers of material used to produce the protective mask will allow the user to visualize the airborne contamination and pathogens filter before inhaling (Shelus et al., 2020). As a result, healthcare workers, employees, or even an average person will learn and gain information from these apps. There was also a study in project development on interactive indoor cycling exercises using virtual video games (Kassim & Said, 2018). This project monitors heartbeat rate, cycle speed, finishing line, traveled distance, and stamina levels (Cho et al., 2009; Lee et al., 2008). Even though this project uses virtual reality, the control system was mobile. Therefore, this project motivated the user to exercise to maintain their health and stamina. Thus, this project could be helpful for this research as it could help this study collect data while using the protective face mask, such as inward leakage test and breathing resistance.

Next total resistance caused by the mask during expiration is called exhalation resistance, and the total resistance caused by the mask during inspiration is called inhalation resistance. Breathing resistance is exhalation and inhalation resistance (Ramirez & O'Shaughnessy, 2016). Striving the performance of the respirator in the workplace by evaluating particle penetration and breathing resistance (BR) of N95 filtering face-piece respirators (FFRs) under simulated air environmental conditions was studied (Ramirez, 2015). The results show that the respirators' breathing resistance will increase relative to humidity and low temperature outside the respirators. Yao et. al (2019) found the effect of structural characteristics on breathing resistance qualities in an overall assessment and characterization of face masks' comfort feeling and performance linked to breathing resistance for healthcare in fog and haze weather. Therefore, even if people are outside and using a face mask, there is a need for time to do a deep breathing method. Breathe and inhale-exhale for eight minutes or open the face mask for a moment to breathe fresh air. It could help improve breathing for a moment and help to release stress (Hariharan et al., 2021; Jerath et al., 2006).

Augmented Reality

AR has a solid potential for development in the future. AR has been implemented in many industries, critically impacting people’s lives: education, healthcare, and security. Furthermore, the mobile phone is a reliable and helpful device for AR to easily access millions of people. Mobile augmented reality (MAR) gadget that uses sensors to determine the position and details of the environment. These combined features resulted in the merging of reality and the virtual world (Ganapathy, 2013). Figure 3 depicts the MAR system’s architecture.

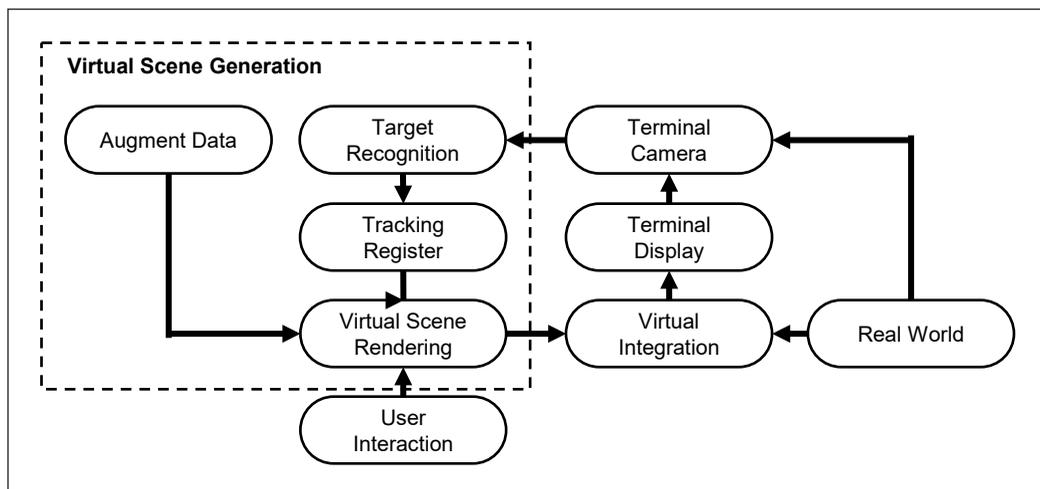


Figure 3. The architecture of the MAR System (Ganapathy, 2013)

A previous study in which the project was to design augmented reality for engineering equipment in education was done by Ghazali et al. (2019). These mobile apps have helped the student and teachers explain the component well when teaching. This interactive learning will attract students to learn and understand better as the image will show 3D information. Therefore, the apps will be able to help our society to know better during this pandemic. Other’s research in which the project was to design a rehabilitation hand exercise system using video games to help stroke patients done by Mazlan (2020). As the system is interactive and unique, stroke patients will take this chance to learn and move around using it, motivating them to start moving their hands. Therefore, it shows that augmented reality will help others learn, train, and play video games.

METHODOLOGY

More than 73 papers have been reviewed from journals or previous research proceedings. There are mainly from IEEE Xplore, ResearchGate, and Google Scholar. Figure 4 shows the review scope in which papers were chosen over five years, beginning in 2015 and 2020.

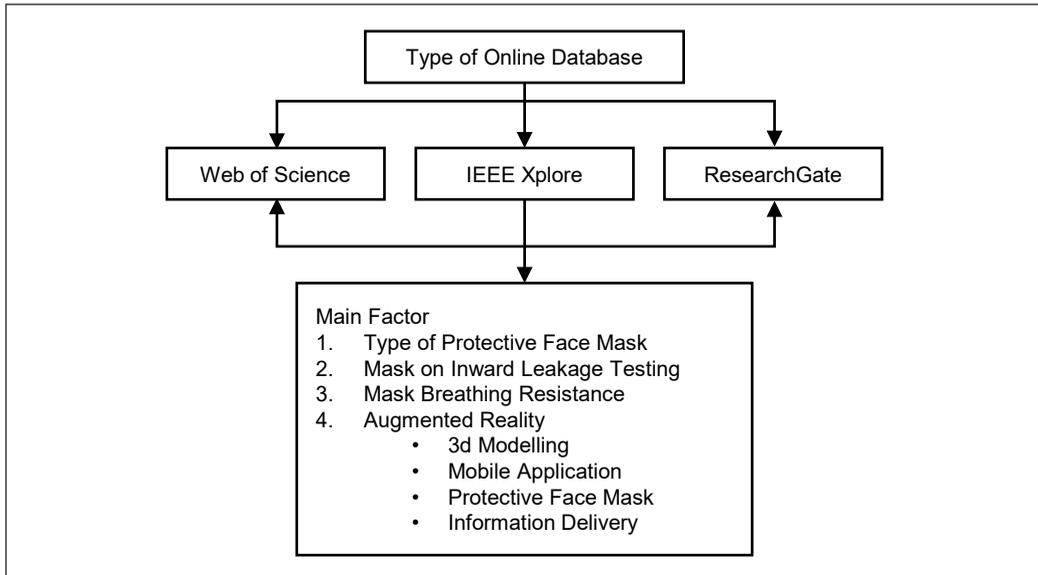


Figure 4. Review structure from publications record

Most of the journals and proceedings are from 2018 to 2021, and only a few journals and proceedings are between 2013 and 2017.

RESULT AND ANALYSIS

This chapter reviews the protective face mask parameter, mask in inward leakage testing parameter, followed by mask breathing resistance parameter, AR's discussion on mobile augmented reality, and 3D Augmented Reality Modelling parameter. Then, the research gap in augmented reality for the protective face mask continued in information delivery and face detection.

Protective Face Mask Parameter

A respirator is a protective device covering the nose and mouth or the entire face or head to guard the wearer against hazardous air environments. Table 1 provides an overview and comparison of several types of protective face masks. Because 0.3 microns is the most demanding particle size, the higher the filter effectiveness, the better the protection against microparticles. As a result of the study review, employing a protective mask primarily on N95 was the best filtering approach.

Mask in Inward Leakage Testing Parameter

Total inward leakage is the combination of contaminated air that leaks through a respirator from various sources, including face seals, valves, and gaskets, and penetration through the

Table 1
Comparison of different protective face masks

Protective Face Mask	Brand	Material Used	Filter Efficiency	Test Agents
	N95 (US)	<u>3 layers</u> 1: non-woven fabric 2: electrostatic fiber cotton 3: skin-friendly lining fabric	0.3 Microns ≥ 95%	NaCl
	FFP2 (EUROPE)	<u>4 layers</u> 1: Spunbound-outer layer 2: Meltblown 3: Airlaid in polypropylene 4: Spunbound skin-friendly	0.3 Microns ≥ 94%	NaCl and paraffin oil
	KN95 (CHINA)	<u>5 layers</u> 1: non-woven fabric 2: filter sponge 3: melt-blown cloth 4: melt-blown cloth 5: pro-muscle non-woven fabric	0.3 Microns ≥ 95%	NaCl
	P2 (AUSTRALIA)	<u>3 layers</u> 1: non-woven fabric 2: electrostatic fiber cotton 3: skin-friendly lining fabric	0.3 Microns ≥ 94%	NaCl
	KF94 (KOREA)	<u>4 layers</u> 1: the outer layer 2: the structural layer 3: melt blown filter 4: hypoallergenic material	0.3 Microns ≥ 94%	NaCl and paraffin oil
	DS2 (JAPAN)	<u>3 layers</u> 1: non-woven fabric 2: electrostatic fiber cotton 3: skin-friendly lining fabric	0.3 Microns ≥ 95%	NaCl

Table 1 (continue)

Protective Face Mask	Brand	Material Used	Filter Efficiency	Test Agents
	Surgical Mask	3 layers 1: non-woven fabric 2: High-density filter layer 3: skin-friendly composite fiber	3.0 Microns ≥ 95% 0.1 Microns ≥ 30%	NaCl
	Airism (Uniqlo)	3 layers 1: fabric mesh 2: micro filter 3: fabric mesh	3.0 Microns ≥ 60% 0.1 Microns ≥ 20%	NaCl

filter (Baugh, 2015). Analysis of protective face masks will test the different groups, such as sex, general facial characteristics, and facial measurements. Table 2 compares several protective face masks used during inward leakage testing. It demonstrates that, in contrast to loose-fitting surgical masks, sealed N95 can significantly minimize inward aerosol leakage and provide functional respiratory protection (Scheepers et al., 2021). The researcher’s methodology included four activities: bending down, talking, moving the head side to side, and moving the head up and down. As a result, the prior research will assist this project plan in determining the technique utilized to match the element in obtaining the information required for the user to comprehend more readily. Therefore, this previous study will help this project plan determine the method used to fit the factor in getting the information needed for users to understand more easily.

Table 2
Comparison of total inward leakage testing results for different protective face mask

Protective Face Mask	Total Inward Leakage Testing
N95	N/A
FFP2	≤ 8% leakage (arithmetic mean)
KN95	≤ 8% leakage (arithmetic mean)
P2	≤ 8% leakage (individual and arithmetic mean)
KF94	≤ 8% leakage (arithmetic mean)
DS2	N/A

Mask Breathing Resistance Parameter

Breathing resistance describes how difficult it is to breathe when wearing a protective face mask (Ramirez & O’Shaughnessy, 2016). The findings indicate that the structural characteristics of a face mask can influence the performance of breathing resistance (Davis & Tsen, 2020). The higher the performance BR, the higher the physical condition of that person to breathe through the protective face mask. People will take this chance not wearing a protective face mask. A review done by a researcher for breathability while

wearing a protective face mask shows that it depends on the materials as it will also affect the filtering efficiency (Kwong et al., 2021). Table 3 compares the inhalation resistance and exhalation resistance of several protective face masks.

Table 3
Comparison of protective face mask based on breathing resistance parameter

Protective Face Mask	Inhalation Resistance	Flow Rate	Exhalation Resistance	Flow Rate
N95	≤ 343 Pa	85 L/min	≤ 245 Pa	85 L/min
FFP2	≤ 70 Pa (at 30 L/min) ≤ 240 Pa (at 95 L/min) ≤ 500 Pa (clogging)	Varied	≤ 300 Pa	160 L/min
KN95	≤ 350 Pa	85 L/min	≤ 250 Pa	85 L/min
P2	≤ 70 Pa (at 30 L/min) ≤ 240 Pa (at 95 L/min)	Varied	≤ 120 Pa	85 L/min
KF94	≤ 70 Pa (at 30 L/min) ≤ 240 Pa (at 95 L/min)	Varied	≤ 300 Pa	160 L/min
DS2	≤ 70 Pa (w/valve) ≤ 50 Pa (no valve)	40 L/min	≤ 70 Pa (w/valve) ≤ 50 Pa (no valve)	40 L/min

The previous study by Amilcar Ramirez strived the performance of the respirator in the workplace by evaluating particle penetration and breathing resistance (BR) of N95 filtering face-piece respirators (FFRs) under simulated air environmental conditions (Ramirez, 2015). The results show that respirators' breathing resistance will increase relative to humidity and low temperature outside the respirators. Next, Yao and Wang proposed research to provide an overall evaluation and characterization of the comfort sensation and performance of face masks related to breathing resistance for healthcare in fog and haze weather and address the influence of structural features on breathing resistance properties (Yao et al., 2019). Twelve face masks in different varieties being an experiment, the results show that the face mask's structural features can affect the performance of breathing resistance (Matuschek et al., 2020). Thus, the researcher hopes to help manufacturers develop and produce high-quality face masks for initial applications.

Mobile Augmented Reality

This review aims to identify the potential use of AR in different fields. Table 4 shows the reviews related to mobile AR applications. Mobile technology has impacted us daily; people use it to work, learn, spend their leisure time, and interact socially. However, the actual implementation of mobile AR is still lacking. Nevertheless, more research needs to happen for future industries in AR and VR.

As shown in summary in Table 4, there are many fields in which AR technology can be adapted for teaching and learning. More research on integrating AR in teaching and learning in terms of protective face masks should be conducted. For example, a mobile

Table 4
Reviews on different types of AR application

Field	AR Features Used	Purpose	Author
Construction Safety Education	Interactive 3D image and guided animation	To provide training in wearing PPE and to have Hazard Inspection Process	(Le et al., 2015)
Medical	Interactive 3D anatomy pictures and feedback	To teach anatomy of the brain and guide in the dissemination of mental health information and self-evaluation	(Bakar et al., 2021; Hossain et al., 2021; Moro et al., 2021)
Art Creation	AR technology in exhibiting the object to image	To provide different multi-finger forms in various manipulation operations and could do the rotate, swipe, pinching, and spreading 3D object	(Bhargava et al., 2017; Kanivets et al., 2019)
Tourism	Interactive 3D AR images and interactive features	To provide information about the country and guide tourists to find the place and food	(Rashid et al., 2017; Acaya et al., 2018; Dangkham, 2018; Demir & Karaarslan, 2018; Safitri et al., 2017)
Vibro Motor Wearable	Head-mounted display and personal interaction panel	To demonstrate how to use the app through tactile human-machine interactions and feeling the tactile feedback impact	(Rumiński & Klinker, 2018)
Animation games	Mobile AR games	To show a 3D augmented image of a virtual pet that lives in the Augmented Reality world	(Costa et al., 2019)
Education	AR technology educational	To teach preschool learning platform and having an interactive layout to make fun with the learning	(Hou et al., 2017; Koca et al., 2019)
	AR learning environment	To overlay the objects and produces the sound of the item name. Helping the daze individuals recognize an object	(Mambu et al., 2019)
	Augmented video, animation, and sound	To provide related phrases and sounds to assist in teaching eating skills to blind people	(Bouaziz et al., 2020)

app uses AR technology-enhanced compared to books for users to visualize the image and find the information about protective face masks.

Augmented Reality Modelling Parameter

Augmented reality is relatively a mixture of both reality and virtual that allows real-world objects to be augmented. Thus, adding a 3D image inside the apps would make it more exciting and visualize better. Researchers anticipated that augmented reality has a very high potential and benefits in teaching and learning, based on augmented reality technology's rapid growth and improvement (Kesim & Ozarslan, 2012). Several researchers have used 3D augmented reality to model from image to object (Cheberyachko et al., 2020; Mahrous

et al., 2021; Reipschläger & Dachsel, 2019; Teng & Peng, 2017). Using 3D augmented reality technology will result in an appealing application for the user. For example, a previous study has used augmented reality to create a 3D picture of the gut for anatomical purposes (Andayani et al., 2019). This result could help the doctors or students explain to the patient if needed for more closure. Therefore, for this project, there will be an interface where the user can play around with the protective face mask and know the layer-by-layer structure of the protective face mask.

Other than that, the design technique for designing the AR marker-based would be based on the parameter. For example, the camera on the smartphone may be used as a scanner to detect the augmented reality marker-based that employs the Vuforia as an image tracker. As the matching marker-based found, the 3D model of the marker-based will overlay on the image. An excellent marker-based is quickly and consistently visible under all conditions. Therefore, specific parameters for choosing a suitable marker-based are needed to pay attention. Table 5 shows the marker-based detection parameter (Vuforia, 2020).

Table 5
Marker detection parameter

Parameter	Explanation
Marker-based Image	<ul style="list-style-type: none"> • 4 to 5 Star Rating Upload on Vuforia to have the app detect. The more structures the marker-based, the faster camera is detected.
Distance	<ul style="list-style-type: none"> • Between 30 cm to 2 meter • 12 × 10 cm images From the marker-based and it should be parallel for the image to overlay on the screen.
Color Image	<ul style="list-style-type: none"> • RGB and Contrast color between the background and the marker-based. • 12 × 10 cm images The camera needs to focus on the marker-based rather than other marker-based.
High Resolution	<ul style="list-style-type: none"> • 8-bit and 24-bit image pixel Clearer image for the user to visualize

Augmented Reality for Information Delivery

AR has significantly assisted in giving visualization and information to users by using identifiers such as picture markers to offer the correct information (Roy & Kanjilal, 2021; Shirazi & Behzadan, 2014). Through AR manipulation, people may quickly obtain information ranging from microscopic things to vast terrains. AR will substantially assist in delivering

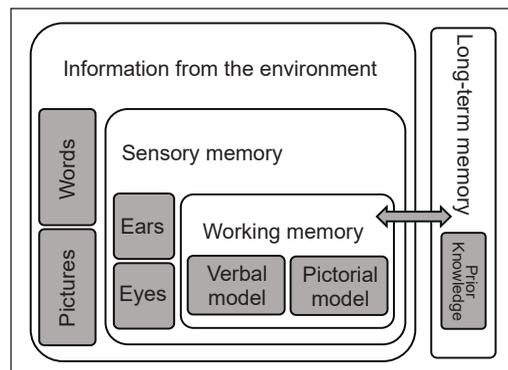


Figure 5. The role of AR in information delivery (Shirazi & Behzadan, 2014)

visualization and information by using identifiers such as picture markers to offer the correct information to the user. Figure 5 depicts the function of augmented reality in information transmission.

Face Detection for Face Mask Sizing

Face masks and respirators come in a variety of shapes and sizes. They are used in the military (e.g., for pilots' oxygen masks), public safety agencies (e.g., firefighters' respirators), medical (e.g., aerosol face masks), and automotive purposes (e.g., paint respirators). Depending on the type of face mask, it intends to provide oxygen or filter air. Most types of face masks must be able to accommodate a wide range of facial shapes. Leakage might result in pollution and inhalation of toxic gases and particles for respirators, which could cause lung illnesses or other health concerns.

Furthermore, ill-fitting oxygen masks that leak into the eyes are irritating, especially to users wearing spectacles. Thus, a tight fit with no leakage is thus essential for constructing an efficient face mask. As a result, automatic face detection systems play a critical role in face identification, facial expression recognition, head-pose estimation, human-computer interaction, and other applications. Face detection is a computer system that locates and sizes a human face in a digital image. Face detection has emerged as a significant issue in the computer vision literature (Kumar et al., 2019). The previous study shows research on designing protective face masks suitable for users to talk or make any motion (Bolkart et al., 2014). Figure 6 shows the landmark point on the face done to measure the size of the protective face mask.

Besides that, a researcher built a semi-automated technique for sizing nasal Positive Airway Pressure (PAP) masks (Johnston et al., 2017). The result shows a 72% accuracy in appropriately sizing a mask and a 96% in sizing within one mask size group. Face detection to measure human faces will help the user buy a suitable and adequate protective face mask. Furthermore, the correct sizing will improve the face seal leakage and filtering percentage when a person inhales. Therefore, it is essential to know the software that will help point the landmark and measure accurate size while integrating it into AR technology.

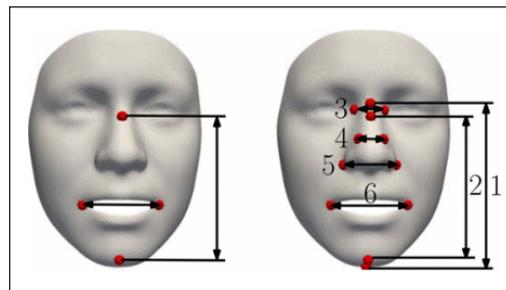


Figure 6. The landmark point (Bolkart et al., 2014)

Analysis of Review Gap in Augmented Reality

Table 6 compares several prior efforts relating to augmented reality use in the mask. The variables being compared include the technology support, parameters, platform, and

software used. These parameters were identified, and new parameters for new developments in this research are included in Table 6.

Table 6
Research gap on augmented reality for a protective mask

Tech Support	Parameter	Platform	Software Used	Author
VR and AR (3D)	Supporting safety training to identify the hazards and training in protecting themselves while working in construction buildings	Mobile	Build AR PRO-2, Revit Architecture 2013, and Blender 2.68	(Le et al., 2015)
VR and AR (3D)	Error-avoidant training approach Automated assessments	Head-mounted display	Unity 5 and custom software integrated with Nintendo Switch	(Eubanks et al., 2016)
AR (3D)	Support practical training in safe donning and doffing Assessment of this training	Mobile	-	(Cahill, 2020)
AR (3D)	Information on fibers of an N95 mask	Website	Spark AR	(Bartzokas et al., 2020)
AR (3D)	Information on a protective face mask and face tracking	Mobile	3D Unity, Vuforia	Current development

CONCLUSION

Augmented reality is essential for future technology in our country as a new medium for creating interactive learning during this pandemic. 3D augmented reality will satisfy the end-user in terms of better visual understanding. At the same time, it reduces the time for the user to find or search information for on various types of protective masks. 3D augmented reality represents a leading technology in our ordinary lives that needs to improve to aid people nowadays in teaching or sharing the information to visualize better. As a result, 0.3 microns is the most challenging particle dust to capture. Thus, having a 95% and above filtered percentage shows that the material used for the protective face masks can protect the wearer from inhaling air pollution. The results also show the comparison on inward leakage testing in which six protective face masks give 8% arithmetic means while at the same time identifying the breathing resistance with the flow rate. Moreover, the result for 3D AR for 3D modeling, mobile application, information delivery, and protective face mask can be analyzed and identified. This comparison of data for different protective face masks helps the research provide better information to the user. The various parameter result will help give the user a better understanding and more knowledge in choosing a suitable and adequate protective mask. This application system will help the country, industries, and employees who require a protective mask's information.

ACKNOWLEDGEMENT

The authors would like to thank the Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), College of Engineering and Research Management Centre (RMC), Universiti Teknologi MARA, Shah Alam, Selangor for the support grant no 600-RMC/GIP 5/3 (105/2021) in supporting the student's research and funding in publishing this paper.

REFERENCES

- Acaya, M. A., Amador, J. P. G., Elinon, E. S., Recaido, R. M. A., Bermudez, J. R. D., Guadanaa, R. R. H., Bangit, A. C., & Ramirez, E. Q. (2018). Fieldtrip Ni Juan: An augmented reality mobile application for the tourist spots in the Philippines for Travel Hub PH. In *2018 IEEE Region 10 Conference (2242-2247)*. IEEE Publishing. <https://doi.org/1109/TENCON.2018.8650364>
- Andayani, U., Syahputra, M., Muchtar, M., Sattar, M., Prayudani, S., & Fahmi, F. (2019). 3D modelling intestine anatomy with augmented reality for interactive medical learning. *IOP Conference Series: Materials Science and Engineering*, *648*, Article 012035. <https://doi.org/10.1088/1757-899X/648/1/012035>
- Ardon-Dryer, K., Warzywoda, J., Tekin, R., Biros, J., Almodovar, S., Weeks, B. L., Hope-Weeks, L. J., & Sacco, A. (2021). Mask material filtration efficiency and mask fitting at the crossroads: Implications during pandemic times. *Aerosol and Air Quality Research*, *21*(7), Article 200571. <https://doi.org/10.4209/aaqr.200571>
- Bakar, W. A. W. A., Man, M., Solehan, M. A., & Sabri, I. A. A. (2021). GAAR: Gross anatomy using augmented reality mobile application. *International Journal of Advanced Computer Science and Applications*, *12*(5), 162-168. <https://doi.org/10.14569/IJACSA.2021.0120520>
- Balamurali, B. T., Enyi, T., Clarke, C. J., Harn, S. Y., & Chen, J. M. (2021). Acoustic effect of face mask design and material choice *Acoustics Australia*, *49*, 505-512. <https://doi.org/10.1007/s40857-021-00245-2>
- Barratt, R., Wyer, M., Hor, S. Y., & Gilbert, L. (2020). Medical interns' reflections on their training in use of personal protective equipment. *BMC Medical Education*, *20*, Article 328. <https://doi.org/10.1186/s12909-020-02238-7>
- Bartzokas, N., Chavar, A. J., Porter, L. Z., Kim, M., & Blufarb, A. (2020). *Developing Animated Sequences for Augmented Reality*. The New York Times Company. <https://rd.nytimes.com/projects/developing-animated-sequences-for-augmented-reality>
- Baugh, L. (2015). *Total inward leakage testing for respirators*. International Safety Equipment Association. <https://safetyequipment.org/knowledge-center-items/total-inward-leakage-testing-for-respirators/>
- Bazaluk, O., Cheberiachko, S., Cheberiachko, Y. I., Deryugin, O. V., Lozynskyi, V., Knysh, I., Saik, P., & Naumov, M. (2021). Development of a dust respirator by improving the half mask frame design. *International Journal of Environmental Research and Public Health*, *18*(10), Article 5482. <https://doi.org/10.3390/ijerph18105482>
- Bhargava, A., Bertrand, J., & Babu, S. V. (2017). AACT: A mobile augmented reality application for art creation. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 254-255). IEEE Publishing. <https://doi.org/10.1109/3DUI.2017.7893368>

- Bolkart, T., Bose, P., Shu, C., & Wuhler, S. (2014). A general framework to generate sizing systems from 3d motion data applied to face mask design. In *2014 2nd International Conference on 3D Vision* (pp. 425-431). IEEE Publishing. <https://doi.org/10.1109/3DV.2014.43>
- Bouaziz, R., Alhejaili, M., Al-Saedi, R., Mihdhar, A., & Alsarrani, J. (2020). Using marker based augmented reality to teach autistic eating skills. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 239-242). IEEE Publishing. <https://doi.org/10.1109/AIVR50618.2020.00050>
- Cahill, D. J. (2020). *Personal protective equipment (PPE) safe*. Centre for Innovative Human Systems (CIHS). https://www.tcd.ie/cihs/projects/ppe_safe.php
- Cheberyachko, S., Cheberyachko, Y., Naumov, M., & Deryugin, O. (2020). Development of an algorithm for effective design of respirator half-masks and encapsulated particle filters. *International Journal of Occupational Safety and Ergonomics*, 28(2), 1145-1159. <https://doi.org/10.1080/10803548.2020.1869429>
- Cherrie, J. W., Apsley, A., Cowie, H., Steinle, S., Mueller, W., Lin, C., Horwell, C. J., Smeuwenhoek, A., & Loh, M. (2018). Effectiveness of face masks used to protect Beijing residents against particulate air pollution. *Occupational and Environmental Medicine*, 75(6), 446-452. <https://doi.org/10.1136/oemed-2017-104765>
- Cho, K. J., Reponen, T., McKay, R., Shukla, R., Haruta, H., Sekar, P., & Grinshpun, S. A. (2009). Large particle penetration through N95 respirator filters and facepiece leaks with cyclic flow. *The Annals of Occupational Hygiene*, 54(1), 68-77. <https://doi.org/10.1093/annhyg/mep062>
- Chua, M. H., Cheng, W., Goh, S., Kong, J., Li, B., Lim, J., Mao, L., Wang, S., Xue, K., Yang, L., Ye, E., Zhang, K., Cheong, W., Tan, B., Li, Z., Tan, B., & Loh, X. J. (2020). Face masks in the new COVID-19 normal: Materials, testing, and perspectives. *Research*, 2020, 1-40. <https://doi.org/10.34133/2020/7286735>
- Costa, A., Lima, R., & Tamayo, S. (2019). Eva: A virtual pet in augmented reality. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)* (pp. 45-51). IEEE Publishing. <https://doi.org/10.1109/SVR.2019.00024>
- Dangkham, P. (2018). Mobile augmented reality on web-based for the tourism using HTML5. In *2018 International Conference on Information Networking (ICOIN)* (pp. 482-485). IEEE Publishing. <https://doi.org/10.1109/ICOIN.2018.8343165>
- Davis, B., & Tsen, L. (2020). Wearing an N95 respiratory mask: An unintended exercise benefit? *Anesthesiology*, 133, 684-686. <https://doi.org/10.1097/ALN.0000000000003421>
- Demir, Ö. F., & Karaarslan, E. (2018). Augmented reality application for smart tourism: GökovAR. In *2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)* (pp.164-167). IEEE Publishing. <https://doi.org/10.1109/SGCF.2018.8408965>
- Eubanks, J. C., Somareddy, V., McMahan, R. P., & Lopez, A. A. (2016). Full-body portable virtual reality for personal protective equipment training. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 490-501). Springer, Cham. https://doi.org/10.1007/978-3-319-39907-2_47
- Ganapathy, S. (2013). Design guidelines for mobile augmented reality: User experience. In W. Huang, L. Alem, & M. A. Livingston (Eds.), *Human Factors in Augmented Reality Environments*, (pp. 165-180). Springer. https://doi.org/10.1007/978-1-4614-4205-9_7

- Ghazali, M. A. A., Kassim, M., Ya'acob, N., Idris, A., & Saaidin, S. (2019). Mobile application on garbage collector tracker using google maps. In *2019 IEEE 9th International Conference on System Engineering and Technology, ICSET 2019* (pp. 287-291). IEEE Publishing. <https://doi.org/10.1109/ICSEngT.2019.8906374>
- Ghosh, M., Bhattacharya, A., Ghosh, K., Bhattacharya, K., & Halder, A. (2020). Saving the savior in COVID19 pandemic: Face masks. *International Journal of Health & Allied Sciences*, *9*(5), 58-61. https://doi.org/10.4103/ijhas.IJHAS_70_20
- Hamid, N. F. I. A., Din, F. A. M., Izham, S., & Isa, S. N. M. (2016). An interactive mobile augmented reality for advertising industry. In *Proceedings of SAI Intelligent Systems Conference* (pp. 763-770). Springer. https://doi.org/10.1007/978-3-319-56994-9_52
- Hao, W., Parasch, A., Williams, S., Li, J., Ma, H., Burken, J., & Wang, Y. (2020). Filtration performances of non-medical materials as candidates for manufacturing facemasks and respirators. *International Journal of Hygiene and Environmental Health*, *229*, Article 113582. <https://doi.org/10.1016/j.ijheh.2020.113582>
- Hariharan, P., Sharma, N., Guha, S., Banerjee, R. K., D'Souza, G., & Myers, M. R. (2021). A computational model for predicting changes in infection dynamics due to leakage through N95 respirators. *Scientific Reports*, *11*, Article 10690. <https://doi.org/10.1038/s41598-021-89604-7>
- Hossain, M. F., Barman, S., Biswas, N., & Bahalul Haque, A. K. M. (2021). Augmented reality in medical education: AR Bones. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 348-353). IEEE Publishing. <https://doi.org/10.1109/ICCCIS51004.2021.9397108>
- Hou, S. M., Liu, Y. Y., Tang, Q. B., & Guo, X. G. (2017). Mobile augmented reality system for preschool education. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)* (pp. 321-323). IEEE Publishing. <https://doi.org/10.1109/ICVRV.2017.00074>
- Huber, T., Goldman, O., Epstein, A., Stella, G., & Sakmar, T. (2021). Principles and practice for SARS-CoV-2 decontamination of N95 masks with UV-C. *Biophysical Journal*, *120*(14), 2927-2942. <https://doi.org/10.1016/j.bpj.2021.02.039>
- Jerath, R., Edry, J. W., Barnes, V. A., & Jerath, V. (2006). Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical Hypotheses*, *67*(3), 566-571. <https://doi.org/https://doi.org/10.1016/j.mehy.2006.02.042>
- Johnson, A. (2016). Respirator masks protect health but impact performance: A review. *Journal of Biological Engineering*, *10*(4). <https://doi.org/10.1186/s13036-016-0025-4>
- Johnston, B., McEwan, A., & Chazal, P. D. (2017). Semi-automated nasal PAP mask sizing using facial photographs. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1214-1217). IEEE Publishing. <https://doi.org/10.1109/EMBC.2017.8037049>
- Kanivets, O. V., Kanivets, I. M., Kononets, N. V., Gorda, T. M., & Shmelter, E. O. (2019). Development of mobile applications of augmented reality for projects with projection drawings. *CEUR Workshop Proceeding*, *2547*, 262-273.
- Karuppasamy, K., & Obuchowski, N. (2021). Comparison of fit for sealed and loose-fitting surgical masks and N95 filtering facepiece respirators. *Annals of Work Exposures and Health*, *65*(4), 463-474. <https://doi.org/10.1093/annweh/wxaa125>

- Kassim, M., & Bakar, A. S. A. (2021). The design of augmented reality using unity 3D image marker detection for smart bus transportation. *International Journal of Interactive Mobile Technologies (IJIM)*, 15(17), 33-48. <https://doi.org/10.3991/ijim.v15i17.22071>
- Kassim, M., & Zubir, M. T. H. M. (2019). Design of augmented reality for engineering equipment in education. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(6), 2773-2781, Article 15. <https://doi.org/10.30534/ijatcse/2019/15862019>
- Kassim, M., & Said, M. N. H. M. (2018). Data analytics on interactive indoor cycling exercises with virtual reality video games. In *2018 4th International Conference on Control, Automation and Robotics (ICCAR)* (pp. 321-326). IEEE Publishing. <https://doi.org/10.1109/ICCAR.2018.8384693>
- Kesim, M., & Ozarslan, Y. (2012). Augmented reality in education: Current technologies and the potential for education. *Procedia - Social and Behavioral Sciences*, 47, 297-302. <https://doi.org/https://doi.org/10.1016/j.sbspro.2012.06.654>
- Koca, B. A., Çubukçu, B., & Yüzgeç, U. (2019). Augmented reality application for preschool children with unity 3D platform. In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-4). IEEE Publishing. <https://doi.org/10.1109/ISMSIT.2019.8932729>
- Kumar, A., Kaur, A., & Kumar, M. (2019). Face Detection Techniques: A Review. *Artificial Intelligence Review*, 52, 927-948. <https://doi.org/10.1007/s10462-018-9650-2>
- Kwong, L. H., Wilson, R., Kumar, S., Crider, Y. S., Sanchez, Y. R., Rempel, D., & Pillarisetti, A. (2021). Review of the breathability and filtration efficiency of common household materials for face masks. *ACS Nano*, 15(4), 5904-5924. <https://doi.org/10.1021/acsnano.0c10146>
- Le, Q. T., Pedro, A., Lim, C., Park, H., Park, C., & Kim, H. (2015). A framework for using mobile based virtual reality and augmented reality for experiential construction safety education. *International Journal of Engineering Education*, 31(3), 713-725.
- Lee, S. A., Grinshpun, S. A., & Reponen, T. (2008). Respiratory performance offered by N95 respirators and surgical masks: Human subject evaluation with NaCl Aerosol representing bacterial and viral particle size range. *The Annals of Occupational Hygiene*, 52(3), 177-185. <https://doi.org/10.1093/annhyg/men005>
- Leung, K. H., Hung, K., Ko, C. P., & Lo, S. F. (2019). Design and development of an augmented reality mobile application for medical training. In *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-4). IEEE Publishing. <https://doi.org/10.1109/ICETAS48360.2019.9117464>
- Long, Y., Hu, T., Liu, L., Chen, R., Guo, Q., Yang, L., Cheng, Y., Huang, J., & Du, L. (2020). Effectiveness of N95 respirators versus surgical masks against influenza: A systematic review and meta analysis. *Journal of Evidence-Based Medicine*, 13(2), 93-101. <https://doi.org/10.1111/jebm.12381>
- Mahrous, A., Elgreatly, A., Qian, F., & Schneider, G. B. (2021). A comparison of pre-clinical instructional technologies: Natural teeth, 3D models, 3D printing, and augmented reality. *Journal of Dental Education*, 85(11), 1795-1801. <https://doi.org/10.1002/jdd.12736>
- Mambu, J. Y., Anderson, E., Wahyudi, A., Keyeh, G., & Dajoh, B. (2019). Blind reader: An object identification mobile- based application for the blind using augmented reality detection. In *2019 1st International*

- Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 138-141). IEEE Publishing. <https://doi.org/10.1109/ICORIS.2019.8874906>
- Matuschek, C., Moll, F., Fangerau, H., Fischer, J. C., Zänker, K., van Griensven, M., Schneider, M., Kindgen-Milles, D., Knoefel, W. T., Lichtenberg, A., Tamaskovics, B., Djiepmo-Njanang, F. J., Budach, W., Corradini, S., Häussinger, D., Feldt, T., Jensen, B., Pelka, R., Orth, K., ... & Hausmann, J. (2020). Face masks: Benefits and risks during the COVID-19 crisis. *European Journal of Medical Research*, 25, Article 32. <https://doi.org/10.1186/s40001-020-00430-5>
- Mazlan, N. (2020). Rehabilitation hand exercise system with video games. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1), 545-551. <https://doi.org/10.30534/ijatse/2020/74912020>
- Moro, C., Phelps, C., Redmond, P., & Stromberga, Z. (2021). HoloLens and mobile augmented reality in medical and health science education: A randomised controlled trial. *British Journal of Educational Technology*, 52(2), 680-694. <https://doi.org/10.1111/bjet.13049>
- Ramirez, J., & O'Shaughnessy, P. (2016). Filter penetration and breathing resistance evaluation of respirators and dust masks. *Journal of Occupational and Environmental Hygiene*, 14(2), 148-157. <https://doi.org/10.1080/15459624.2016.1237027>
- Ramirez, J. A. (2015). *Evaluation of particle penetration and breathing resistance of N95 filtering face-piece respirators and uncertified dust masks* (Doctoral dissertation). The University of Iowa, USA. <https://www.proquest.com/dissertations-theses/evaluation-particle-penetration-breathing/docview/1767779432/se-2?accountid=27932>
- Rashid, R. A., Mohamed, H., & Hussin, A. R. C. (2017). Mobile Augmented Reality Tourism Application Framework. In *International Conference of Reliable Information and Communication Technology* (pp. 108-115). Springer. https://doi.org/10.1007/978-3-319-59427-9_12
- Reipschläger, P., & Dachselt, R. (2019). Designar: Immersive 3d-modeling combining augmented reality with interactive displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (pp. 29-41). ACM Publishing. <https://doi.org/10.1145/3343055.3359718>
- Roy, S. G., & Kanjilal, U. (2021). Web-based augmented reality for information delivery services: A performance study. *Desidoc Journal of Library & Information Technology*, 41(3), 167-174.
- Rumiński, D., & Klinker, G. (2018). Augmenting mixed reality applications with the vibro motors wearable. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 401-402). IEEE Publishing. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00116>
- Safitri, R., Yusra, D. S., Hermawan, D., Ripmiatin, E., & Pradani, W. (2017). Mobile tourism application using augmented reality. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-6). IEEE Publishing. <https://doi.org/10.1109/CITSM.2017.8089305>
- Santarsiero, A., Giustini, M., Quadrini, F., D'Alessandro, D., & Fara, G. M. (2021). Effectiveness of face masks for the population. *Annali Di Igiene Medicina Preventiva E Di Comunita*, 33(4), 347-359. <https://doi.org/10.7416/ai.2020.2390>
- Scheepers, P. T. J., Wertheim, H. F. L., van Dael, M., Anzion, R., Holterman, H. J., Teerenstra, S., de Groot, M., Voss, A., & Hopman, J. (2021). Comparative performance testing of respirator versus surgical mask

- using a water droplet spray model. *International Journal of Environmental Research and Public Health*, 18(4), Article 1599. <https://doi.org/10.3390/ijerph18041599>
- Shelus, V., Frank, S., Lazard, A., Higgins, I., Pulido, M., Richter, A., Vandegrift, S., Vereen, R., Ribisl, K., & Hall, M. (2020). Motivations and barriers for the use of face coverings during the COVID-19 pandemic: Messaging insights from focus groups. *International Journal of Environmental Research and Public Health*, 17(24), Article 9298. <https://doi.org/10.3390/ijerph17249298>
- Shirazi, A., & Behzadan, A. (2014). Design and assessment of a mobile augmented reality-based information delivery tool for construction and civil engineering curriculum. *Journal of Professional Issues in Engineering Education and Practice*, 141(3), Article 04014012. [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000229](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000229)
- Sik-Lanyi, C. (2017). Virtual reality healthcare system could be a potential future of health consultations. In *2017 IEEE 30th Neumann Colloquium (NC)* (pp. 000015-000020). IEEE Publishing. <https://doi.org/10.1109/NC.2017.8263275>
- Steinle, S., Sleuwenhoek, A., Mueller, W., Horwell, C. J., Apsley, A., Davis, A., Cherrie, J. W., & Galea, K. S. (2018). The effectiveness of respiratory protection worn by communities to protect from volcanic ash inhalation. Part II: Total inward leakage tests. *International Journal of Hygiene and Environmental Health*, 221(6), 977-984. <https://doi.org/https://doi.org/10.1016/j.ijheh.2018.03.011>
- Teng, C. H., & Peng, S. S. (2017). Augmented-reality-based 3D Modeling system using tangible interface. *Sensors and Materials*, 29(11), 1545-1554. <https://doi.org/10.18494/sam.2017.1703>
- Vuforia. (2020). *Increasing detection distance*. Vuforia Engine Developer Portal. <https://developer.vuforia.com/forum/vumarks/increasing-detection-distance>
- WHO. (2020). *Malaysia: WHO Coronavirus disease (COVID-19) dashboard with vaccination data*. World Health Organization. <https://covid19.who.int/region/wpro/country/my>
- Yao, B. G., Wang, Y. X., Ye, X. Y., Zhang, F., & Peng, Y. L. (2019). Impact of structural features on dynamic breathing resistance of healthcare face mask. *Science of The Total Environment*, 689, 743-753. <https://doi.org/10.1016/j.scitotenv.2019.06.463>
- Zhu, J., He, X., Guffey, S., Wang, L., Wang, H., & Cheng, J. (2020). Performance comparison of N95 and P100 filtering facepiece respirators with presence of artificial leakage. *Annals of Work Exposures and Health*, 64(2), 202-216. <https://doi.org/10.1093/annweh/wxz086>

Suppression of Coffee-Ring Effect on Nitrocellulose Membrane: Effect of Polyethylene Glycol

Sarah Sorfeena Shahrudin, Norhidayah Ideris* and Nur Atikah Kamarulzaman

School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

In the development of the diagnostic kit, it was favorable to have a low antigen concentration due to the difficulty of antigen preparedness and purification. However, it can cause the coffee-ring effect, producing different pattern formations on the selected membrane. It can lead to a false interpretation of the result. Thus, the immobilization of protein solution (lysozyme) as a model protein for antigen, with the addition of hydrosoluble polymer additive onto a membrane, was evaluated to suppress the coffee-ring effect. This research aims to evaluate the effect of polyethylene glycol on the protein solution for coffee-ring effect suppression and to analyze the image of the coffee-ring effect. From the experimental studies, 5 different concentrations (v/v%) of PEG which are 3.0, 2.0, 1.0, 0.1 and 0.01 v/v% is added at 4.0 mg/mL of lysozyme solution before being spotted onto nitrocellulose membrane. The color intensity of the dried spot, together with the formation of the coffee-ring effect, is analyzed by Image-J software. It is the approach to measure the suppression of the ring effect, in which 0.01 v/v% concentration portrays the most faded ring effect on nitrocellulose membrane. This effect occurs due to a surface tension gradient that causes the solute particles to accumulate at the edge of the droplet. As Marangoni flow has been altered, the coffee-ring effect is successfully suppressed; thus, uniform pattern deposition is achieved.

Keywords: Coffee-ring effect, lysozyme protein, membrane, polyethylene glycol (PEG)

ARTICLE INFO

Article history:

Received: 13 October 2021

Accepted: 29 October 2021

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.20>

E-mail addresses:

sarahsorfeena07@gmail.com (Sarah Sorfeena Shahrudin)

norhidayah7229@uitm.edu.my (Norhidayah Ideris)

Ikakamal96@gmail.com (Nur Atikah Kamarulzaman)

* Corresponding author

INTRODUCTION

In the medical field, biosensor technology is developing vastly along with modern discoveries, including protein, fluorescent microarray, and nanomaterials. Although

it has a wide range of applications, one of the properties that define the performance of the biosensor is the sensitivity, which is the limit of detection (LOD) for the biological analyte. Nevertheless, the traditional method is still applicable, as it is for general mass screening and unrequired specialist & high-end instruments such as the dot-blot technique. However, along with the advantages, this technique also has some drawbacks where the signal produced through colorimetric can be faded or discolored. In addition, less antigen used during the preparation can cause a ring-like structure, leading to misinterpretation of the result (Devineau et al., 2016).

The drawbacks stated before often appear in a ring-like pattern, formally known as the coffee-ring effect (CRE), a ubiquitous event formed during the evaporation of sessile droplets on a flat substrate, thus leading to solute particle distribution. It is initiated through the combination of several natural events, including the formation of capillary flow, followed by the pinning of the contact line of the droplet on the surface (Bansal et al., 2018). Driven by the flow, the particles that cannot evaporate and accumulate at the periphery of the contact line in a ring shape. Much research has been conducted to suppress CRE due to its typical challenges involving reduced accuracy and quality of the product. A common example of this phenomenon is pigment ink and inkjet printing drying, where the end product is low quality.

Some methods to suppress CRE are by introducing Marangoni flow, controlling the temperature of the substrate (Carreón et al., 2021), and generating electric fields in the droplet (Mampallil et al., 2015). However, most of the methods mentioned have a complex requirement where they need thermal conductivity of substrate or require conductive liquid (Eral et al., 2011); thus, the addition of hydrosoluble polymer additive, which is polyethylene glycol (PEG), was chosen as a method to suppress the coffee-ring phenomenon in this study. In general, the previous work has been demonstrated with the PEG incorporated into the liquid to minimize the ring-like effect (Cui et al., 2012; Seo et al., 2017). The experiment was conducted on glass slide surfaces where the ring structure is visible before being characterized using a high-end instrument for better understanding. Here, this research work also utilized such PEG since the PEG molecule in the water solution where the solution becomes a surfactant-like solution (Cao & Kim, 1994) causes a decline in surface tension while increases in solution viscosity (Nowak et al., 2016, Hu et al., 2013). Therefore, adding PEG to protein solution is expected to produce a homogenized protein-surface interaction, resulting in a uniform deposition of a suspended particle on the nitrocellulose membrane.

In the medical sector, the membrane technology has been widely used because it can act as an adsorptive surface for the detection of antigens in the blood sample and as a diagnostic tool to detect protein present in the patient sample due to the simplicity of the design and the speed of the information taken (Ahmad et al., 2016b). A low antigen concentration

was preferable because of the complexity of antigen preparedness and purification, but it can cause CRE to occur, producing a different pattern distribution. In addition, it can lead to false-negative test results. These problems have become an issue since it involves the well-being of a person and the health status that was concerned.

Nitrocellulose, nylon, or polyvinylidene fluoride (PVDF) membranes are commonly used for the detection of protein in the medical industry, but nitrocellulose membrane is preferable to use due to its high protein binding affinity and offers higher efficiency of protein transfer compared to a nylon membrane and PVDF membrane (Kurien & Scofield, 2015). Throughout the discovery of this phenomenon, investigations regarding CRE suppression are widely performed where most are being done on a solid surface such as glass slide and silica. Despite that, the study evaluating CRE suppression on membrane surfaces is still unavailable. Therefore, this experiment was conducted to evaluate the effect of PEG on the protein solution to suppress the ring-like effect along with analyzing the formation of nitrocellulose membrane by Image-J software.

MATERIALS AND METHOD

Material

The membrane used in this experiment is nitrocellulose membrane (Hi-Flow™Plus, Merck) with the pore size of HF 135 with 191µm thickness. Polyethylene Glycol (PEG) was chosen as a polymer additive with an average molecular weight of 500-600. In contrast, lysozyme from chicken egg white acts as a reference protein. Ponceau S enhances the colorimetric signal since it is compatible with the chosen membrane. All solutions used were supplied from Sigma-Aldrich.

Membrane Characterization

In characterizing the surface morphology and roughness of nitrocellulose membrane, Scanning Electron Microscope, SEM (VP-SEM SU1510, Hitachi, Japan) with accelerating voltage of 5.0kV and Atomic Force Microscopy, AFM (Ntegra NT MDT, Russia), are used respectively. For water contact angle, the images were captured using a live video camera of the Contact Angle Goniometer (VCA-3000S, AST, USA) and the sessile drop method at room temperature with normal humidity. In addition, the contact angles of the membrane were measured using distilled water on the upper side of the membrane, where the live video camera would capture the image immediately.

Protein Immobilization

Immobilization of Protein Solution onto Membrane. The nitrocellulose membrane was cut into several test strips with the dimension of a 1 × 5 cm rectangle where the upper

surface of the membrane was marked. In this experiment, the lysozyme protein is chosen as the reference protein. The lysozyme protein was chosen due to its low price and had more stability (Du et al., 2014). The concentration of reference protein used was constant which was 4.0 mg/mL while different concentration of polyethylene glycol (PEG) concentration (0.01, 0.1, 1.0, 2.0 and 3.0 vol/vol%) were prepared, respectively.

The reference protein is mixed with 0.01 vol/vol% of PEG solution in a 1:1 ratio. 1 μ L of the mixture was then spotted onto the membrane using a micropipette in a triplicate manner. Next, the membrane was let to air-dry in the laboratory with well-circulating air at the normal room temperature between 20–24 °C for about 20 minutes. Once the membrane had completely dried, it was immersed in approximately 10 mL of Ponceau S solution for 5 minutes for signal enhancement. The membrane was then rinsed with distilled water to wash off the undesired membrane background and air-dried at room temperature before proceeding to the analysis. These steps are repeated using different concentrations of PEG.

Coffee-Ring Effect (CRE) Analysis

The protein blotting pattern distribution image can be seen using the EPSON Perfection L220 scanner with 16 > 8-bit grayscale and 9600 dpi resolution settings. In addition, the characteristics of the pattern deposition on the surface of the nitrocellulose membrane were analyzed using Image-J software (1.8.0_172, USA) to measure the color intensity of the pattern.

RESULTS AND DISCUSSION

Effect of PEG on Suppression of CRE

The mechanism of droplet deposition induced by specific flow refers to Figure 1, where the schematic diagram shows the droplet deposition induced by CRE and Marangoni flow, respectively. During the evaporation process, the temperature of the surface droplet was reduced in a non-homogenous manner. It is due to the low temperature of liquid-air interaction at the droplet's upper and the high surface tension. Therefore, the flow pattern at the specific area changed into an inward direction near the droplet surface, where the shear stress stabilizes the Marangoni stress (Marin et al., 2016). In the CRE phenomenon, the observed flow was radial outflow that forms due to the absence of the circulating flow.

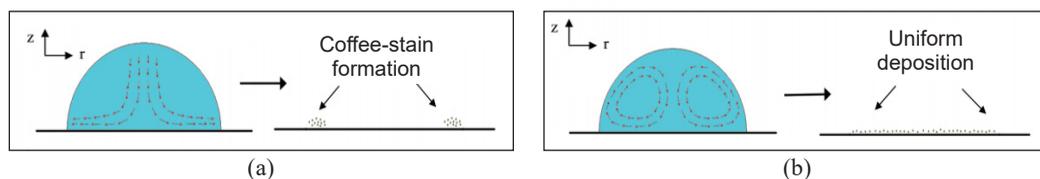


Figure 1. Schematic diagram of droplet deposition induced by (a) coffee-ring effect and (b) Marangoni flow (Majumder et al., 2012).

As the particle flow in a droplet has a differential evaporation rate thus, it drives the particle to move to the periphery of the droplet leading to the formation of CRE. As a result, it was observed that strong outflow occurred at the bottom of the surface, as shown in Figure 1(a).

Meanwhile, the suspended particle of PEG solution has a radial inward flow known as Marangoni flow caused by differences in surface tension. Together with the air and liquid surface interfacial, the surface tension will be maximized, which leads to the formation of radial inward flow near the air and liquid surface interfacial. The presence of higher surface tension at the top of the droplet brings the protein particle to the droplet's upper area before it circulates downwards, where the particle either will be absorbed into the substrate or transferred to the edge of the droplet. This produced uniform deposition where the recirculating flow will homogenize the particle concentration (Majumder et al., 2012). As both flows reach equilibrium, a closed-loop flow known as the Marangoni vortex is formed (Seo et al., 2017) as both flows have opposite directional flow characteristics, as shown in Figure 1(b).

Figure 2 shows protein blotting on a nitrocellulose membrane with the constant lysozyme protein solution concentration of 4.0 mg/mL with the addition of manipulated PEG concentration. The blank solution, Figure 2(a), acts as a reference protein to compare with the other PEG-lysozyme solution to measure the level of CRE suppression. By looking at this comparison, adding PEG in the concentrations of 3.0, 2.0, and 1.0 v/v% shows a ring structure similar to the reference protein. These occur due to the uneven evaporation process on the membrane (Zhuang et al., 2019, Nilghaz et al., 2015). As the evaporation process begins at the edge of the droplet, the capillary flow would direct the particle to accumulate from the center to the periphery of the droplet. For the concentration of 0.1 and 0.01 v/v%, the ring structure form is in a lower intensity and almost similar manner. As 0.01 v/v% utilizes a lower concentration of PEG additive, which is cost-effective for a larger-scale implementation, the analysis proceeds with the selected concentration.

Therefore, when the evaporation process is completed, the particles that are now highly concentrated on the droplet's edge create a ring-like shape known as CRE, as in Figure 3(a). In order to suppress the coffee-ring effect, the PEG solution is mixed with the protein liquid to get a uniform pattern deposition, as in Figure 3(b). In general, the Marangoni

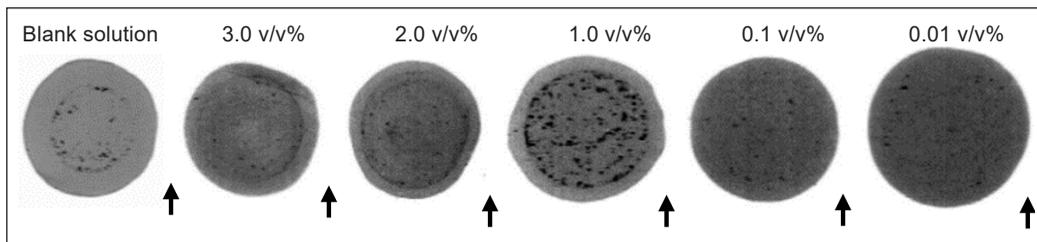


Figure 2. Protein blotting on nitrocellulose membrane (scanned with the setting of 16 > 8-bit grayscale and resolution of 9600 dpi) at 4.0 mg/mL of protein solution with varying PEG concentration

flow in PEG solution produced a slow-paced inward and outward radial motion, leading to an even distribution of the particle on the membrane surface (Seo et al., 2017).

As PEG solution is highly compatible with nitrocellulose membrane due to existing polar and non-polar function groups, this leads to a high rate of hydrophilic interaction. Using AFM to characterize the membrane surface in Figure 4(b), the rough surface of nitrocellulose aids in a faster evaporation process at the edge of the droplet along with highly porous and homogenous throughout the membrane in Figure 4(a). It is also supported by the low receding angle of the membrane when the water contact angle (WCA) test is performed in Figure 4(c). These factors contribute to the accumulation of lysozyme particles at the contact line of the droplet.

Image Analysis of CRE Pattern Formation

The spotted solution on the membrane was first scanned before being imported into Image-J software to analyze using the line profile tool. The image analysis was performed to quantitatively measure the intensity of the CRE formed on the nitrocellulose membrane. The graph was plotted with a y-axis denoted as a grey scale value, which stands for the color intensity of the observed spot for the deposition pattern of the particle on the membrane in grey scale value. A higher value of the grey scale represents the low intensity of the spot (Shahrudin et al., 2021). Figure 5 shows the differences in the color intensity profile of

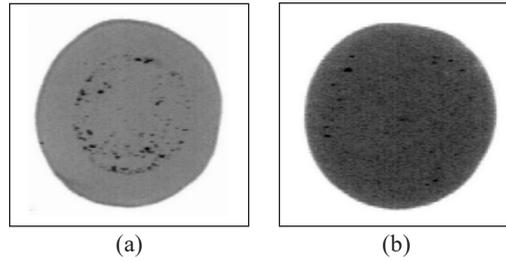
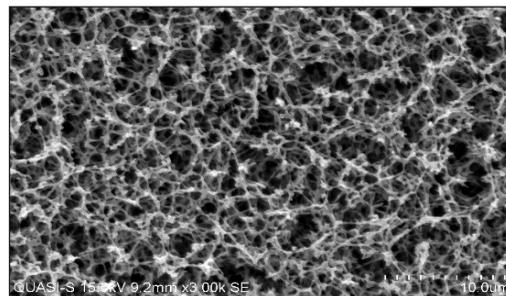
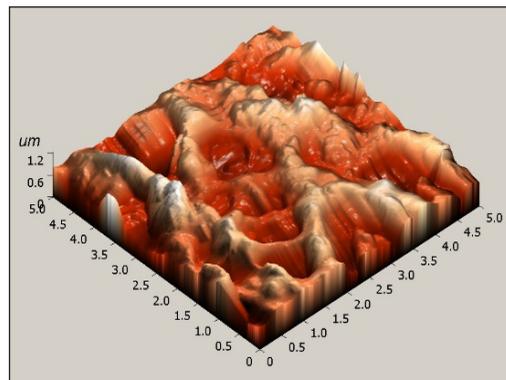


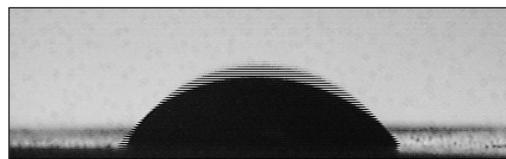
Figure 3. Image of the deposition pattern of the droplet (a) blank solution (lysozyme solution: 4.0 mg/mL) and (b) PEG solution (0.01 v/v%). The blank solution has the presence of a coffee-ring effect, while the solution with the added PEG does not has a coffee-ring effect.



(a)



(b)



(c)

Figure 4. Nitrocellulose membrane images under (a) SEM, (b) AFM, and (c) water-contact angle test

the reference protein solution and the addition of 0.01 v/v% PEG, respectively. For the reference protein, the formation of two valley-like shapes is denoted with a circle in Figure 5(a), indicating CRE's presence. The low grey scale value with a sharp peak shows the presence of CRE, while the grey scale value with a smooth line profile and the absence of apparent valley formation represents no CRE formation on the nitrocellulose membrane. This formation of CRE is due to the suspension of a particle at the edge of the droplet that leads to the formation of CRE at a given time. As for the concentration of 0.01 v/v% of PEG produces a high grey scale value without the obvious valley-like formation. It indicates no deposition of lysozyme protein particles at the edge of the dried spot.

In Figure 5(a), the reading of the grey scale value that portrays the formation of the valley structure in the graph was recorded at 107.574 and 105.694. Besides that, in Figure 5(b), the lowest and highest value of the grey scale value were 27 and 96, respectively. Comparing these readings, the low value obtained from the graph for a solution with a PEG profile supports the image of dried the spots where the CRE is formed in a reference protein. The deposition of protein particles in Figure 5(a) is mostly concentrated at the edge of the spot, whereas in (b), the evaporation is evenly throughout the spot, resulting in a more uniform and dark appearance. It also shows that the protein particles do not accumulate as much at the edge of the spot as at the earlier spot (Ahmad et al., 2016a). However, the CRE produced was not significant enough in Figure 5(b), even though naked eye can be observed the ring-like effect. In addition, the membrane surface roughness can also affect the deposition pattern of the particle. Therefore, it was preferable to use a smooth surface area to increase the even distribution of particles.

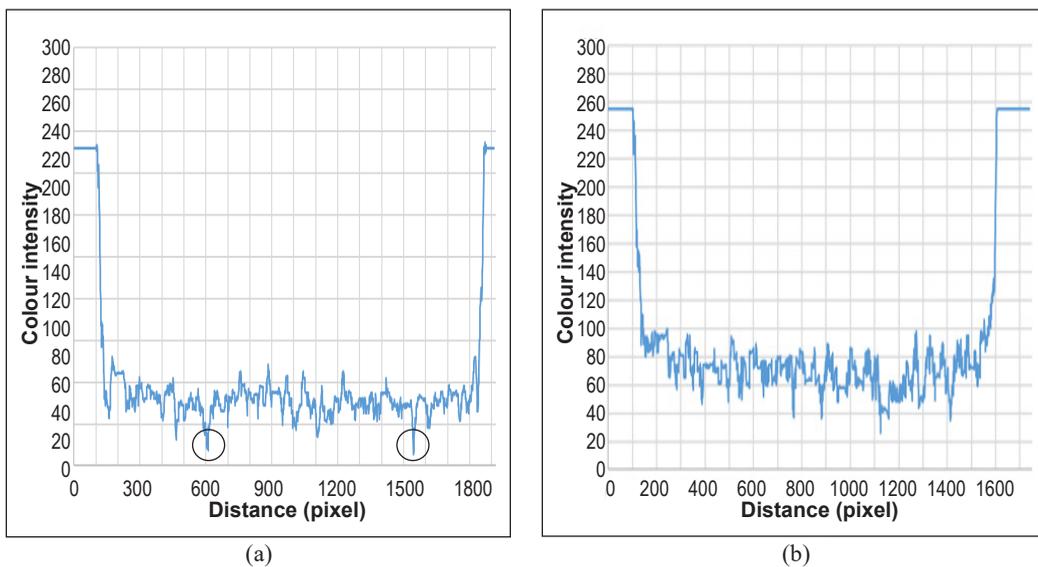


Figure 5. Side profile of greyscale intensity of dried protein spot for (a) blank solution and (b) addition of 0.01 v/v% of PEG solution

CONCLUSION

In conclusion, the uniform pattern deposition can be achieved by adding hydrosoluble polymer additive such as PEG into the lysozyme protein solution before spotting it on the nitrocellulose membrane. The varying concentration of PEG will induce the main contributor to the formation of CRE, which is the Marangoni flow; thus, the CRE can be suppressed. The low concentration of PEG, which was 0.01 v/v%, was favorable to be used in diagnostic testing as it proves to aid in forming a more uniform protein spot with even evaporation. Moreover, the low concentration of PEG may increase the number of diagnostic tools employed in resource-limited countries to include the number of transmittable disease cases, thus improving global healthcare. This study serves as the pipeline for other researchers to include additional high-end equipment for a more detailed study of the interaction of polymeric membrane and polymer additives. Both show excellent results in suppressing this phenomenon.

ACKNOWLEDGMENT

The author gratefully acknowledges the Fundamental Research Grant Scheme (FRGS) funding from the Ministry of Higher Education, Malaysia, and Universiti Teknologi Mara (UiTM). The grant number 600-IRMI/FRGS 5/3 (021/2019).

REFERENCES

- Ahmad, A. L., Ideris, N., Ooi, B. S., Low, S. C., & Ismail, A. (2016a). Impact of membrane pore structure on protein detection sensitivity of affinity-based immunoassay. *Polish Journal of Chemical Technology*, 18(2), 97-103. <https://doi.org/10.1515/pjct-2016-0035>
- Ahmad, A. L., Ideris, N., Ooi, B. S., Low, S. C., & Ismail, A. (2016b). Optimization of polyvinylidene fluoride (PVDF) membrane fabrication for protein binding using statistical experimental design. *Journal of Immunoassay and Immunochemistry*, 37(4), 421-437. <https://doi.org/10.1080/15321819.2016.1157489>
- Bansal, L., Seth, P., Murugappan, B., & Basu, S. (2018). Suppression of coffee ring: (Particle) size matters. *Applied Physics Letters*, 112(21), Article 211605. <https://doi.org/10.1063/1.5034119>
- Cao, B. H., & Kim, M. W. (1994). Molecular weight dependence of the surface tension of aqueous poly(ethylene oxide) solutions. *Faraday Discussions*, 98, 245-252. <https://doi.org/10.1039/FD9949800245>
- Carreón, Y. J. P., Ríos-Ramírez, M., Vázquez-Vergara, P., Salinas-Almaguer, S., Cipriano-Urbano, I., Briones-Aranda, A., Díaz-Hernández, O., Santos, G. J. E., & González-Gutiérrez, J. (2021). Effects of substrate temperature on patterns produced by dried droplets of proteins. *Colloids and Surfaces B: Biointerfaces*, 203, Article 111763. <https://doi.org/10.1016/j.colsurfb.2021.111763>
- Cui, L., Zhang, J., Zhang, X., Huang, L., Wang, Z., Li, Y., Gao, H., Zhu, S., Wang, T., & Yang, B. (2012). Suppression of the coffee ring effect by hydrosoluble polymer additives. *ACS Applied Materials and Interfaces*, 4(5), 2775-2780. <https://doi.org/10.1021/am300423p>

- Devineau, S., Anyfantakis, M., Marichal, L., Kiger, L., Morel, M., Rudiuk, S., & Baigl, D. (2016). Protein adsorption and reorganization on nanoparticles probed by the coffee-ring effect: Application to single point mutation detection. *Journal of the American Chemical Society*, *138*(36), 11623-11632. <https://doi.org/10.1021/jacs.6b04833>
- Du, P., Zhao, J., Mashayekhi, H., & Xing, B. (2014). Adsorption of bovine serum albumin and lysozyme on functionalized carbon nanotubes. *Journal of Physical Chemistry C*, *118*(38), 22249-22257. <https://doi.org/10.1021/jp5044943>
- Eral, H. B., Augustine, D. M., Duits, M. H. G., & Mugele, F. (2011). Suppressing the coffee stain effect: How to control colloidal self-assembly in evaporating drops using electrowetting. *Soft Matter*, *7*(10), 4954-4958. <https://doi.org/10.1039/c1sm05183k>
- Hu, Y. C., Zhou, Q., Ye, H. M., Wang, Y. F., & Cui, L. S. (2013). Peculiar surface profile of poly(ethylene oxide) film with ring-like nucleation distribution induced by Marangoni flow effect. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *428*, 39-46. <https://doi.org/10.1016/j.colsurfa.2013.03.035>
- Kurien, B. T., & Scofield, R. H. (2015). Detection of blotted proteins: Methods and protocols. *Detection of Blotted Proteins: Methods and Protocols*, *1314*, 1-386. <https://doi.org/10.1007/978-1-4939-2718-0>
- Majumder, M., Rendall, C. S., Eukel, J. A., Wang, J. Y. L., Behabtu, N., Pint, C. L., Liu, T. Y., Orbaek, A. W., Mirri, F., Nam, J., Barron, A. R., Hauge, R. H., Schmidt, H. K., & Pasquali, M. (2012). Overcoming the “coffee-stain” effect by compositional marangoni-flow-assisted drop-drying. *Journal of Physical Chemistry B*, *116*(22), 6536-6542. <https://doi.org/10.1021/jp3009628>
- Mampallil, D., Reboud, J., Wilson, R., Wylie, D., Klug, D. R., & Cooper, J. M. (2015). Acoustic suppression of the coffee-ring effect. *Soft Matter*, *11*(36), 7207-7213. <https://doi.org/10.1039/c5sm01196e>
- Marin, A., Liepelt, R., Rossi, M., & Kähler, C. J. (2016). Surfactant-driven flow transitions in evaporating droplets. *Soft Matter*, *12*(5), 1593-1600. <https://doi.org/10.1039/c5sm02354h>
- Nilghaz, A., Zhang, L., & Shen, W. (2015). Coffee stains on paper. *Chemical Engineering Science*, *129*, 34-41. <https://doi.org/10.1016/j.ces.2015.02.017>
- Nowak, E., Kovalchuk, N. M., Che, Z., & Simmons, M. J. H. (2016). Effect of surfactant concentration and viscosity of outer phase during the coalescence of a surfactant-laden drop with a surfactant-free drop. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *505*, 124-131. <https://doi.org/10.1016/J.COLSURFA.2016.02.016>
- Seo, C., Jang, D., Chae, J., & Shin, S. (2017). Altering the coffee-ring effect by adding a surfactant-like viscous polymer solution. *Scientific Reports*, *7*(1), 1-9. <https://doi.org/10.1038/s41598-017-00497-x>
- Shahrudin, S. S., Ideris, N., Abu Bakar, N. F., Ahmad, A. L., & Lah, N. F. C. (2021). Influence of membrane character on suppression of coffee-ring effect. *Materials Today: Proceedings*, *46*, 1870-1874. <https://doi.org/10.1016/j.matpr.2021.01.761>
- Zhuang, J. L., Zhang, Y., Liu, X. Y., Wang, C., Mao, H. L., Du, X., & Tang, J. (2019). Edge enriched self-assembly of Au nanoparticles: Coffee-ring effect during microcontact printing via agarose stamps. *Applied Surface Science*, *469*, 90-97. <https://doi.org/10.1016/j.apsusc.2018.10.261>



Multi-Objective Optimal Control of Autocatalytic Esterification Process Using Control Vector Parameterization (CVP) and Hybrid Strategy (HS)

Fakhrony Sholahudin Rohman^{1,2}, Dinie Muhammad¹, Iylia Idris³, Muhamad Nazri Murat¹ and Ashraf Azmi^{3*}

¹School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 USM, Nibong Tebal, Pulau Pinang, Malaysia

²Department of Chemical Engineering, Universitas Brawijaya, Malang 65145, Indonesia

³School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

The semi-batch esterification of propionic anhydride (PA) with 2-butanol (BT) in the presence of catalyst can be optimised using an optimal control strategy, which utilises the reactor temperature (TR) and feed (FR) flowrate. However, the opposing objective functions, which are maximum conversion (XM) and minimum process time (t_f) in the autocatalytic esterification process, could complicate the optimisation strategy. Simultaneous optimisation of various objectives results in a multi-objective optimal control (MOOC) problem with numerous solutions known as non-dominated (ND) points. In this paper, control vector parameterisation (CVP) and hybrid strategy (HS) are utilised to form Pareto Front (PF) for two opposite targets, which are first to increase XM and secondly to

reduce t_f . Each ND point comprises variant optimal dynamic tracks of TR and FR, which results in various targets of XM and t_f . These solutions provide numerous options for evaluating trade-offs and deciding on the most efficient operating strategy. It is found that the ND point in zone II can be selected as the trade-off of the optimal TR and FR in this study.

ARTICLE INFO

Article history:

Received: 16 October 2021

Accepted: 14 February 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.21>

E-mail addresses:

fathkrk@yahoo.com (Fakhrony Sholahudin Rohman)

annursi@gmail.com (Dinie Muhammad)

iyliaidris@uitm.edu.my (Iylia Idris)

chnazri@usm.my (Muhamad Nazri Murat)

ashraf.azmi@uitm.edu.my (Ashraf Azmi)

* Corresponding author

Keywords: Autocatalytic esterification, multi-objective optimisation, optimal control, Pareto Front

INTRODUCTION

The flavour and scent reagents for the food industry consume ester from the esterification reaction. Esterification is also an essential mechanism in medicine and cosmetics production (Zulkeflee et al., 2021). In industrial practice, the ester is a salience product from batch processes generally utilised to generate high quality and particular products in small quantities while controlling the waste products and raw material losses (Rohman et al., 2021a). The sec-butyl propionate ester (SBP) is synthesised via a reaction involving propionic anhydride (PA) and 2-butanol (BT). The reaction is catalysed by sulphuric acid (Zaldivar et al., 1993).

The mathematical models for optimising esterification conditions can contribute to decision-makers evaluating and executing a wide range of options with fewer attempts (De et al., 2019). The advantages of mathematical modelling include quickly identifying the characteristics and behaviour of the esterification process, low operating costs compared to experimental trials, and increased effectiveness. However, the mathematical modelling of semi-batch autocatalytic esterification working under an unsteady state system requires a composition of differential equation system with technical limitations. Furthermore, the autocatalytic esterification process contains intrinsic nonlinear equations. Thus, optimal control provides the solution for the most effective operating strategy for the optimum time-varying feed rate (FR) and reactor temperature (TR), ensuring that production and efficiency are maximised (Faust et al., 2019).

Presently, the Single Objective Optimisation (SOO) problem is used to solve the specific reference on the optimal control of the esterification between PA and BT for SBP ester synthesis (Rohman et al., 2021a). However, opposing target functions, namely maximum conversion (XM) and lowest process time (t_f), exist in the esterification process optimisation, resulting in numerous compositions of optimum operating conditions. The optimal outcomes of the SOO problem cannot explain the relation between counteracting target functions and cannot give various sets of optimal profiles. Therefore, it is difficult to find a single solution that is best to meet all the targets. For this reason, the Multi-Objective Optimisation (MOO) approach is proposed. It is proven that by using a MOO approach, optimal strategy enhancement can provide a greater way of searching for performance trade-offs arising from opposing targets for the ester production. Nevertheless, the application of MOO in the SBP esterification process has not been reported elsewhere. Furthermore, Multi-Objective Optimal Control (MOOC) in autocatalytic ester production can address the optimisation study void for the SBP ester production process.

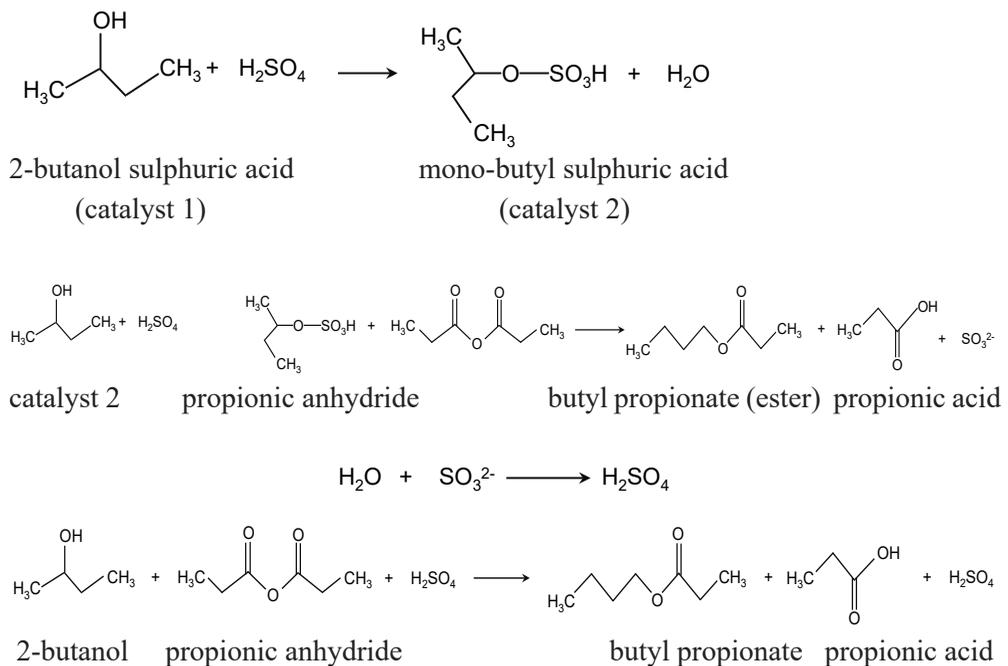
This work addresses the MOOC issue in the esterification process between BT and PA. The most effective strategy of FR and TR is determined to minimise t_f and maximise conversion, XM. Control vector parameterisation (CVP) and a hybrid strategy (HS) are utilised to execute the optimal control problem. The novelty of this research is that this

is the first time for the CVP and hybrid strategy to be applied as multi-objective optimal control of the semi-batch esterification process.

METHODOLOGY

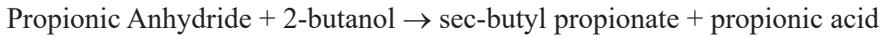
Modelling of PA and BT Autocatalytic Esterification

PA reacts with BT to yield SBP and propionic acid (Pac). The reaction occurs in a homogeneously moderate exothermic reaction. The esterification reaction is catalysed by a strong acid, such as sulphuric acid. Without the sulphuric acid catalyst, the reaction rate between propionic acid and 2-butanol is negligible in the presence of propionic anhydride. Thus, the reaction follows second-order kinetics. According to Zaldivar et al. (1993), autocatalytic behaviour occurs when a catalyst is added. He discovered that in the existence of a catalyst, the PAC increases linearly towards the reaction rate. As PAC concentration increases, the reaction rate also rises, resulting in autocatalytic behaviour. However, PAC does not influence the reaction rate once a certain concentration level is attained. Due to the complexity of the numerous theoretical autocatalytic mechanism, a model is created based on the presumption of two catalysts (cat1, cat2). The former accelerates the second-order reaction, and the latter produces a first-order reaction expression. Furthermore, the transformation of the catalysts was correlated with the acidity function and the concentration of BT. The reaction scheme for autocatalytic esterification propionic anhydride with 2-butanol used in the present reactor modelling is presented as follows:



The esterification reaction mechanisms are represented as the main reaction and reaction for catalyst formation (Zaldivar et al., 1993):

Main Reaction:



Reaction for catalyst formation:



The reaction rate of the main reaction can be written as Equation 1:

$$r_1 = (k_1 + k_2 C_{cat1}) C_A C_B + k_3 C_{cat2} C_B \quad (1)$$

The reaction rate due to the formation of the second catalyst is also considered in Equation 2:

$$r_{cat} = k_4 10^{-H} C_{cat1} C_A \quad (2)$$

These assumptions considered for the construction of the model are constant responding heat capacity, reaction mixture transport characteristics, effectual overall heat transfer coefficient, and density variation exist; negligible heat losses to the environment; even distribution TR and perfect mixing: heat aggregation in the reactor wall is diminished; there are no secondary heating effects, and there is no pressure effect; BT has been designated as the limiting reactant. The acidity function expression is an empirical model which is expressed as Equation 3 (Zaldivar et al., 1993):

$$H = -(p_1 C_{cat1} + p_2 C_c) \left(p_3 + \frac{p_4}{T} \right) \quad (3)$$

Where p_{1-4} is the parameter of an acidity function.

The mass balance equations considered in the optimal control task can be assessed to denote the concentration profile in the autocatalytic esterification reaction.

The mass balances for the semi-batch autocatalytic esterification reactor are presented by Equations (4-10) as shown in Equations 4 to 10 (Ubrich, 2000):

$$\frac{dC_A}{dt} = -r_1 - \frac{F_0 C_A}{V} \quad (4)$$

$$\frac{dC_B}{dt} = -r_1 - \frac{F_0}{V} (C_{B0} - C_B) \quad (5)$$

$$\frac{dC_C}{dt} = r_1 - \frac{F_0 C_C}{V} \quad (6)$$

$$\frac{dC_D}{dt} = r_1 - \frac{F_0 C_D}{V} \quad (7)$$

$$\frac{dC_{cat1}}{dt} = -r_{cat} - \frac{F_0 C_{cat1}}{V} \quad (8)$$

$$\frac{dC_{cat2}}{dt} = r_{cat} - \frac{F_0 C_{cat2}}{V} \quad (9)$$

$$\frac{dV}{dt} = F_0 \quad (10)$$

Concentrations of BT, PA, PAc, SBP, sulphuric acid and mono-butyl sulphuric acid are denoted as C_A , C_B , C_C , C_D , C_{cat1} , and C_{cat2} , respectively. The FR, volume of solution, and TR within the reactor are denoted by F_0 , V , and T , respectively. The initial values of C_A , C_B , C_C , C_D , C_{cat1} , C_{cat2} , and V_j are 3.4M, 0M, 0M, 0M, 1.02×10^{-2} M, 0M, and 1L, respectively.

Reaction rate constants follow Arrhenius law in Equation 11:

$$k_i = k_{0i} \exp\left(\frac{E_{ai}}{RT}\right) \quad (11)$$

The reaction kinetics of this ester production have been demonstrated, and the value of kinetics information is elaborated in Table 1 from Zaldivar et al. (1993). k_1 , k_2 and k_3 are the reaction rate constant for the primary reaction. Meanwhile, k_4 represents the reaction rate constant for the formation of the second catalyst.

Table 1
Kinetic parameter equations (Zaldivar et al., 1993)

Subscript i	E_{ai} (J mol ⁻¹)	Parameter p_i	k_{0i}
1	80,478.64	2.002×10^{-1}	5.36178×10^7 L mol ⁻¹ s ⁻¹
2	79,159.5	3.205×10^{-2}	2.8074×10^{10} L ² mol ⁻² s ⁻¹
3	69,974.6	-21.3754	3.9480×10^{10} L mol ⁻¹ s ⁻¹
4	76,6172.2	12706	1.4031×10^8 L mol ⁻¹ s ⁻¹

Multi-Objective Optimal Control Technique

The optimal control method executed in this study is CVP. The AMIGO2 package developed by Balsa-Canto et al. (2016) using a MATLAB environment is utilised for the CVP method. The CVP approach algorithm in AMIGO2 is constructed from Vassiliadis et al.'s (1994) research. It is based on discretising the control profiles, whereas state profiles remain in continuous form (Azmi et al., 2021). Therefore, the ODE solver first calculates the differential equation.

The original optimal control equations are then re-formulated into a finite-dimensional Nonlinear Programming problem for solution searching by the static optimiser. An appropriate gradient search with the Nonlinear Programming based algorithm is also required. The differential equations are solved at every iteration of the optimal solution searching. The inputs are frequently parameterised using a piecewise-constant (Rohman & Aziz, 2020; Azmi et al., 2020). The technique of parameterisation over finite elements (discretisation) and formation into Nonlinear Programming for CVP is depicted in Figure 1. The CVP method’s standard procedure is described in Equations 12 to 15:

Problem:

$\underset{x,u(t),v(t)}{Min} \mu[d, U(t), V(t)]$	Objective function
$s. t. g[d, U(t), V(t)] \leq 0$	Constraints (Inequality)
$c[d, U(t), V(t)] = 0$	Constraints (Equality)
$\frac{dz}{dt} = F[x, U(t), V(t), t], \quad t \in [0,1]$	Process equation
$V(0) = V_0$	Initial condition (states)
$d^L \leq d \leq d^U$	Decision variables (bounds)
$U^L \leq U \leq U^U$	Control profile (bounds)
$V^L \leq V(t) \leq V^U$	State profile (bounds) (12)

Step 1: The Lagrange Interpolating Polynomial is performed to discretise the control variables (CVs).

$$u_K(t) = \prod_{i=1}^k u_i \delta_i(t) \text{ where } \delta_i(t) = \prod_{1,i}^K \frac{(t-t_k)}{(t_i-t_k)}, \quad u_K(t_i) = u_i, \quad i = 1, \dots, K \quad (13)$$

Step 2: The discretised (CVs) shifted to the ODE model.

$$\frac{dV}{dt} = F(d, V(t), u_K(t), t) \text{ with } V(0) = V_0 \quad (14)$$

Step 3: The discretised - optimisation problem using the CVP method is represented by Equation (10).

$$\underset{x,u(t),v(t)}{Min} \mu[d, U(t), V(t)]$$

$$s. t. g[d, U(t), V(t)] \leq 0$$

$$c[d, U(t), V(t)] = 0$$

$$\begin{aligned} \frac{dV}{dt} &= F(d, V(t), u_K(t), t) \text{ with } V(0) = V_0 \\ d^L &\leq d \leq d^U \\ U^L &\leq U \leq U^U \\ V^L &\leq V(t) \leq V^U \end{aligned} \quad (15)$$

Step 4: In the updated dynamic model, the starting prediction for interval time t_i and decision variables d are substituted, which are then assessed using Runge-Kutta 4th order ODE solver.

Step 5: The hybrid strategy (HS) is utilised to compute the target and constraints in Equation 15 that are evaluated in Step 4 based on the d values. Steps 4–5 are reiterated until convergence is reached.

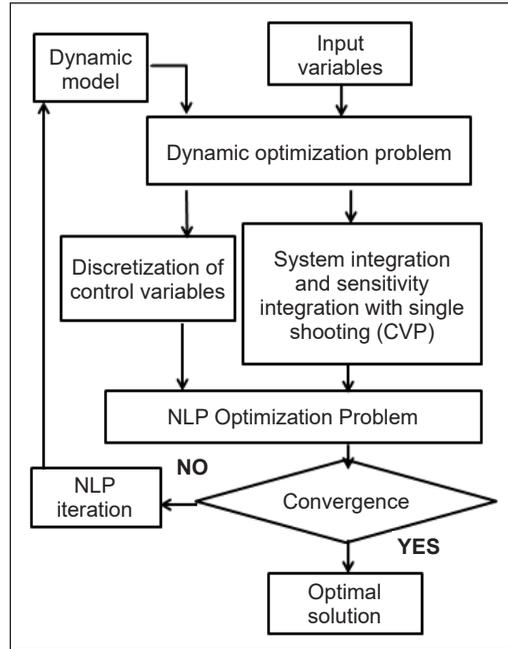


Figure 1. Optimal control method basic procedure: CVP

Hybrid Based Nonlinear Programming Solver. As a Nonlinear Programming solver, two phases of stochastic and deterministic-based optimisation are utilised. First, the stochastic-based Nonlinear Programming solver selects Differential Evolution (DE) (Storn and Price, 1997). In the meantime, sequential quadratic programming (SQP) is considered a deterministic Nonlinear Programming solver. However, it is worth noting that SQP algorithms are tended to multimodal solutions and drag to a premature solution, especially if they start further away from the optimal solution. As a result, the stochastic method, such as differential evolution (DE), is an excellent strategy to resolve the optimisation problem because it can drift away from finding local solutions while locating the optimal solution in reasonable computation time (Storn and Price, 1997). On the other hand, the DE method usually leads to subliminate solutions at a high computational cost. As there is always a counterbalance between convergence rapidity and robustness in both DE and SQP approaches, the HS is established by merging the main parts of the DE and SQP method, leveraging their complementing properties (Banga et al., 2015; Rohman et al., 2016).

The HS is the work of Banga et al. (2014). It is composed of two phases. In the first phase, the DE solver sought the near-optimal solution. When a convergence criterion (SC1) based on the rapid searching distance for each repetition succeeded, the population searching iteration is switched to the 2nd phase. In the final phase, this solution is computed as a near-optimal searching point for the DE solver. Finally, when a convergence criterion

(SC2) is met to achieve a better optimal solution, the optimiser has reached the final convergence and ends the searching process. The SC1 and SC2 convergence values of 0.02 and 10^{-6} , respectively, are determined by empirical data (Storn and Price, 1997; Banga et al., 2015; Rohman et al., 2016). Figure 2 depicts the general steps of the HS method.

Multi-Objective Optimisation (MOO) Technique. The optimum results for MOOs are an arrangement of trade-off values known as the non-dominated (ND) set (Maiti et al., 2011). A set is ND set if it is impossible to enhance one target without

loosening the other's value; the ND set is the most optimum solution for all target functions. The MOO is addressed at each repetition in this method to provide an ND set. The PF and possible counterbalance between target functions can be constructed by updating a set of discontinuous points and asserting ND points obtained from numerous runs.

The ϵ -constraint approach is one of the MOO techniques studied. There is no accumulation in the single target function in the ϵ -constraint technique; the HS optimiser solves the first target. The second target is a constraint using convergence's tolerance values ϵ . The ND points on the PF are updated by gradually changing the ϵ for multiple runs (Rohman et al., 2016). Therefore, the problem can be written as Equation 16:

$$\begin{aligned}
 & \text{Min}_{x,u(t),z(t)} \mu_1[d, U(t), V(t)] \\
 & \text{Subject to} \\
 & \mu_2[d, U(t), V(t)] \leq \epsilon \\
 & F[x, U(t), V(t), t] \in S
 \end{aligned} \tag{16}$$

The ϵ is progressively changed in multiple runs to provide the optimal points on the PF (Rohman et al., 2016). For each optimiser run, HS is utilised for the optimal solution searching.

Problem Optimisation Formulation. The decision variable is the piecewise constants of TR and FR. The catalyst, reactant, and product concentrations are regarded as state

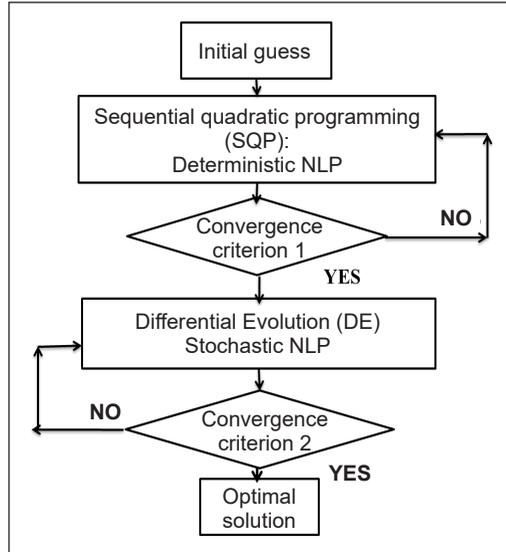


Figure 2. General steps of the HS

variables, and their values are demonstrated as process dynamics for the semi-batch course. As a simplified process model, the catalysed esterification process equations are considered. It only requires the state's differential Equations 4 to 10. This model states optimal profiles of FR and TR, which are employed to accomplish the reactor's target performance. The bounds are specified based on the TR's range capacity (303K- 343K) and pump flow rate ($0 - 5 \times 10^{-4}$ L/s), respectively. The total process time is made up of six intervals of time Δt , which are counted as free final time. As a result, the length of the interval time, Δt , is also optimised, with a range of 10min–30min.

The bi-target functions are to maximise XM while decreasing t_f . The target function is expressed to be of the form min function (to minimise). To maximise $\frac{C_{A0}V_0 - C_A V}{C_{A0}V}$, the max function is expressed as $\min - \left(\frac{C_{A0}V_0 - C_A V}{C_{A0}V} \right)$. The solution volume in the final time, 2L, served as the inequality constraint. The optimal control problem is mathematically formulated as Equation 17:

$$\begin{aligned} \min_{T(t), Fo(t)} \phi_1 &= - \left(\frac{C_{A0}V_0 - C_A V}{C_{A0}V} \right) \\ \min_{T(t), Fo(t), \Delta t} \phi_2 &= t_f \\ \text{Subject to: } Mdx/dt &= f(x(t), u(t), p, t) && \text{(model equation)} \\ 0 \leq F_0 &\leq 5 \times 10^{-4} \text{ L/s} && \\ 303\text{K} \leq T &\leq 343\text{K} && \text{(Lower and upper bounds)} \\ 10 \text{ min} \leq \Delta t &\leq 30 \text{ min} && \text{(Final inequality constraint)} \end{aligned} \quad (17)$$

RESULTS AND DISCUSSION

The Pareto Front (PF) chosen, i.e., ϵ -constraint, as shown in Figure 3, is segregated into three zones. The lower end of the PF (zone 1) is denoted by a relatively short t_f and a lower XM rate. The upper end of the PF (zone 3) is indicated by a relatively long t_f and a high XM rate. Finally, zone 2, positioned between zones 1 and 3, is designated as having a medium t_f and XM rate.

Each point of PF in Figure 3 is associated with a unique FR and TR profile. From Figures 4 to 6, ND points of A, B, and C, positioned in zones 1, 2, and 3, respectively, show a different trend of profiles. Table 2 shows the optimal control results for ND points A, B, and C, based on two reactor performances, regarded as t_f and XM.

Table 2 explains that the t_f for ND points A, B, and C is 46 min, 55 min, and 67 min, respectively. The XM for ND points A, B, and C is 0.994, 0.998, and 0.999, respectively. The variation in the conversion. i.e., 0.992 to 0.999 (Figure. 3: Pareto Front results) is significant

to be explored. The objectives obtained from PF show significant improvement of SOO results where the combination of t_f and XM obtained was 80 min, 0.999 (for maximising XM problem) and 60 min, 0.970 (for minimising t_f problem) (Rohman et al., 2021b). Furthermore, the variation in the conversion. i.e., 0.992 to 0.999 (Figure 3: Pareto Front results) is significant to be explored. It is because the profit value obtained from SOO results varied from RM/year 4.55×10^6 (maximise XM) to RM/year 5.51×10^6 (minimise t_f) (Rohman et al., 2021b); thereby, the increment of 0.001 XM with different t_f will differ profit value significantly.

TR is a decision variable that has a prominent effect on t_f and XM. The reaction rates for the reactant, product, and catalyst escalated with elevating TR (Ubrich, 2000). As a result of the disparate TR profiles, the amount of t_f and XM varies, as shown in Figures 4b, 5b, and 6b. Due to the lowest value of the optimal profile of TR acquired, point A produced the shortest XM and t_f , and vice versa for point C.

The optimal TR profiles (Figures 4b, 5b and 6b) and FR (Figures 4a, 5a and 6a) displayed a significantly different trend and complementary effect in promoting reaction rate. As the TR profile decreased in value, the FR profile equilibrated to sustain the reaction rate by raising the FR value. A higher FR of PA can induce the autocatalytic reaction,

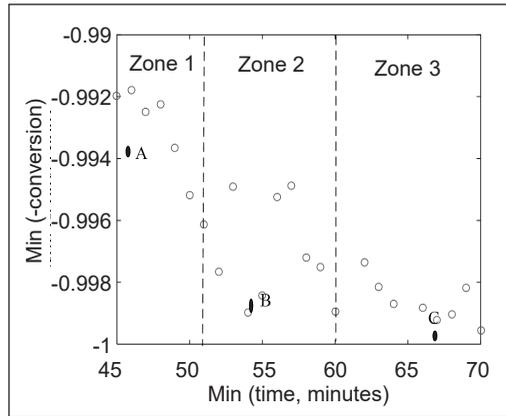
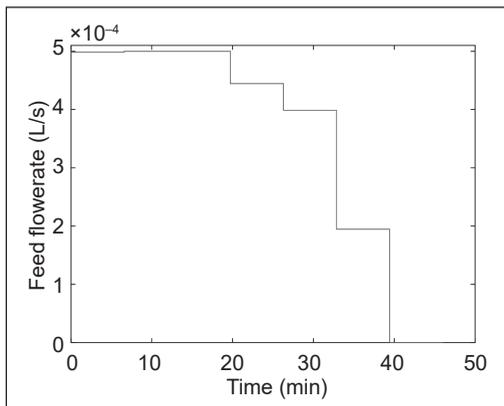


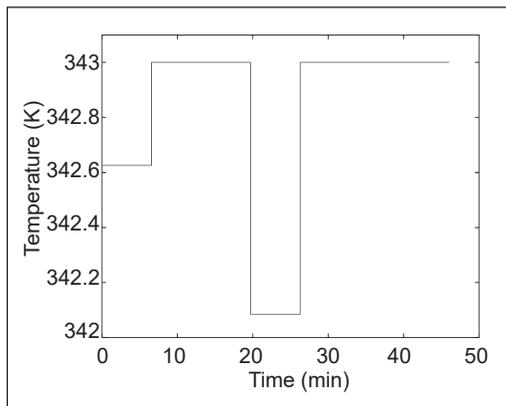
Figure 3. Pareto Front results

Table 2
Final time and conversion in points A, B, and C

Non-dominated point	A	B	C
Final time (t_f , minutes)	46	55	67
Conversion (XM)	0.994	0.998	0.999



(a)



(b)

Figure 4. Optimal trajectories in point A: (a) Feed flowrate; and (b) Temperature

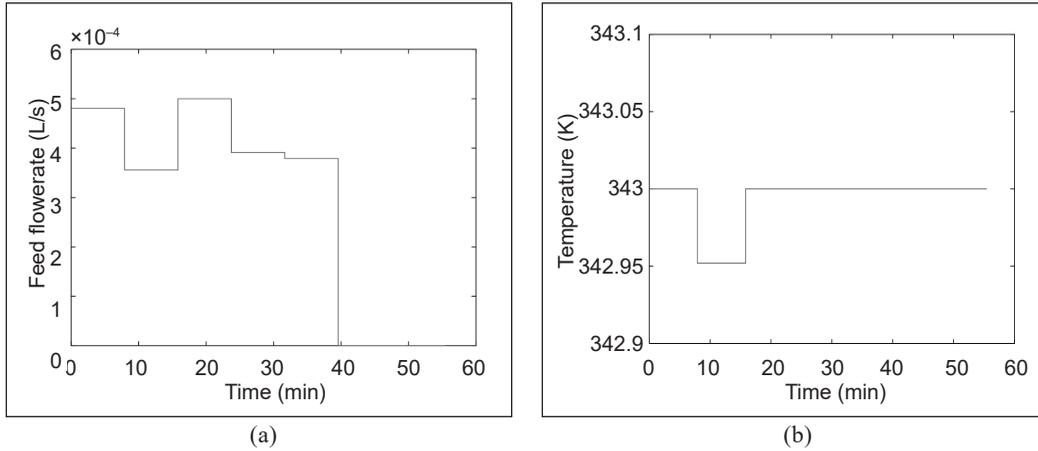


Figure 5. Optimal trajectories in point B: (a) Feed flowrate; and (b) Temperature

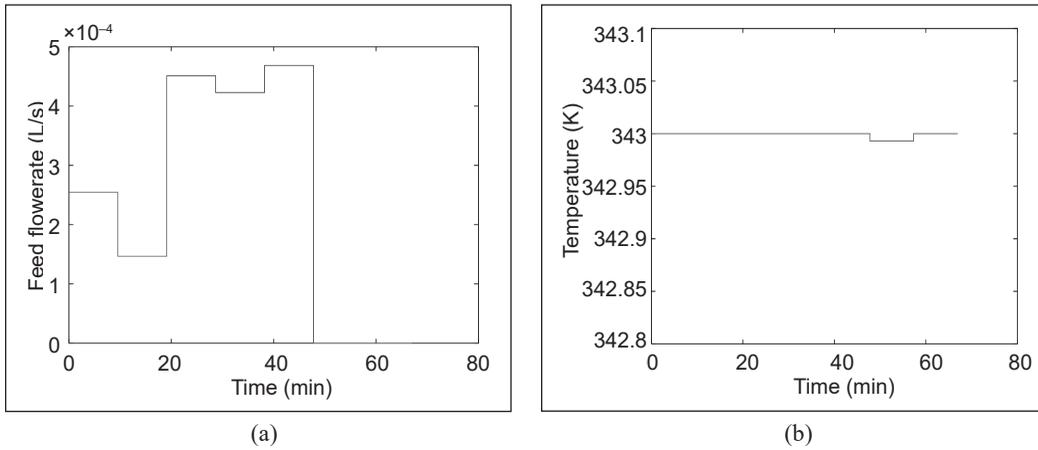


Figure 6. Optimal trajectories in point C: (a) Feed flowrate; and (b) Temperature

thereby escalating the XM rate. Furthermore, the longer the t_f , the less value of the FR. The usage of obtained PF provides the process analyst with several preferences for practical application, which refers to the trade-off between the capacity of ester production and optimal profiles extracted from PF. The increase of XM and the lessened t_f leads to profit. However, the higher TR and FR yield a rise in energy consumption and material cost, respectively. Therefore, it is found that the ND point in zone II (point B) can be selected as the trade-off of the optimal TR and FR in this study.

CONCLUSION

In a semi-batch reactor, the ester production process often has multiple performance targets, some conflicting with one another. For optimal control solver, the CVP and HS have been utilised. The ϵ -constraint has been performed to find PF solutions for the constrained

MOOC problem of minimising t_f and maximising XM. Each ND points along the PF have a variation of optimal FR and TR profiles, resulting in various t_f and XM values. The information contained within the PF enables the process analyst to examine the trade-offs between distinct target functions and to select an appropriate optimal strategy for the ester production. The critical study is that the optimal TR profile resulting from the chosen ND point becomes a pre-specified set point. Point B can be considered the most efficient of optimal TR and FR. The tracking controller then maintains the performance reactor in practice. The multi-objective optimisation study in terms of economic function can be considered a future study.

ACKNOWLEDGEMENT

This research received financial support from the College of Engineering, Universiti Teknologi MARA, Shah Alam through Synergy 2021 Grant No. 600-TNCPI 5/3/DDF (FKK) (009/2021).

REFERENCES

- Azmi, A., Sata, S. A., Rohman, F. S., & Aziz, N. (2020). Optimization studies of low-density polyethylene process: Effect of different interval numbers. *Chemical Product and Process Modeling*, 15(4), Article 20190125. <https://doi.org/10.1515/cppm-2019-0125>
- Azmi, A., Sata, S. A., Rohman, F. S., & Aziz, N. (2021). Dynamic optimization of low-density polyethylene production in tubular reactor under thermal safety constraint. *Chemical Industry and Chemical Engineering Quarterly*, 27(1), 85-97. <https://doi.org/10.2298/CICEQ190108027A>
- Balsa-Canto, E., Henriques, D., Gábor, A., & Banga, J. R. (2016). AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*, 32(21), 3357-3359. <https://doi.org/10.1093/bioinformatics/btw411>
- Banga, J. R., Balsa-Canto, E., Moles, C. G., & Alonso, A. A. (2005). Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. *Journal of Biotechnology*, 117(4), 407-419. <https://doi.org/10.1016/j.jbiotec.2005.02.013>
- De, R., Bhartiya, S., & Shastri, Y. (2019). Multi-objective optimization of integrated biodiesel production and separation system. *Fuel*, 243, 519-532. <https://doi.org/10.1016/j.fuel.2019.01.132>
- Faust, J. M., Hamzehlou, S., Leiza, J. R., Asua, J. M., Mhamdi, A., & Mitsos, A. (2019). Dynamic optimization of a two-stage emulsion polymerization to obtain desired particle morphologies. *Chemical Engineering Journal*, 359, 1035-1045. <https://doi.org/10.1016/j.cej.2018.11.081>
- Maiti, S. K., Lantz, A. E., Bhushan, M., & Wangikar, P. P. (2011). Multi-objective optimization of glycopeptide antibiotic production in batch and fed batch processes. *Bioresource Technology*, 102(13), 6951-6958. <https://doi.org/10.1016/j.biortech.2011.03.095>
- Rohman, F. S., & Aziz, N. (2020). Dynamic multi-objective optimization of autocatalytic esterification in semi batch by using control vector parameterization (CVP) and non-dominated sorting genetic algorithm

- (NSGA-II). *IOP Conference Series: Materials Science and Engineering*, 778, Article 012081. <https://doi.org/10.1088/1757-899X/778/1/012081>
- Rohman, F. S., Sata, S. A., & Aziz, N. (2016). Online dynamic optimization strategy for handling disturbance in semi batch autocatalytic esterification process: Application of hybrid optimizer and simple re-optimization activator. *Advanced Science Letters*, 22(10), 2729-2733. <https://doi.org/10.1166/asl.2016.7021>
- Rohman, F. S., Sata, S. A., Othman, M. R., & Aziz, N. (2021a). Dynamic optimization of autocatalytic esterification in a semi-batch reactor. *Chemical Engineering & Technology*, 44(4), 648-660. <https://doi.org/10.1002/ceat.202000308>
- Rohman, F. S., Sata, S. A., Othman, M. R., & Aziz, N. (2021b). Optimizing autocatalysis with uncertainty by derivative-free estimators. *Optimal Control Applications and Methods*, 42(1), 180-194. <https://doi.org/10.1002/oca.2668>
- Storn, R., & Price, K. (1997). Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341-359. <https://doi.org/10.1023/A:1008202821328>
- Ubrich O. (2000). *Improving safety and productivity of isothermal semi batch reactor by modulating feed rate* (Doctoral dissertation). Swiss Federal Institute of Technology Lausanne, Switzerland.
- Vassiliadis, V. S., Sargent, R. W., & Pantelides, C. C. (1994). Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints. *Industrial & Engineering Chemistry Research*, 33(9), 2111-2122. <https://doi.org/10.1021/ie00033a014>
- Zaldivar, J. M., Hernandez, H., Molga, E., Galvan, I. M., & Panetsos, F. (1993). The use of neural networks for the identification of kinetic functions of complex reactions. In *Proceedings of the third European symposium on computer aided process engineering, ESCAPE* (Vol. 3). Oxford Press.
- Zulkeflee, S. A., Rohman, F. S., Sata, S. A., & Aziz, N. (2021). Autoregressive exogenous input modelling for lipase catalysed esterification process. *Mathematics and Computers in Simulation*, 182, 325-339. <https://doi.org/10.1016/j.matcom.2020.11.006>



Automated Cryptocurrency Trading Bot Implementing DRL

Aisha Peng*, Sau Loong Ang and Chia Yean Lim

Department of Computing, UOW Malaysia KDU Penang University College, 10400 George Town, Pulau Pinang, Malaysia

ABSTRACT

A year ago, one thousand USD invested in Bitcoin (BTC) alone would have appreciated to three thousand five hundred USD. Deep reinforcement learning (DRL) recent outstanding performance has opened up the possibilities to predict price fluctuations in changing markets and determine effective trading points, making a significant contribution to the finance sector. Several DRL methods have been tested in the trading domain. However, this research proposes implementing the proximal policy optimisation (PPO) algorithm, which has not been integrated into an automated trading system (ATS). Furthermore, behavioural biases in human decision-making often cloud one's judgement to perform emotionally. ATS may alleviate these problems by identifying and using the best potential strategy for maximising profit over time. Motivated by the factors mentioned, this research aims to develop a stable, accurate, and robust automated trading system that implements a deep neural network and reinforcement learning to predict price movements to maximise investment returns by performing optimal trading points. Experiments and evaluations illustrated that this research model has outperformed the baseline buy and hold method and exceeded models of other similar works.

Keywords: Automated trading system, deep neural network, reinforcement learning

ARTICLE INFO

Article history:

Received: 28 October 2021

Accepted: 05 April 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.22>

E-mail addresses:

aishaaishapeng@gmail.com (Aisha Peng)

sauloong.ang@kdupg.edu.my (Sau Loong Ang)

cylim@usm.my (Chia Yean Lim)

* Corresponding author

INTRODUCTION

Cryptocurrencies and cryptocurrency trading have significantly succeeded as an emerging business and research topic, with a noticeable uptick in interest and activity. For example, Fang et al. (2022) survey found

Current affiliation:

Chia Yean Lim

School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia

that there are emergent trading technologies, including machine learning technology and other emergent trading methods, portfolio and cryptocurrency assets research, and market condition research since 2018.

There are a few distinct advantages of trading cryptocurrency. For instance, its drastic fluctuations, its 24/7 availability thanks to its decentralised market, low transaction costs using peer-to-peer transactions, and programmable “smart” capabilities bring benefits to holders.

However, trading activities can be controversial as some people relate them to “legal” gambling. However, a good trader looks for consistency, whereas a gambler looks for quick profits. Thus, this research aims to investigate a suitable architecture to integrate deep reinforcement learning (DRL) into a trading bot.

The motive of this trading bot is to train a model to learn from carefully selected technical indicators to learn to trade for profits consistently and accurately. This research aims to expose more people to cryptocurrency applications and provide an option for beginner investors who have no prior knowledge of trading but would like to earn passive income. Additionally, this research hopes to affiliate with people interested in fintech and would like to explore the potential of DRL in trading.

The reason for undertaking this research includes the lack of a baseline. Despite much work and research published on applying DRL to video gameplay and robotics, there is relatively little work on applying DRL to financial trading (Li, 2017). No standard baseline or publicly acknowledged architecture exists to create the proposed model (Huang, 2018). Additionally, this research aims to fill the absence of commercially available products. The current commercially available product commonly uses a trading strategy based on technical indicators to control the trading points, with no application of artificial intelligence (AI).

In addition, simple trading systems may be sufficient for capturing recurrent patterns. However, they are incompetent for the dynamically random cryptocurrency market as they assume a linear relationship with no feedback from the environment. At the same time, the cryptocurrency market has many exception values and is largely affected by external factors, including political and geographical news. Chaos theory—a conceptual framework of nonlinear dynamic systems resolves the unpredictability with distinctive patterns (Cartwright, 1991). Neurons in deep learning architecture can exhibit chaotic dynamics; hence this research aims to utilise deep learning techniques to reconcile the unpredictability of the chaotic trading problems.

This research also aspires to resolve the human hindrance of being always available. Humans have limitations; they cannot focus on the market all the time and always think rationally. Although there are automated functions such as grid trading, these functionalities are simply fixed positions with no flexibility and AI integrated. Not to mention, humans tend to be their own greatest enemy when it comes to financial trading due to emotions. For example, humans fear missing trading opportunities during market euphoria and panic selling in a market downtrend.

LITERATURE REVIEW

Automated Trading System

Automated trading systems (ATS) take past market prices as inputs and generate a trading signal suggesting the optimum. In other words, it is a decision-making system where the dependability and performance of a system are subjective to the efficiency of big data analysis and modelling (Huang et al., 2019).

Many attempts have been made to develop a consistently profitable system in different fields ranging from fundamental analysis and econometric modelling of financial markets to machine learning (Dempster & Romahi, 2002; Moody et al., 1999). Nonetheless, few attempts were successful, and most could not be used to trade actual markets due to associated practical disadvantages such as large draw-downs in profits and excessive switching behaviour resulting in very high transaction costs (Dempster & Leemans, 2006).

Generally, the main advantages of ATS include (1) emotional trading elimination, (2) greater discipline in following strategy rules and more consistent behaviour, (3) virtually guaranteed participation in every important trend, and (4) minimised losses. Regardless, the known cons of ATS are (1) trend-following systems rely on major trends to be profitable, (2) non-trending markets are non-profitable, and (3) unable to recognise that market is not trending (Tucnik, 2010).

Reinforcement Learning

The RL framework includes an agent that optimises its behaviour by interacting with its environment. After taking action in some state, the agent receives a scalar reward from the environment, indicating the quality of that action. The agent's main goal is to find a policy that maximises the total accumulated reward. By iteratively interacting with the environment, the agent uses past experiences to decide on future actions to take in or around a certain state (Grondman et al., 2012).

Proximal Policy Optimisation Algorithm

PPO algorithm is a policy optimisation method where the agent directly learns the policy function that maps state to action. The policy optimisation approach maintains a parameterised action-selection policy. It updates the policy parameters by moving them toward an estimate of the gradient of a performance measure (Van Otterlo & Wiering, 2012). The most commonly used gradient estimator is obtained by differentiating the policy gradient objective function in Equation 1:

$$L^{PG}(\theta) = \hat{E}_t[\log \pi_\theta(a_t|s_t)\hat{A}_t] \quad [1]$$

where θ represents the parameter values (weights), $\pi_{\theta}(a_t|s_t)$ is the policy, and \hat{A}_t is an estimator of the advantage function at timestep t .

However, the advantage estimate \hat{A}_t uses a value function to estimate how good it is for the agent to be in a certain state (Schulman et al., 2015). Since this is only an estimate of a neural network, the estimation will be noisy. Thus, multiple optimisation steps based on the policy gradient loss often lead to destructively large policy updates (Schulman et al., 2017). Furthermore, it will push the policy network into a region of parameter space where the subsequent data will be collected under a very poor policy, causing it never to recover again.

Therefore, the team of OpenAI designed the PPO algorithm based on the trust region policy optimisation (TRPO) to resolve the mentioned issue (Schulman et al., 2017). The core idea is to include a constraint directly into the optimisation objective so the updated policy does not move too far away from the old policy, sticking close to a region where everything works fine. Hence, the objective function for PPO is in Equation 2.

$$L^{CLIP}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad [2]$$

$r_t(\theta)$ is simply the probability ratio between the updated policy network outputs and the old policy network outputs:

$$r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \quad [3]$$

In Equation 3, when the estimated advantage is positive, the L^{CLIP} function clips the r value to limit the effect of gradient update. Similarly, when r goes near 0. Similarly, the update will be limited to prevent reducing the action probability to 0. Since the estimated advantage is noisy, destroying a policy based on a single estimate is unwise.

To summarise, PPO is a family of policy optimisation methods that use multiple epochs of stochastic gradient ascent to perform each policy update. As a result, PPO provides stability and reliability and is simple to implement, requiring only a few lines of code change to a vanilla policy gradient implementation (Schulman et al., 2017).

Deep Neural Network

In 1957, Frank Rosenblatt created the perceptron. It is the prototype of the now-known neural network (NN) (Rosenblatt, 1958). NN is an ML technique inspired by the human brain. It consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Simply put, NN is nothing more than a certain type of nonlinear function (Anthony et al., 1999).

A deep neural network (DNN) is a type of NN modelled as a multilayer perceptron (MLP). DNN is trained to learn representations from data sets without manual feature extractors (Shrestha & Mahmood, 2019). It is used in RL to approximate the value function (Li, 2017). It allows RL to be applied in larger problems as the state action pair grows, unlike non-deep RL, which requires storing the values via a tabular function.

Convolutional Neural Network

Convolutional Neural Network (CNN) is used in this research to leverage its ability to extract useful knowledge and learn the internal representation of the cryptocurrency market data (Livieris et al., 2020).

CNN is mainly composed of two parts: the convolution layer and the pooling layer. Each convolution layer contains a plurality of convolution kernels. The convolution kernel (filter) can be considered a tiny window that “slides” all over the input matrix, applying convolution operation on each subregion (patch) that this specified window “meets” across the input matrix.

After the convolution operation of the convolution layer, the extracted feature dimensions are very high. Therefore, a pooling layer is added after the convolution layer to reduce the feature dimension (Lu et al., 2020). A pooling layer is a subsampling technique to extract certain values from the convolved features to produce a lower dimension matrix. As a result, the pooling layer produces new matrices, which can be considered summarised versions of the convolved features produced by the convolutional layer. The pooling operation can help the system to be more robust since small changes in the input will not change the pooled output values (Livieris et al., 2020).

Long Short-Term Memory

Long short-term memory (LSTM) is a network model designed to solve the longstanding problems of gradient explosion and gradient disappearance in recurrent neural networks (RNN) (Hochreiter & Schmidhuber, 1997).

The LSTM networks are composed of an input layer, one or more hidden layers, and an output layer. Each LSTM memory cell has three main gates (input, output, and forget) to maintain and adjust its cell state. The purpose of the forget gates is to provide a way for the memory cells to reset themselves, which is proved to be important for tasks that require the network to “forget” previous inputs (Graves, 2012). The input gate specifies which information is added to the memory (cell state). In contrast, the output gate specifies which information from memory (cell state) will be used as output information (Wu et al., 2018).

As CNNs are not usually adapted to manage complex and long temporal dependencies, the LSTM network is used in this research to cope with temporal correlations. However, simultaneously, CNN exploits features provided in the training set, which LSTM cannot do.

Therefore, a time-series model which exploits the benefits of both deep learning techniques could improve the prediction performance (Livieris et al., 2020).

Related Works

Several past works have implemented DRL into an ATS. All the results have proven that DRL techniques can outperform baseline methods in the stock market or cryptocurrency market. However, none of the mentioned works implemented a CNN-LSTM in their model. In contrast, this research leverages CNN's high sensitivity to undergo feature extraction from the historical data and technical indicators of the crypto market and LSTM time delay characteristics to build a robust and consistent trading bot.

Sattarov et al. (2020) developed a cryptocurrency trading points recommending system using DRL. Their experimental results on the BTC historical prices are positive, with a 74.6% profit within a month (19 March 2019 to 21 April 2019). However, the trend in that month is generally trending upwards. At the same time, this research will be evaluated in a complete business cycle, where a specific downtrend period is included to ensure fairness and truly evaluate the robustness and stability of the model.

Li et al. (2019) implemented the deep Q-network (DQN) and the asynchronous advantage actor-critic (A3C) as their DRL algorithm to autonomously make trading decisions and gain profits in the dynamic financial markets. They also used an LSTM model as a part of the function approximator. As expected, the A3C outperformed the DQN algorithm as it is too complex to learn the Q function with a value-based algorithm. On the other hand, the policy-based algorithm can learn a good policy since it directly operates in the policy space. Therefore, this research implemented the policy-based PPO algorithm.

Xiong et al. (2018) also implemented a Deep Deterministic Policy Gradient (DDPG) to find the best trading strategy in the dynamic stock market. The results show the DDPG algorithm to achieve a higher return than the traditional min-variance portfolio allocation method and the Dow Jones Industrial Average (DJIA), with an annualised return of 22.24%. The model also achieved a much higher Sharpe ratio indicating it is more robust than the others in balancing risk and return. Nonetheless, there is no mention of data transformation. Cryptocurrencies as an activity in a payment system are influenced by seasonal effects (Haferkorn & Diaz, 2014). Thus, this research implemented differencing techniques to remove seasonality from the historical prices dataset. Furthermore, technical indicators are introduced into the dataset to capture different types of information about the market. It is proven that combining information from technical indicators and macroeconomic variables produces superior forecasts to better track the market's substantial countercyclical fluctuations (Neely et al., 2014).

Lucarelli and Borrotti (2019) made a comparison of Double Deep Q-Network (D-DQN), Dueling Double Deep Q-Network (DD-DQN) and DQN in implementing a BTC

ATS. Their proposed system has an intact stop-loss and take-profit strategy applied upon a certain threshold. Besides, each model is built with two different settings: profit reward function and Sharpe ratio reward function. Their Sharpe D-DQN model shows the best result with an average return of 5.81% and a maximum return of 26.14% from 1 December 2014 to 27 June 2018. Nevertheless, the fact that the model is trained in the 1-minute interval may not be practical for real-time trading as the markets are too volatile, and the model may not be able to consider time delays. Besides, the function implemented was the standard mean squared error (MSE) loss function, which, as mentioned, might cause large inaccurate updates to the model due to the noisy value function, which is a serious issue in RL as updates are carried out iteratively. Unlike the PPO algorithm, which clips the estimated loss, the model is updated in a trusted, safe region.

As of the writing of this journal, there has yet to be the implementation of deep reinforcement learning, especially with the PPO algorithm for trading the cryptocurrency market. Besides, using two different deep learning models is also an upcoming noteworthy area of research. It has not been thoroughly configured to be applied in predicting trading signals in the cryptocurrency markets. This research aims to revolutionise the implementation of deep reinforcement learning on fintech domains, as well as showcase the outstanding trading performance of deep reinforcement learning in trading problems.

METHODS

Research Methodology

There are four main stages conducted in this research. The first stage is the data acquisition from Binance exchange REST API v3. The second stage is the data transformation applied to the dataset before training. Technical indicators and differences, as well as normalisation techniques, are applied to transform the input data. The third stage is the model's training, where RL and DNN architecture is used to update the policy network. The fourth stage is where evaluation analysis is conducted on the trained model to evaluate the accuracy and robustness of the model.

Dataset Interpretation

The dataset of historical k-line is obtained from Binance exchange REST API. The obtained dataset is separated into market cycles, as seen in Figure 1, for training and testing purposes and to allow the model to identify the market cycle phase in different timesteps. It is important to split the training and testing dataset. Just because a model performs well on its training data does not mean it will perform well on data it has never seen, and what matters in this study is that the model can predict in real-time unseen data. (Chollet, 2017). The dataset is first split on Bitcoin's halving date (11 May 2020). Cycles 1 and 2 will be

used to train the model, whereas cycle 3 is the testing dataset. The imbalanced ratio (70/30) ensures the model has enough training dataset to learn for an effective mapping of inputs to outputs. As there is no limitation for setting a splitting ratio, 70/30 is a common dataset splitting proportion. Nonetheless, the optimal splitting ratio varies on the characteristics of different datasets (Vrigazova, 2021).

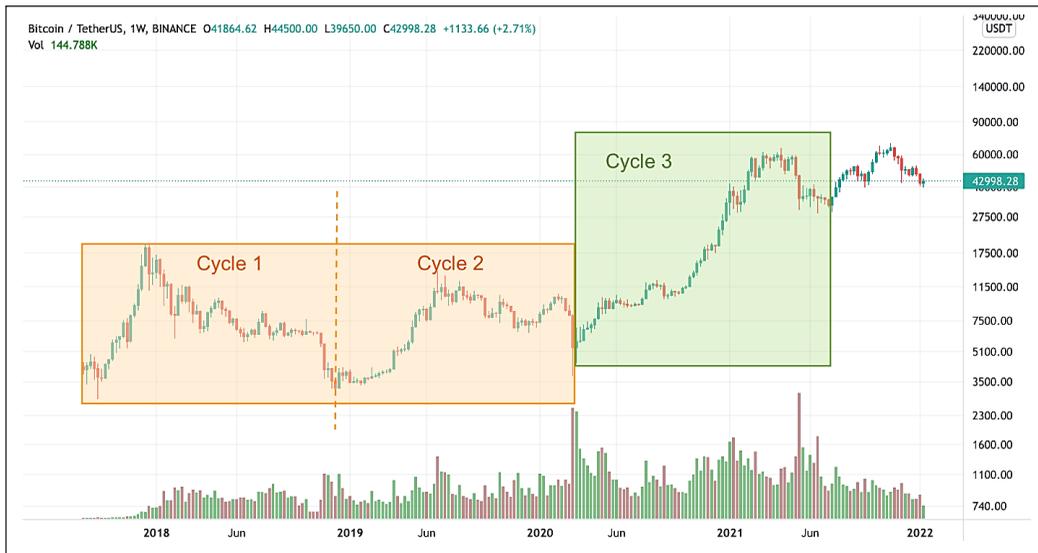


Figure 1. Splitting training and testing dataset

Technical Indicators

Technical indicators are included in the dataset to capture various market information and reduce the noise of the volatile cryptocurrency markets. The chosen technical indicators are true average range (volatility indicator), relative strength index (momentum indicator), and Chaikin money flow indicator (volume indicator). It shall be observed that different categories of technical indicators are used. It improves the model by minimising highly correlated dataset features, as high feature correlation increases redundancy in representations, which is undesirable (Zhang et al., 2018). Furthermore, correlated features add noise and redundancy to the model, making knowledge discovery during the training phase more difficult (Kotsiantis et al., 2006).

These indicators are also chosen because their values are in a limited range. For example, the relative strength index ranges from 0 to 100, and the Chaikin money flow indicator has a zero-centred value. If other indicators were to be used, such as Bollinger Bands, the values are highly dependent on seasonal variations. As PPO uses experience replay to train the model at the end of each episode, having time-dependent values will introduce bias to certain states; as the model goes into implementation, it might not be able to perform well as it does not recognise the values anymore as time changes. For example,

it would be hard for the model to predict a downtrend if it was learned on uptrend data while training.

Data Pre-Processing

In practice, it is nearly always advantageous to apply pre-processing transformations to the input data before it is presented to a network (Bishop, 1995). As stated, time-series data is not stationary, as the model also uses the historical close prices to act as a trend indicator; the data needs to be processed.

The first stage in any analysis should be to see if there is any indication of a trend or seasonal impacts, and if so, remove them. Therefore, the data fed to the stationary model are a realisation of a stationary process (Cowpertwait & Metcalfe, 2009). Figures 2 and 3 show the difference before and after processing the closing data prices. Before normalising the data, a technique known as differencing is applied. Differencing can help stabilise the mean of the time series by removing changes in the level of a time series and eliminating or reducing trend and seasonality (Hyndman & Athanasopoulos, 2018).

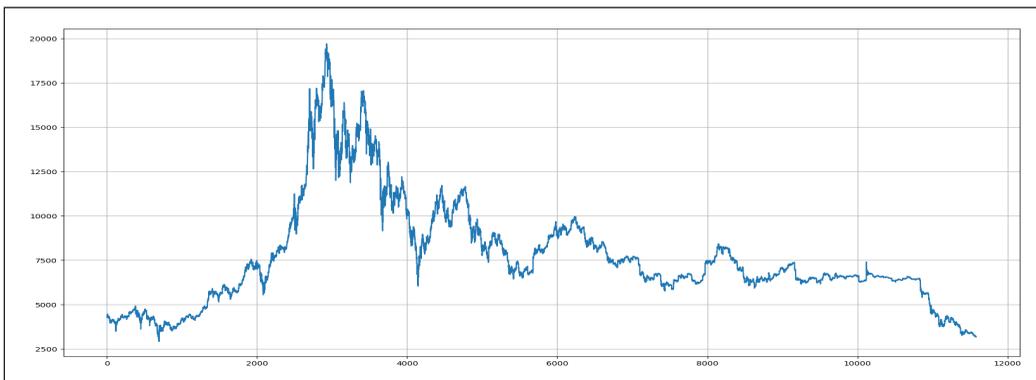


Figure 2. Data before differencing

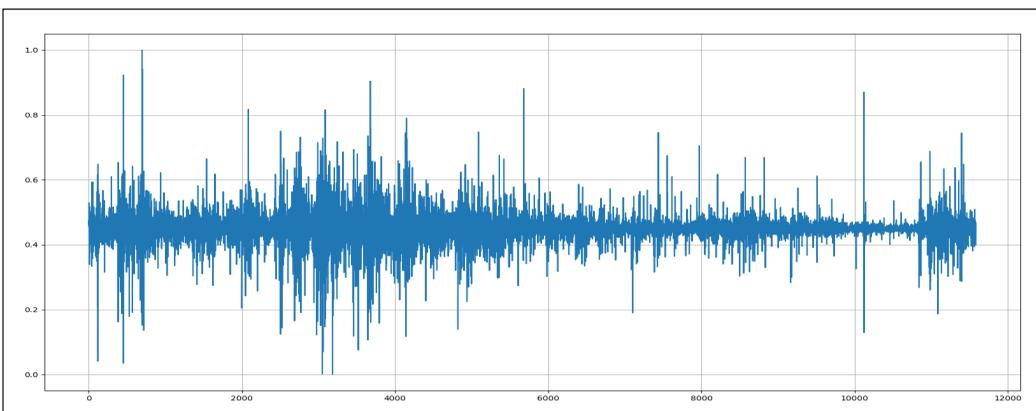


Figure 3. Data after differencing

After applying differencing to the closing data prices, all the feature data, including the technical indicators, are normalised. Research has found that input data normalisation prior to training is crucial to obtaining good results and significantly fastening the calculations (Sola & Sevilla, 1997).

Reinforcement Learning

PPO is chosen as the RL algorithm due to a few factors. Firstly, it must be a model-free algorithm. It is due to the volatility of the crypto market, where the prices are extremely dynamic, and short-term decline can happen unexpectedly. Implementing the model-free approach allows the agent in this research to rapidly adapt when the environment changes its way of reacting to the agent's actions.

Furthermore, PPO uses the actor-critic framework. Typically, RL methods fall under two categories: policy-based (actor) or value-based (critic). Actor-critic methods combine the best of both worlds. The critic uses an approximation architecture to learn a value function, which is then used to update the actor's policy parameters in the direction of performance improvement. As a result, it reduces the high variance of the policy networks, resulting in a more robust model. Additionally, actor-critic methods promise variance reduction, thus resulting in faster convergence (Konda & Tsitsiklis, 1999).

As the trading environment is a continuous state-action space, the actor can produce continuous actions without optimisation procedures on a value function because the critic is there to evaluate the current actor's policy by approximating and updating the value function using samples, which also speeds up the learning process (Grondman et al., 2012). As a result, the updated policy analyses how good the action is and how much better it could be. Hence, the stability provided by the characteristics of the actor-critic framework makes it a great model for the financial trading environment (Yang, 2020).

Keeping in mind that this research aims to build a stable and robust training bot, PPO is a good fit to adapt to a trading strategy that incurs the most profit in the long term instead of an unsustainable strategy that only works for a rapid gain. Generally, PPO is a simple to implement yet reliable RL algorithm; thus, it is a competent algorithm to start for an automated trading domain (Schulman et al., 2017). Additionally, the RL trading environment simulation improves the agent's generalisation and robustness to the environment's changes (Ganesh et al., 2019). A good, simulated environment provides a high degree of reduction to simulate multiple policies for evaluation and improvement, including various cases such as the different phases of market cycles (Zhang, 2021).

Given the cryptocurrency market domain, the RL environment is custom-built to adapt to the cryptocurrency market domain. This RL environment trains the agent through positive reinforcement by interacting with the environment. Essentially, the agent conducts "trial and error" to maximise its reward (Kolm & Ritter, 2019). The agent will be interacting

with the environment using state space, action space and reward functions that are specifically designed. The state-space specifies the observations the agent receives from the environment. The action space specifies the possible interactions between the agent and the environment. Finally, the reward function incentivises to drive the agent to learn a better action (Liu et al., 2020).

Table 1

Example of input states to model at each step. Each state has 100 hours of market information containing the closing price and three different technical indicators

Timestamp	Closing price	Relative strength index indicator	Normalised average true range indicator	On-balance volume indicator
1502942400	4261.32	49.48	1.42	-0.14
1502946000	4291.37	51.89	1.46	-0.07
1502949600	4309.37	56.01	1.48	0.01
...
1503302400	4036.3	48.80	1.33	-0.02

As shown in Table 1, with each step, the states are a lookback window of 100 hours consisting of the market information, including the technical indicator. The technical indicators are derived from the unprocessed dataset's open, high, low, close, and volume (OHLCV). In other words, the model will predict the action based on the 100 hours of information generated by the environment at each step.

Table 2

Possible agent actions as predicted by the model

Actions	Effect
Buy	Buy crypto assets at current price with all the current net worth. Deduct net worth and add holding crypto amount.
Hold	Do nothing.
Sell	Sell all crypto assets. Deduct holding crypto amount and add net worth.

As depicted in Table 2, the possible actions by the agent are buying, holding, or selling. In the RL environment, there is a "net worth" variable that keeps track of the agent's current net worth and another "crypto holding" variable that keeps track of the agent's current holding amount of crypto assets. To simulate a real trading environment, the agent will perform a "buy" action each time the model predicts a buying signal. As a result, the environment will deduct the "net worth" variable and add the amount of crypto bought to the "crypto holding" variable. The amount bought is calculated by Equation 4.

$$\text{Amount bought} = \frac{\text{Current net worth}}{\text{Current crypto closing price}} \quad [4]$$

If the predicted action is held, the agent will not do anything. Whereas, if the predicted action is sold, the environment will now deduct all the current amount of the “crypto holding” variable and adds the equivalent amount of money to the “net worth” variable, which can be calculated by Equation 5.

$$\text{Amount sold} = \text{Current crypto amount held} \times \text{Current crypto closing price} \quad [5]$$

After taking action, the reward for the predicted action is computed using the formula in Equation 6.

$$\text{Reward} = \text{Current net worth} - \text{Previous net worth} \quad [6]$$

For instance, if the agent’s selling action results in a smaller net worth value than its previous net worth, the reward will be negative, incurring a penalty to the model. On the contrary, if the action taken results in a greater net worth value than its previous net worth, the reward will be a positive value. As the model training aims to increase the agent’s net worth, this reward system will drive the model to learn by attempting to maximise the reward and avoid negative rewards.

Figure 4 depicts the model’s flowchart of the training and predicting process. As the actor-critic framework is implemented, the flowchart shows how the critic model is

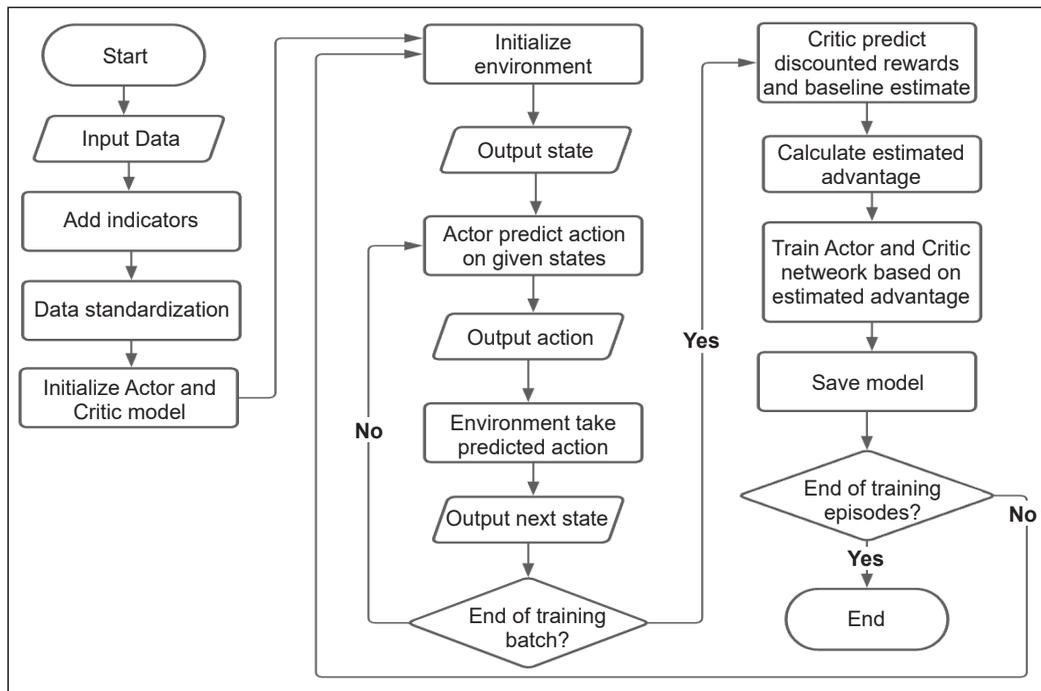


Figure 4. Flowchart of training and predicting process

responsible for predicting the discounted rewards, which will be used to calculate the estimated advantage. The estimated advantage is then used to train the actor and critic models.

Figure 5 depicts the pseudocode for how RL model training is conducted. The environment is reset for each training batch, and the trajectories of the end of each training batch are collected. The collected trajectories are used to compute the estimated advantage and update the policy by training the actor and critic models.

Figure 6 illustrates how RL is integrated with deep learning architecture. The RL environment passes the states and generates the rewards for the deep learning model, and the deep learning model will return predicted action onto the RL environment. In addition, the model loss will be clipped by the PPO loss function to restrict the weight updates during model training.

```

for i = 1, 2, ..., training episodes do
    Environment reset
    for j = 1, 2, ..., training batch size do
        Collect trajectories by running old policy in environment
    end for
    Compute estimated advantage
    Update policy
end for
    
```

Figure 5. Pseudocode for model training

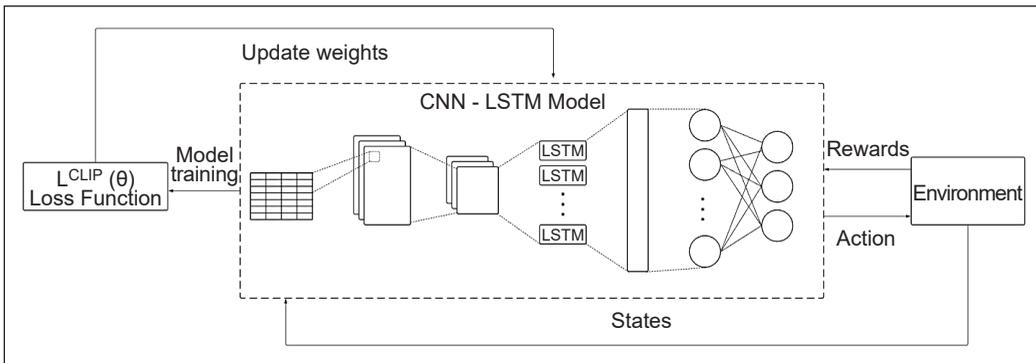


Figure 6. Integration of reinforcement learning and deep learning

Deep Learning Architecture

The CNN-LSTM architecture in this research is designed as shown in Figure 7. As CNNs are not usually adapted to manage complex and long temporal dependencies, the LSTM network is used in this research to cope with temporal correlations. Furthermore, CNN is paying attention to the most obvious features in the line of sight, so it is used in this model to perform feature engineering. Along with its convolutional and pooling layers, it

can filter the input data and extract useful information as an input, enhancing the model's performance (Livieris et al., 2020). Aside from feature engineering, CNN is also superior in its ability to deliberate the relationship between the extracted features and hence perform classification (Liao et al., 2017). Hence, CNN is popular in applications such as image analysis and financial time-series prediction.

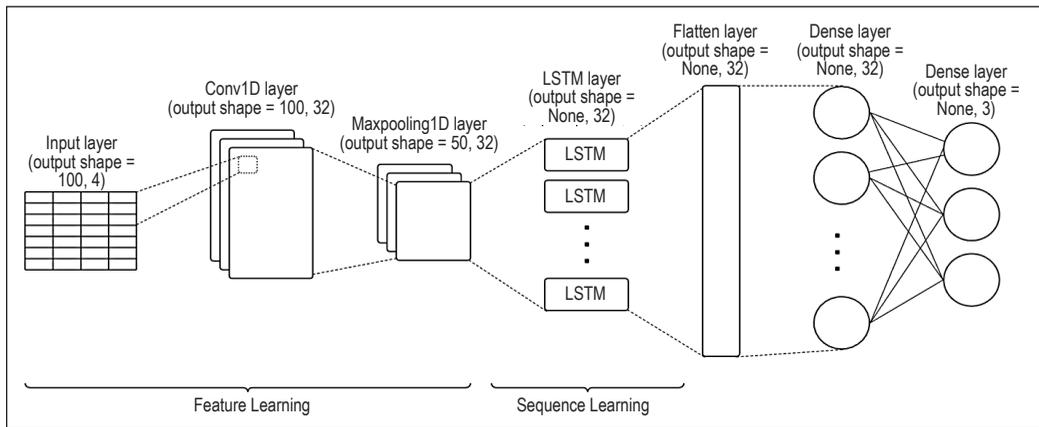


Figure 7. Actor deep learning model architecture

However, CNN is a highly sensitive model as even small translations or rescaling of the input data can drastically change the network's prediction (Azulay & Weiss, 2018). However, it might introduce large variance into the function approximator, causing major updates to the policy network due to the noisy value function. By integrating LSTM with CNN, LSTM can tone down the high sensitivity of CNN by introducing a delay into the model. In addition, LSTM allows the model to be more stable by considering past market information with its long memory by storing the information in its cell state as compared to shorter memory RNN. It allows the model to avoid pitfalls such as a sudden market spike.

The research uses transfer learning to overcome the problem of hard-to-train financial data due to high volatility and insufficient data (Jeong & Kim, 2019). However, this learning approach is not implemented in this research as it may hurt the learning performance in the target domain, a situation often referred to as negative transfer (Pan & Yang, 2009). Furthermore, it prevents the pre-trained model from introducing additional bias into the model, causing the learning process to side-track, especially when BTC holds unique market characteristics due to extreme price movements. Not to mention that investors' attention to the crypto market drastically changed over time (Gronwald, 2014).

The main parameter involved in this proposed model is its weights, which will be optimised through iterative updates during experience replay in the training process. Furthermore, the proposed model integrates with RL to perform optimal weight updating based on the stable PPO loss function instead of the conventional noisy estimated value function.

Evaluation Metric

The evaluation metric is important for discriminating and obtaining the optimal model architecture (Hossin & Sulaiman, 2015). The performance of the proposed model will be evaluated against the baseline buy-and-hold method. The single-period rate of return of the proposed method and the buy and hold method will be compared to determine the profitability of the trained model. Additionally, the model's accuracy will be determined to assess the correctness of the model's trading decision. Finally, the model will be tested on unseen data to determine the system's performance and ensure fairness against the buy-and-hold method.

RESULTS AND DISCUSSIONS

Training Process

The training process will take in the pre-processed training dataset containing two market cycles; as mentioned, the model will interact with input states and react by acting on the predicted action. Therefore, a successful training process for this experiment will decrease model loss and increase net worth. Normally, the model loss will be high initially as the actions predicted are inconsistent with the target output but should decrease in time as the model learns to predict actions based on input states. As a result, the model is expected to yield an increase in net worth from generating accurate trading signals.

The model loss, net worth, and order per episode are recorded to visualise the training process better. The training process took approximately 100 hours to complete 4000 training episodes. Each episode is made up of 1000-time steps. The training parameters for model training in each episode are five epochs with a batch size of 32. In other words, the model will go over each episode five times, where the model will update its weight parameters every 32 batches. As observed in Figure 8, the model can learn as the loss per replay decreases.

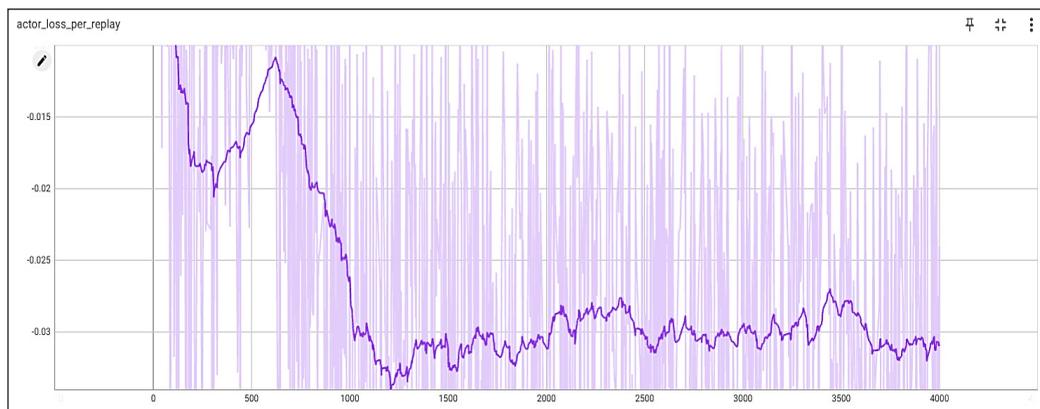


Figure 8. Actor loss per replay

As a result, the model has proven its ability to perform profitable trades as its net worth increases as more training episodes continue, as shown in Figure 9. Notably, the model also learned to perform more trading orders per episode, as illustrated in Figure 10, which is believed to reflect how it can achieve more profits.

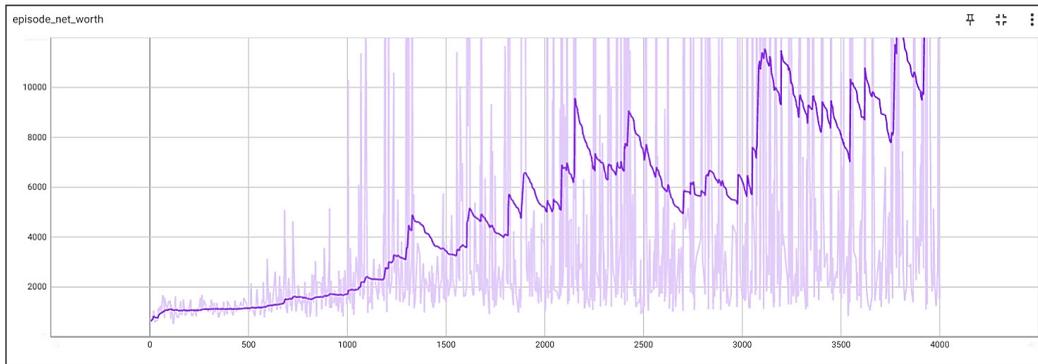


Figure 9. Net worth per episode

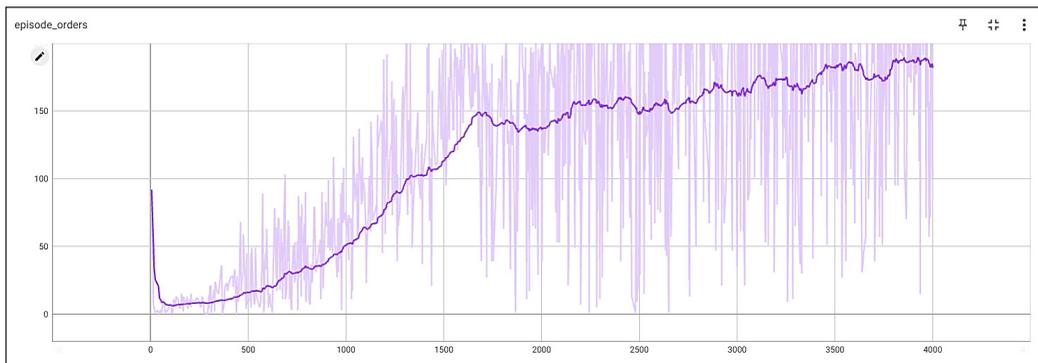


Figure 10. Orders made per episode

Evaluation and Findings

Model Accuracy. The trained model is tested with an unseen dataset with 436 days, ranging from 10 May 2020 to 20 July 2021 for all the cryptocurrencies, except for DOT, as it was only launched on 18 August 2020. Table 3 shows the model's performance, including the number of trades, number of profitable trades, accuracy, total return rate given the time, and annualised return. The model has proved its capability to make profitable trades.

The model's accuracy is evaluated by the number of profitable trades made over the total number of trades made. A trade is considered profitable if the current selling price of the trade is higher than the previous buying price. As the model is initially trained using BTC historical dataset, the accuracy for BTC is the highest among the 10 cryptocurrencies. Nonetheless, the model can perform well with all the cryptocurrencies with high accuracy. Although, this may be due to the nature of cryptocurrencies generally trending upwards.

Table 3
Model performance on the testing dataset

	ADA	BCH	BNB	BTC	DOT
Trades	850	652	1100	1184	313
Profitable trades	763	569	1012	1145	263
Accuracy	89.76%	87.27%	92.00%	96.71%	84.03%
Total return rate	49835618642.99%	15199675.56%	38527152540.10%	169543702.22%	647835.54%
Annualised return	2246218700.00%	2397972.87%	1807081748.09%	18416004.14%	1570920.32%
	ETH	LINK	LTC	XLM	XRP
Trades	824	632	592	895	1152
Profitable trades	741	536	514	804	1096
Accuracy	89.93%	84.81%	86.82%	89.83%	95.14%
Total return rate	421040489.53%	410081019.47%	26941402.38%	4071603763.66%	1494947924270.84%
Annualised return	39728250.52%	2397972.87%	1807081748.09%	18416004.14%	1570920.32%

Table 4
Baseline buy-and-hold method performance on the testing dataset

	ADA	BCH	BNB	BTC	DOT
Total return rate	2280.73%	78.02%	1723.91%	258.64%	253.65%
Annualised return	64632.43%	3777.43%	51117.58%	10395.64%	41160.71%
	ETH	LINK	LTC	XLM	XRP
Total return rate	886.64%	286.06%	170.68%	248.98%	182.74%
Annualised return	29269.75%	11316.05%	7323.75%	10067.89%	7758.00%

The performance of the baseline buys and hold method is shown in Table 4. As compared to the baseline buy and hold method on BTCUSDT, a trader would have acquired a total return rate of 258.64%. Therefore, the model has significantly outshone the baseline method and outperformed previous researchers' models. This evaluation result also substantiates humans' significant limitations of not always being able to trade and missing out on trading opportunities without an ATS.

Table 5 shows the historical annualised return of BTC from 2017 to 2021. Despite having an exceptionally high return rate, the trained model can outmatch the performance of the baseline buy and hold method of all historical annualised returns of BTC with an annualised return of 10395.64% on the testing dataset.

Table 5
Historical annualised return of BTC with baseline buy and hold method

Year	Annualised Return
2017	1331%
2018	-73%
2019	95%
2020	301%
2021	58%

Model Performance on Market Downtrend

To evaluate the performance of the model in a market downtrend, the model is tested on the date ranging from 9 May 2021 to 19 July 2021 as seen in Figure 11. The evaluation results are depicted in Table 6. Despite the falling prices, the model can trade with outstanding results. The accuracy for trading BTC is once again the highest, with 88.79%. It showcased the model's adaptability in different market conditions.



Figure 11. Downtrend in 2021 market cycle

Table 6
Model performance on a market downtrend

	ADA	BCH	BNB	BTC	DOT
Trades	83	66	86	107	41
Profitable trades	58	50	57	95	20
Accuracy	69.88%	75.76%	66.28%	88.79%	48.78%
Total return rate	147.14%	329.53%	324.93%	316.64%	7.30%
Annualised return	13937.88%	287727.96%	271293.66%	243593.93%	46.95%
	ETH	LINK	LTC	XLM	XRP
Trades	90	64	76	77	67
Profitable trades	76	44	44	60	34
Accuracy	84.44%	68.75%	57.89%	77.92%	50.75%
Total return rate	410.91%	266.72%	67.73%	142.45%	-37.47%
Annualised return	742776.57%	121225.78%	1588.02%	12543.58%	-92.31%

However, the results of trading DOTUSDT and XRPUSDT were not as good as others. It may be due to the model's inability to adapt to the different market characteristics compared to BTCUSDT. As a result, its low accuracy causes it to predict an incorrect action, getting caught in a market down spike. Table 7 shows the total return rate of the baseline buy and hold method; all the crypto pairings result in negative results. Comparably, implementing the baseline buy and hold method for XRPUSDT would have lost -63.04%, while the model was able to cut loss to -37.47%. In conclusion, the proposed model is appropriately evaluated and accepted as effective.

Table 7
Baseline buy-and-hold method performance on a market downtrend

	ADA	BCH	BNB	BTC	DOT
Total return rate	-26.87%	-68.91%	-53.17%	-46.00%	-68.78%
	ETH	LINK	LTC	XLM	XRP
Total return rate	-50.85%	-68.13%	-65.86%	-62.93%	-63.04%

CONCLUSION AND RECOMMENDATIONS

This research proposes a new automated cryptocurrency trading system integrated with DRL. This research has perfectly demonstrated its capability to resolve the stated problem statement – overcoming human hindrance and bridging the gap between human and automated trading. The goal of presenting the concept and use of cryptocurrencies to all audiences and utilising the power of machine learning to earn profits has been accomplished. With that said, the experimental analysis of the model manifested exceptional results, even

surpassing similar works by other researchers, thus again fortifying the implementation of DRL in the finance sector.

Undeniably, there are underlying limitations in this research. Different training parameters have not been explored since the time constraint arises from hardware constraints. In the future, this work hopes to explore different architecture to induce a comparison, thus reaching a consensus on the best-performing architecture. Furthermore, the input data features can be further improved, and different technical indicators can be tested. Correlation analysis such as the Pearson correlation coefficient can be further conducted to drop correlated features to achieve optimal noise reduction (Benesty et al., 2009). To effectively address each cryptocurrency's market behaviour, models should also be trained on each cryptocurrency's historical data to attain even greater accuracy and adaptability.

ACKNOWLEDGEMENT

This research is supported through the KDU Penang University College Internal Research Grant.

REFERENCES

- Anthony, M., Bartlett, P. L., & Bartlett, P. L. (1999). *Neural network learning: Theoretical foundations*. Cambridge University Press.
- Azulay, A., & Weiss, Y. (2018). Why do deep convolutional networks generalize so poorly to small image transformations? *Journal of Machine Learning*, 20(184), 1-25. <https://doi.org/10.48550/arXiv.1805.12177>
- Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). On the importance of the Pearson correlation coefficient in noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(4), 757-765. <https://doi.org/10.1109/tasl.2008.919072>
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Cartwright, T. J. (1991). Planning and chaos theory. *Journal of the American Planning Association*, 57(1), 44-56. <https://doi.org/10.1080/01944369108975471>
- Chollet, F. (2017). *Deep learning with Python*. Simon and Schuster.
- Cowpertwait, P. S. P., & Metcalfe, A. V. (2009) *Time series data*. In *Introductory time series with R* (pp. 1-25). Springer. https://doi.org/10.1007/978-0-387-88698-5_1
- Dempster, M. A. H., & Romahi, Y. S. (2002). Intraday FX trading: An evolutionary reinforcement learning approach. In H. Yin, N. Allinson, R. Freeman, J. Keane & S. Hubbard (Eds.), *Intelligent Data Engineering and Automated Learning - IDEAL 2002* (pp. 347-358). Springer. https://doi.org/10.1007/3-540-45675-9_52
- Dempster, M. A., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications*, 30(3), 543-552. <https://doi.org/10.1016/j.eswa.2005.10.012>

- Fang, F., Ventre, C., Basios, M., Kong, H., Kanthan, L., Li, L., Martinez-Regoband, D., & Wu, F. (2022). Cryptocurrency trading: A comprehensive survey. *Financial Innovation*, 8(13). <https://doi.org/10.1186/s40854-021-00321-6>
- Ganesh, S., Vadori, N., Xu, M., Zheng, H., Reddy, P., & Veloso, M. (2019). *Reinforcement learning for market making in a multi-agent dealer market*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1911.05892>
- Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37-45). Springer. https://doi.org/10.1007/978-3-642-24797-2_4
- Grondman, I., Busoniu, L., Lopes, G. A., & Babuska, R. (2012). A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1291-1307. <https://doi.org/10.1109/tsmcc.2012.2218595>
- Gronwald, M. (2014). *The economics of bitcoins - Market characteristics and price jumps*. (Working Paper No. 5121). https://www.cesifo.org/DocDL/cesifo1_wp5121.pdf
- Haferkorn, M., & Diaz, J. M. Q. (2014). Seasonality and interconnectivity within cryptocurrencies - An analysis on the basis of bitcoin, litecoin and namecoin. In A. Lugmayr (Ed). *International Workshop on Enterprise Applications and Services in the Finance Industry* (pp. 106-120). Springer. https://doi.org/10.1007/978-3-319-28151-3_8
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2). <https://doi.org/10.5121/ijdkp.2015.5201>
- Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. *Enterprise Information Systems*, 13(1), 132-144. <https://doi.org/10.1080/17517575.2018.1493145>
- Huang, C. Y. (2018). *Financial trading as a game: A deep reinforcement learning approach*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1807.02787>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
- Jeong, G., & Kim, H. Y. (2019). Improving financial trading decisions using deep Q-learning: Predicting the number of shares, action strategies, and transfer learning. *Expert Systems with Applications*, 117, 125-138. <https://doi.org/10.1016/j.eswa.2018.09.036>
- Kolm, P. N., & Ritter, G. (2019). Modern perspectives on reinforcement learning in finance. *Journal of Machine Learning in Finance*, 1(1).
- Konda, V. R., & Tsitsiklis, J. N. (1999). Actor-critic algorithms. In S. Solla, T. Leen & K. Müller (Eds.), *NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems* (pp. 1008-1014). MIT Press.
- Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International Journal of Computer and Information Engineering*, 1(12), 4104-4109. <https://doi.org/10.5281/zenodo.1082415>

- Li, Y. (2017). *Deep reinforcement learning: An overview*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1701.07274>
- Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. *Procedia Computer Science*, 111, 376-381. <https://doi.org/10.1016/j.procs.2017.06.037>
- Liu, X. Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., & Wang, C. (2020). *FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2011.09607>
- Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN-LSTM model for gold price time-series forecasting. *Neural Computing and Applications*, 32, 17351-17360. <https://doi.org/10.1007/s00521-020-04867-x>
- Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Artificial Intelligence for Smart System Simulation, 2020*, Article 6622927 <https://doi.org/10.1155/2020/6622927>
- Lucarelli, G., & Borrotti, M. (2019). A deep reinforcement learning approach for automated cryptocurrency trading. In J. MacIntyre, I. Maglogiannis, L. Iliadis & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 247-258). Springer. https://doi.org/10.1007/978-3-030-19823-7_20
- Moody, J., Saffell, M., Andrew, W. L., Abu-Mostafa, Y. S., LeBaron, B., & Weigend, A. S. (1999). Minimizing downside risk via stochastic dynamic programming. *Computational Finance*, 403-415.
- Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, 60(7), 1772-1791. <https://doi.org/10.1287/mnsc.2013.1838>
- Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. <https://doi.org/10.1109/tkde.2009.191>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/h0042519>
- Sattarov, O., Muminov, A., Lee, C. W., Kang, H. K., Oh, R., Ahn, J., Oh, H. J., & Jeon, H. S. (2020). Recommending cryptocurrency trading points with deep reinforcement learning approach. *Applied Sciences*, 10(4), Article 1506. <https://doi.org/10.3390/app10041506>
- Schulman, J., Moritz, P., Levine, S., Jordan, M., & Abbeel, P. (2015). *High-dimensional continuous control using generalized advantage estimation*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1506.02438>
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1707.06347>
- Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, 7, 53040-53065. <https://doi.org/10.1109/access.2019.2912200>
- Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3), 1464-1468. <https://doi.org/10.1109/23.589532>

- Tucnik, P. (2010). Optimization of automated trading system's interaction with market environment. In P. Forbrig & H. Günther (Eds.), *Perspectives in Business Informatics Research* (pp. 55-61). Springer. https://doi.org/10.1007/978-3-642-16101-8_5
- Van Otterlo, M., & Wiering, M. (2012). Reinforcement learning and Markov decision processes. In M. Wiering & M. Van Otterlo (Eds.), *Reinforcement Learning. Adaptation, Learning, and Optimization* (pp. 3-42). Springer. https://doi.org/10.1007/978-3-642-27645-3_1
- Vrigazova, B. (2021). The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research Journal*, 12(1) 228-242. <https://doi.org/10.2478/bsrj-2021-0015>
- Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018). A new forecasting framework for bitcoin price with LSTM. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 168-175). IEEE Publishing. <https://doi.org/10.1109/icdmw.2018.00032>
- Xiong, Z., Liu, X. Y., Zhong, S., Yang, H., & Walid, A. (2018). *Practical deep reinforcement learning approach for stock trading*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1811.07522>
- Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020, October 15-16). Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the First ACM International Conference on AI in Finance* (pp. 1-8). ACM Publishing. <https://doi.org/10.1145/3383455.3422540>
- Zhang, W., Yang, Z., Shen, J., Liu, M., Huang, Y., Zhang, X., Tang, R., & Li, Z. (2021). Learning to build high-fidelity and robust environment models. In N. Oliver, F. Pérez-Cruz, S. Kramer, J. Read & J. A. Lozano (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 104-121). Springer. https://doi.org/10.1007/978-3-030-86486-6_7
- Zhang, Z., Zhang, Y., & Li, Z. (2018). *Removing the feature correlation effect of multiplicative noise*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1809.07023>



RNA Isolation from Environmental Samples of a Harmful Algal Bloom for Metatranscriptome Next-Generation Sequencing

Diana Lorons¹, Kenneth Francis Rodrigues¹, Madihah Jafar Sidik² and Grace Joy Wei Lie Chin^{1*}

¹Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

²Borneo Marine Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

ABSTRACT

During a harmful algal bloom (HAB), the seawater contains a high abundance of microorganisms and elemental ions. Such components can interfere with RNA isolation, leading to RNA degradation. The complex HAB seawater property makes isolating high-quality RNA for metatranscriptomic sequencing difficult, which is required for effective RNA sequencing and transcriptome profiling. This study used three isolation techniques to find the optimal strategy for isolating total RNA from bloom samples. One of the isolation techniques was the phenol-chloroform extraction method, which uses organic solvents to isolate RNA. The remaining two isolation techniques used the same commercial RNA extraction kit, *TransZol Up Plus* RNA kit (TransGen Biotech, China). One followed the extraction kit's protocol, while the other modified the protocol. Total RNA was extracted from three seawater samples of three occasions of HAB in Sepanggar Bay. The most effective approach used to extract high-quality RNA from the environmental samples of the HABs was the *TransZol Up Plus* RNA kit, with modified protocol. Results of the modified

protocol generated a high-purity total RNA, ranging from 2.081 to 2.474 for both the absorbance ratios $A_{260/280}$ and $A_{260/230}$. The RNA integrity number value ranged from 6.2 to 7.6. All of the samples resulted in concentrations up to 91 ng/ μ l. We concluded that the modified protocol of *TransZol Up Plus* RNA kit yielded the highest quality total RNA for metatranscriptome next-generation sequencing (NGS). Apart from NGS, the high-quality RNA can also be used

ARTICLE INFO

Article history:

Received: 28 October 2021

Accepted: 26 January 2022

Published: 21 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.23>

E-mail addresses:

dianaloronsmorris@gmail.com (Diana Lorons)

kennethr@ums.edu.my (Kenneth Francis Rodrigues)

madihah@ums.edu.my (Madihah Jafar Sidik)

gracejoy@ums.edu.my (Grace Joy Wei Lie Chin)

* Corresponding author

for various downstream applications, including real-time PCR, RNA cloning, and RNA microarray analysis.

Keywords: Harmful algal bloom, next-generation sequencing, RIN number, total RNA extraction

INTRODUCTION

Harmful algal bloom (HAB), also known as “red tide,” is a common phenomenon in coastal waters worldwide. Under certain favourable environmental conditions, a unicellular microalga may proliferate and aggregate to form dense concentrations of cells or “blooms,” resulting in a HAB outbreak. Such an outbreak severely affects the environment and human health and causes substantial economic losses to the aquaculture, fisheries, and tourism industries (Roberts et al., 2020). For example, since the first HAB report of the harmful alga, *Pyrodinium bahamense* var. *compressum*, in 1976, *Pyrodinium* blooms have occurred regularly for over four decades in the coastal waters of western Sabah, Malaysia (Roy, 1977; Anton et al., 2000; Law et al., 2020). Additionally, HABs caused by another harmful alga, *Margalefidinium* (formerly *Cochlodinium*) *polykrikoides*, have occurred every year since its discovery in 2005 (Anton et al., 2008). The former harmful alga produces saxitoxin, a neurotoxin that causes illnesses and fatalities in humans due to consumption of toxin-contaminated seafood (Usup et al., 2012). In contrast, the latter harmful alga causes water discolouration and mass fish mortality due to oxygen deficiency caused by the algae clogging the fish gills (Harun et al., 2015).

HABs in Sabah occur annually, usually from June to July (Northeast Monsoon) and December to January (Southwest Monsoon). The prevailing wind and surface current patterns may have contributed to the bloom’s initiation (Usup et al., 1989). Another factor that contributed to the formation of HAB in Sabah was the accumulation of high nitrogen and phosphorus concentrations in the waters as a result of wastewater and pollutants being discharged from the mainland during heavy rainfall and surface runoff events. This factor helped promote the growth of the dormant harmful algae below the seabed (Anton et al., 2008). Compared to normal seawater, seawater during a HAB event usually comprises 90 % harmful algal species, produces a foul smell, and appears discoloured due to the pigmentation of the bloom-causing microalgae (Barsanti & Gualtieri, 2014).

Metatranscriptomic next-generation sequencing (NGS) is a relatively new field that attempts to investigate the functional properties of a complex environmental community (Jiang et al., 2011). The approach was designed originally to study bacteria and archaea, and its application to eukaryotic microorganisms has yet to be extensively documented (Lin et al., 2012). Furthermore, unlike metagenomics analysis, which provides information about the presence of microbes in an environmental community, the study of the metatranscriptome using NGS techniques provides information about gene activities, including gene diversity,

abundance, and differential gene expression between different conditions (Mutz et al., 2013). Therefore, isolation of total RNA from environmental samples with acceptable quality and intact RNA for effective RNA sequencing and transcriptome profiling is critical for acquiring good transcript data readings for metatranscriptomic analysis (Kukurba & Montgomery, 2015).

RNA quality is defined by the sample's RNA purity and integrity (Becker et al., 2010; Die & Roman, 2012). Total RNA is considered high-quality if it meets the NGS minimum criteria of (a) an absorbance ratio (A260:A280 and A260:230) of more than 1.8, (b) an RNA integrity number (RIN) score of greater than seven, and (c) a minimum concentration of 50 ng/ μ l (Die & Roman, 2012). The ratios of 260 nm to 280 nm and 260 nm to 230 nm absorbances are used to determine the purity of nucleic acids. For RNA, an A260:A280 ratio of approximately 2.0 is considered "pure". Protein, phenol, or other contaminants that absorb heavily at or near 280 nm may be present if the ratio is significantly lower. For the A260:A230 ratio, the value is often higher than the respective A260:A280 value, so the expected value is commonly in the 2.0 to 2.2. If the ratio is significantly lower than expected, it may indicate the presence of contaminants that absorb at 230 nm (Lucena-Agular et al., 2016).

In addition, the integrity of an RNA molecule is critical for investigations attempting to capture a snapshot of gene expression at the time of RNA extraction. The RNA integrity number (RIN) is the gold standard for determining the integrity of an RNA molecule. RIN is an algorithm for assigning integrity values to RNA measurements, traditionally evaluated using the 28S to 18S rRNA ratio (Schroeder et al., 2006). RNA integrity is measured using microcapillary electrophoresis that calculates the RIN value from 1 (fully degraded RNA) to 10 (intact RNA). The recommended RIN value for NGS is 7 to 8, depending on the type of samples (Thompson et al., 2007; Jeffries et al., 2014). The gradual degradation of ribosomal RNA is reflected by a continuous shift towards shorter fragment sizes (Schroeder et al., 2006). Furthermore, the concentration and amount of RNA depend on the type of RNA-seq library preparation method. Typically, the recommended RNA concentration is 10 to 100 ng/ μ l, while the total RNA amount is 5 ng to 10 μ g. It is important because using a more significant amount of starting total RNA can reduce the number of PCR amplification cycles during library preparation, which will result in more even distribution of mapped reads within target genes (Schmid et al., 2012; He & Jiao, 2014).

RNA extraction is frequently employed in molecular biology research to determine the metabolic activity of a sample. The input RNA's concentration and quality significantly influence gene expression consistency (Tavares et al., 2011). As a result, RNA purity and integrity have become the primary criterion for obtaining high-quality output data. It is critical to ensure that the extraction process is thoroughly managed because low-quality RNA can sabotage the results of NGS applications, such as RNA sequencing (Kukurba &

Montgomery, 2015). RNA extraction from a single species specimen is difficult enough, but environmental samples, such as harmful algal bloom (HAB) samples, are more challenging since they contain numerous species. The samples are also surrounded by various nutrients and ions, which can interfere with the RNA extraction process (Wang et al., 2012).

The RNA of HAB species is typically isolated by harvesting the monoclonal microalgal cultures that have reached their mid-exponential growth stage (Palani-Velu et al., 2013). However, RNA extracted from the environmental sample of HAB is unpredictable because HAB seawater contained many contaminants rather than a clean cell culture from *in-vitro* cell culture. Human activities, urbanisation, and industrialisation near the coastal area have resulted in the free flow of macronutrients, such as nitrogen and phosphorus, in abundance in a HAB environmental sample. Other factors like ions, sodium, debris, and suspended particulate matter will also significantly contribute to the instability of the RNA to be extracted from the HAB samples (Tan & Ransangan, 2016; Anderson et al., 2002). Due to some microalgae species in the HAB sample being made up of rigid cell walls, effective cell lysis for RNA isolation and retrieval of unbiased nucleic acid is substantially more challenging (Mäki et al., 2017).

Previous studies have reported using a variety of RNA extraction techniques to extract RNA from environmental samples of HAB for metatranscriptomic sequencing. For example, Cooper et al. (2014) used two different RNA extraction kits: the Ambion RNAqueous extraction kit (Life Technologies) to extract the total RNA of an environmental sample of *Prorocentrum minimum* bloom and the Nucleospin RNA Plant extraction kit (Machery-Nagel) to extraction the total RNA of clonal cultures of *P. minimum*. Meanwhile, total RNA from a *Levanderina fissa* bloom sample was extracted using the Ambion ToTALLY RNA kit (Life Technologies) after the sample was vortex with glass beads (Gong et al., 2017). In China, S. Xu et al. (2021) extracted the total environmental RNA of *Phaeocystis globosa* bloom by using the TRIzol Reagent Kit (Thermo Scientific) and then further concentrated the RNA sample with the RNeasy MinElute Kit (Qiagen). In addition, X. Xu et al. (2020) studied the microbial communities during HAB by extracting the environmental HAB sample using the RNeasy Isolation Kit (Qiagen).

This study compared three different RNA isolation techniques to extract the RNA from the environmental seawater samples during HAB: (a) phenol-chloroform method (Chomczynski & Sacchi, 2006), (b) *Transzol* Up Plus RNA kit (TransGen Biotech, China), and (c) *Transzol* Up Plus RNA kit (TransGen Biotech, China), with minor modification on the protocol. We aim to obtain high-quality RNA that meets the minimum NGS requirements, which include (a) RNA integrity number (RIN) value greater than 6.3, (b) minimum concentration of 50 ng/μl, (c) total RNA amount of greater than 1.3 μg, (d) absorbance ratio (A_{260}/A_{280} ; A_{260}/A_{230}) of more than 1.8, and (e) two distinct intact 18S and 28S ribosomal RNA bands. Although there are other means for extracting RNA from

Table 1
Details of harmful algal bloom samples used in the study

Bloom Sample	Month, year	Cell Count (cells/l)	Species	Coordinates
BE1	November 2018	4.5×10^2	<i>Margalefidinium polykrikoides</i>	06°02'25.44" N 116°06'33.84" E
BE2	July 2019	1.2×10^2	<i>Margalefidinium polykrikoides</i>	06°02'51.9" N 116°05'57.6" E
BE3	October 2019	10.0×10^2	<i>Margalefidinium polykrikoides</i>	06°05'24.5" N 116°07'30.7" E
		4.2×10^2	<i>Pyrodinium bahamense</i> var. <i>compressum</i>	

Seawater Sample Preparation

The collected HAB seawater samples were promptly transported to the laboratory for filtration, which had to be completed within 24 hours of sampling. Cellulose nitrate membrane filter paper (0.45 μm) (Whatman, United Kingdom) was used to filter the two litres of water sample to retain eukaryotes. Each filter was individually wrapped in aluminium foil to minimize repeated thawing of the sample during RNA extraction and cross-contamination. The wrapped samples were placed in resealable bags and stored in a $-80\text{ }^\circ\text{C}$ freezer (Thermo Fisher Scientific, USA). Right before RNA extraction, filters were briefly thawed on ice. Filtered samples of the same bloom event were placed in a 50 ml centrifuge tube and rinsed with sterile seawater. A total of five pieces of membrane filter paper were rinsed into the 50 ml centrifuge tube. The remaining sample on the cellulose nitrate membrane, which did not rinse off, was scrapped lightly to ensure all samples were ultimately rinsed into the falcon tube. The samples were then pelleted using a refrigerated centrifuge (Eppendorf, Germany) at $1,085 \times g$ for 5 min at $4\text{ }^\circ\text{C}$. After discarding the supernatant, the pelleted samples were ready for RNA extraction.

Total RNA Isolation Using Phenol-Chloroform Extraction Method (Chomczynski & Sacchi, 2006)

The complete procedure of the phenol-chloroform extraction method can be referred to in the original publication by Chomczynski and Sacchi (2006). After discarding the supernatant, the pelleted cells were lysed and homogenised with 1 ml of Solution D (Denaturing solution). Solution D contains 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5 % (w/v) *N*-laurosylsarcosine (sarkosyl) and 0.1 M 2-mercaptoethanol. The lysate was resuspended at least ten times with gentle pipetting, and the cell lysate was placed in a 4-ml polypropylene tube. In that order, 0.1 ml of 2 M sodium acetate (pH 4.0), 1 ml saturated phenol solution, and 0.2 ml chloroform/isoamyl alcohol (49:1) were added to the lysate. Before adding the chloroform/isoamyl alcohol, the tube was gently inverted

several times by hand for 10 s after adding each of the previously stated solutions. After thoroughly mixing, the tube was chilled on ice for 15 min before centrifuging at $10,000 \times g$ for 20 min at 4°C using a refrigerated centrifuge (Eppendorf, Germany).

Upon centrifugation, the solution in the tube was partitioned into three different phases: a lower cloudy white organic phase, pale brownish interphase and an upper colourless aqueous phase containing the RNA. The upper aqueous layer was transferred into a fresh 1.5 ml microcentrifuge tube before adding 1 ml of ice-cold isopropanol solution for RNA precipitation. The tube was incubated in a -20°C freezer (Thermo Fisher Scientific, USA) for 1 h, followed by centrifugation at $10,000 \times g$ for 20 min at 4°C . The supernatant was discarded, and the RNA pellet was dissolved in 0.3 ml of solution D before being transferred to a new 1.5 ml microcentrifuge tube. A total of 0.3 ml of isopropanol solution was added into the tube, followed by incubation at -20°C in a freezer (Thermo Fisher Scientific, USA) for 30 min. Next, the mixture was centrifuged at $10,000 \times g$ for 10 min at 4°C . After the supernatant was discarded, the RNA pellet was washed with 1 ml of 75 % ethanol solution. The RNA was gently mixed and incubated for 15 min at room temperature to completely dissolve any remaining residual salt. Next, the tube was centrifuged at $10,000 \times g$ for 5 min at 4°C , and the supernatant was discarded. The RNA pellet was air-dried at room temperature for 10 min before being dissolved in 150 μl of diethylpyrocarbonate (DEPC)-treated water. The sample was then incubated for 15 min at 60°C before storing the RNA sample in a -80°C freezer (Thermo Fischer Scientific, USA).

Total RNA Isolation Using *TransZol Up Plus* RNA Kit (TransGen Biotech, China)

The *TransZol Up Plus* RNA kit was used to isolate the total RNA content from the HAB environmental samples according to instructions provided by the manufacturer. The cell pellet was added with 1 ml of *TransZol Up* solution and gently pipetted until there were no visible residues in the lysate. After 5 minutes of incubation at room temperature, 0.2 ml of chloroform was added to the lysate according to the ratio (0.2 ml: 1 ml *TransZol Up*). The lysate was shaken vigorously for 30 s and incubated again for 3 min at room temperature before centrifuging at $10,000 \times g$ for 15 min at 4°C using a refrigerated centrifuge (Eppendorf, Germany). Three distinct phases of the lower pink organic phase, interphase, and a colourless upper aqueous phase containing the RNA were observed. The colourless, upper phase containing RNA was transferred into a fresh RNase-free tube, and an equal volume of 100% ethanol (25 μl) was added to the tube; the tube was inverted gently to ensure it mixed thoroughly.

All centrifugations in the following steps were carried out at $12,000 \times g$ at room temperature using a refrigerated centrifuge (Eppendorf, Germany). Once mixed, the solution and precipitates were transferred into a spin column prior to centrifugation for 30 s, and the flow-through was discarded. A total of 500 μl of CB9 (Clean Buffer 9) was added into

the spin column and centrifuged for 30 s, and the flow-through was discarded. Next, the previous step was repeated, adding 500 µl of WB9 (Wash Buffer 9) into the spin column and centrifuging for 30 s. After that, the spin column was centrifuged for 2 min to remove any remaining ethanol. Once completed, the spin column was transferred to a clean 1.5 ml microcentrifuge tube. In the spin column, 50 µl of RNase-free water was added and incubated at room temperature for 1 min. Finally, the same tube was centrifuged for 1 min to elute the RNA and immediately stored in a -80 °C freezer (Thermo Fisher Scientific, USA) to avoid RNA degradation.

Total RNA Isolation Using a Modified Protocol of *TransZol Up Plus* RNA Kit (TransGen Biotech, China)

Minor modifications were made to the original protocol of the *TransZol Up Plus* RNA kit *TransZol Up Plus* RNA kit described by the manufacturer to improve RNA yield and quality. The detailed steps modified from the original protocol are listed in Table 2.

Table 2

Summary of modifications made to the original protocol of *TransZol Up Plus* RNA Kit

Components/Steps	Original Step	Modified Step
<i>TransZol Up</i> solution (Cell lysis)	1 ml <i>TransZol Up</i> solution was added and mixed by gentle pipetting until no residues were visible in the lysate.	If the sample still appeared saturated after adding the initial <i>TransZol Up</i> solution, additional 1 ml of <i>TransZol Up</i> solution was added to ensure complete lysis of the sample.
Clean Buffer 9 (CB9)	The washing step using CB9 was performed twice.	The washing step using CB9 was performed four times, where CB9 was re-added, and the sample was re-centrifuged four times.
Wash Buffer 9 (WB9)	The washing step using WB9 was performed twice.	The washing step using WB9 was performed four times, where WB9 was re-added, and the sample was re-centrifuged four times.
Washing step	CB9 and WB9 solutions were added directly to the bottom of the spin column.	CB9 and WB9 solutions were added entirely to the whole spin column.
RNA precipitation step	The RNA sample was precipitated for 30 minutes before the RNA elution step.	The RNA sample was precipitated for 2 hours before the RNA elution step.
RNA elution step	The RNA sample was directly eluted using RNase-free water.	The spin column was air-dried for 10 minutes to remove the excess ethanol before the RNase-free water was added for RNA elution.

RNA Quality Assessment

The total RNA's initial integrity and overall quality were observed with 1.5 % (w/v) agarose gel electrophoresis by inspection of the 28S and 18S rRNA bands. The extracted RNA sample was visually analysed following electrophoresis conducted at 75 V for 45 min, with

a 1 kb GeneRuler 1kb DNA ladder (Thermo Fisher Scientific, USA). The concentration (absorbance at 260 nm) and purity (absorbance $A_{260/280}$ and $A_{260/230}$) of the RNA were measured with a Nanodrop™ spectrophotometer (Thermo Fisher Scientific, USA). The RNA amount was measured using Qubit® 2.0 fluorometer (Life Technologies, USA). The RNA integrity number (RIN) was analysed with Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Table 3 summarises the required criteria for NGS RNA-seq for the eukaryotic sample.

Table 3

Criteria required for next-generation sequencing RNA-seq for microalgal sample: RNA concentration, RNA amount, RNA purity, RNA integrity number (RIN) and ribosomal RNA integrity on agarose gel

Sample Type	RNA Concentration (He & Jiao, 2014)	RNA Amount (He & Jiao, 2014)	RNA Purity (Die & Roman, 2012)	RNA Integrity Number (RIN) (Schroeder et al., 2006)	Ribosomal RNA (rRNA) (Die & Roman, 2012)
Total RNA (Microalgae)	≥ 50 ng/μl	≥ 1.3 μg	$A_{260/280} = 1.8-2.2$ $A_{260/230} ≥ 2.0$	≥ 6.3 smooth base line	Two intact distinct rRNA bands

RESULTS AND DISCUSSION

The study area, Sepanggar Bay, Kota Kinabalu, is located on the south-eastern coast of the South China Sea and in the northwest part of Sabah, Malaysia (Figure 1). Sepanggar Bay receives freshwater inflows from the Inanam and Menggatal rivers and domestic sewage and factory wastes from an industrial area further inland (Anton et al., 2008). One of the eight international seaports in Sabah is located at Sepanggar Bay, where the port manages a few million tonnes of cargo annually (Goh & Lee, 2010). Besides that, resorts, coastal housing, and water villages can be found along the bay area. In addition, the bay supports fisheries' resources by providing farmed and fished seafood (Gallagher et al., 2016).

Three seawater samplings of bloom occurrences were conducted during the study period (October 2018 to November 2019), as shown in Table 1. *M. polykrikoides* was the primary cause of the first two HAB outbreaks, a dinoflagellate that clogs and damages the gills of wild and farmed fish, resulting in mass mortality due to lack of oxygen (Palani-Velu et al., 2013). Two harmful dinoflagellates, *M. polykrikoides* and *P. bahamense* var. *compressum*, were responsible for this study's most recent HAB event. *Pyrodinium bahamense* is a dinoflagellate that produces saxitoxin, a neurotoxin that causes paralytic shellfish poisoning in humans after consuming the toxin-contaminated shellfish (Jipanin et al., 2019). The HAB poses a severe threat to the Sabah's economy because of the state's dependency on marine and agricultural resources. Marine aquaculture accounts for more than 43% by volume and 73% by value of total aquaculture production in Sabah, equating to 11,846 metric tonnes valued at around USD\$ 50 million (Liaw & Fung, 2000; Anton et al., 2008).

In Malaysia, *P. bahamense* var. *compressum* blooms were first documented in 1976 in Sabah, where the whole stretch of the west coast of Sabah, including Sepanggar Bay, was affected (Roy, 1977). Since then, *P. bahamense* bloom has been an annual occurrence and can be observed along the coastal waters of west Sabah, not exclusively at Sepanggar Bay (Usup et al., 1989; Usup & Azanza, 1998; Leaw et al., 2005; Chin et al., 2013; Anton et al., 2008; Law et al., 2020). Aside from *P. bahamense*, another two causative species were also identified, which were *M. polykrikoides* (Anton et al., 2008) and *Gymnodinium catenatum* (Mohammad-Noor et al., 2010). The *G. catenatum* and *M. polykrikoides* were first observed in Sepanggar Bay in 2003 and 2005, respectively, though no blooms of the former HAB species have been reported since then. Since the *M. polykrikoides* bloomed in 2005, the *P. bahamense* bloom has not occurred again until 2013 (Law et al., 2020). Blooms of *P. bahamense* and *M. polykrikoides* have frequently occurred solely or co-occurred since then. Understanding the time of occurrence and factors contributing to the HAB occurrence is still unknown.

The study initially performed the total RNA extraction using the phenol-chloroform method developed by Chomczynski and Sacchi (2006). Unfortunately, RNA extracted from HAB seawater samples using this method was inefficient, and the resultant RNA often appeared to be degraded or absent, as seen in Figure 2(a). The agarose gel photo in Figure 2(a) shows traces of degraded RNA in Lane 1 and no RNA in Lanes 2 and 3. Unsuccessful extraction of RNA by using this similar approach was also observed in *M. polykrikoides* monoalgal cultures, where RNA was frequently degraded or absent (Palani-Velu et al., 2013). RNA with low yield was obtained only a few times throughout the many trials performed by Palani-Velu et al. (2013), where barely visible large subunit (LSU) and small subunit (SSU) rRNA bands were observed on the agarose gel. The method failed maybe because the method was not suitable to eliminate the numerous impurities in the environmental samples, which were also accompanied by a high abundance of eukaryotic and prokaryotic organisms. Even though the phenol-chloroform method was unsuccessful for this study, this method is still widely used because of its high yield, purity, and speed, which makes it ideal for working with small numbers of samples (Shin, 2012).

After multiple failed attempts to extract total RNA using the phenol-chloroform extraction method (Chomczynski & Sacchi, 2006), a manufactured RNA extraction kit, *TransZol* Up Plus RNA Kit (TransGen Biotech, China), was used to extract total RNA from the HAB environmental samples. The *TransZol* Up Plus RNA kit is a ready-to-use reagent for isolating total RNA from cells and tissues. The silica-based spin column not only holds the RNA together but also prevents a significant volume of RNA from being spun out from the column, yielding a higher volume of RNA during the elution step (Karp et al., 1998). Attempts to extract RNA following the manufacturer's exact protocol have resulted in degraded and no visible RNA, as shown in Figure 2(b). However, in subsequent trials,

the RNA kit successfully produced two intact RNA bands with low yield, and some RNA degradation was observed at the bottom of the agarose gel [Figure 2(c)]. Figure 2(c) results showed that the RNA kit has the potential to extract RNA from the HAB environmental seawater samples; however, some modifications to the original protocol are required to improve the chances of isolating high-quality total RNA content.

The *TransZol Up Plus* RNA kit can be used on any organism, including humans, animals, plants, and bacteria. Therefore, it is not limited to a particular type of sample to be extracted. However, the main concern would be a mixture of organisms in one extraction process. Therefore, it is advisable to take several measures to change the protocols to suit the sample type better. HAB seawater environmental sample contains a combination of ions, planktons (phytoplankton, zooplankton), bacteria, and other contaminants or macronutrients. Therefore, HAB seawater samples' properties are vastly different compared to the cell-cultured samples, where the parameters of the cell cultures can be manipulated and controlled. Other factors such as excessive ions and pH might also become a restriction to yielding a good quality RNA (Tan & Ransangan, 2016; Anderson et al., 2002). As a result, modifications were focused on thoroughly removing all the impurities in the HAB seawater samples, ensuring a good quality RNA was obtained at the end of the extraction method. After several modifications to the *TransZol Up Plus* RNA kit (Table 2), total RNA was effectively recovered from the HAB environmental samples.

The successfully isolated RNA samples were of high quality and intact, as evidenced by visual observation of the ethidium bromide post-stained electrophoresed RNA samples (Figure 2d). Two distinct RNA bands were observed in Figure 2(d), representing the LSU rRNA and SSU rRNA, along with minor smearing at the lower molecular weight range. The LSU and SSU rRNAs were consistently observed in the RNA extracted from all HAB environmental seawater samples. The expected sizes of the rRNAs were 2,000 bp for SSU and 2,500 bp for LSU. However, based on the gel figure (Figure 2d), the sizes of LSU and SSU rRNAs observed were 750 bp and 500 bp, respectively. The size difference is because, in native agarose gels, the secondary structure of RNA affects its migration pattern, causing it to travel differently than its actual size. Native agarose gels separate RNA molecules not only by their size but also by the shape of their native structure (Wu & Tinoco, 1998; Buchmueller & Weeks, 2003). It was recommended to use a denaturing agarose gel system for better RNA quality assessment, as the denaturing gels run under the conditions that cause the natural structure of RNA to be disrupted, resulting in the formation of linear chains (Woodson & Koculi, 2009). In this study, the agarose gel electrophoresis was sufficient to judge the integrity and overall quality of the total RNA preparation by observing the SSU and LSU bands.

Due to the nature of the HAB seawater samples, the modifications made to the original protocol were to eliminate contaminants or impurities that would co-purified

together with the RNA. The modifications would therefore improve the possibilities of isolating high-quality RNA from the environmental seawater sample to be suitable for metatranscriptomic sequencing. Tan and Yiap (2009) documented that the target nucleic acid should be segregated from the contaminants, including protein, carbohydrate, lipids, or other nucleic acids. As a result, special attention was paid to the washing step of the protocol, in which RNA will bind to the silica. At the same time, impurities such as DNA, cellular proteins, ions, and metabolites are washed away by using the washing buffers, CB9 and WB9. The samples were washed twice with the washing buffers, yielding two RNA bands, but significant RNA degradation was observed on the agarose gel (Figure 2c). When the RNA quality was checked, the $A_{260/280}$ and $A_{260/230}$ ratios showed that all samples had a value less than 1.8. Following that, the number of washings was increased so that it was in stoichiometry with the volume of contaminants or impurities in the samples. The samples washed four times gave positive results (Figure 2d). The additional washings ensured that the RNA was cleaned thoroughly, and all the impurities were removed entirely through the flow-through of the spin column. In addition, when the washing solutions were added to the spin column, the solutions were added to the entire column rather than just the bottom of the spin column. This action was taken to ensure that all impurities on the interior wall of the column were also washed away, leaving no impurities in the column to be eluted out along with the RNA.

Other minor changes were also carried out to increase the extracted RNA's quality further. One of the significant changes was the extra addition of 1 ml *TransZol* Up solution when the sample lysate still appeared saturated due to incomplete lysis. The additional *TransZol* Up solution, a lysis solution, was necessary due to the high quantity of microorganisms in the HAB seawater samples. For example, sample BE3 (Table 1) contained a higher concentration of harmful algae. Also, it consisted of two types of microalgae, *M. polykrikoides* and *P. bahamense*, which required the extra addition of the lysis solution for the complete lysis of the sample. The second lysis solution would guarantee complete lysis of all the cells present in the samples. Complete cell lysis is essential for releasing RNA from the samples, where insufficient lysis will lead to low RNA yields (Islam et al., 2017).

Besides that, the incubation period during RNA precipitation by using absolute ethanol was also determined. The protocol of the RNA kit did not provide a specific time range for the RNA precipitation step. Usually, precipitation of RNA requires several hours to overnight, depending on the concentration (Green & Sambrook, 2020). Since the kit's protocol stated that the whole RNA extraction process could be completed within an hour, it was concluded that the RNA should only be precipitated for 10 to 30 minutes. The precipitation was initially carried out for 30 minutes; however, the RNA yield obtained was very low. Therefore, the precipitation incubation period was optimised, and the study

found that 2 hours of incubation time was sufficient to provide the needed RNA output. In the RNA precipitation process, the RNA needs to be recovered from the aqueous phase using alcohol; usually, isopropanol is preferred for RNA precipitation (Sambrook & Russell, 2006). Kaser et al. (2009) implied that there are different precipitation periods for different samples, and the optimal precipitation period for a given sample will succeed in producing the two rRNA bands. Complete precipitation will help retain and pool all the RNA molecules together, resulting in a much higher volume, which will reflect the quality of the samples (Johnson, 2012).

The last change on the kit's protocol was to air-dry the spin column before the elution step. This change ensured that any residual ethanol in the spin column would evaporate, preventing ethanol carryover during RNA elution. However, ethanol carryover could lead to the low performance of RNA in downstream applications. In addition, the presence of ethanol may interfere during library preparation, resulting in lower library concentrations than the acceptable threshold (Dehghani et al., 2019). Lastly, throughout the study, a precaution was taken to ensure that all reagents, such as chloroform and absolute ethanol (which were not included in the kit) and consumables, such as microcentrifuge tubes and micropipette tips used, were RNase-free.

After successfully obtaining two distinctive bands of 18S and 28S rRNA in the RNA extracted from all three bloom samples using the modified protocol of the *TransZol Up Plus* RNA kit, the quality of the RNA samples was evaluated. The quality of the extracted RNA was analysed using a UV-Vis spectrophotometer, showing a total RNA concentration of between 72 ng/ μ l to 91 ng/ μ l, with the RNA purity $A_{260/280}$ of 2.16 on average and $A_{260/230}$ of 2.08 to 2.47. The quality of the extracted RNA was further verified using an automated gel electrophoresis system, demonstrating a RIN ranging from 6.2 to 7.6 (Figure 3), with RNA amounts of 1.1 μ g to 2.1 μ g. The bioanalyzer also validated the RNA integrity by observing two peaks showing the 28S and 18S rRNA. Table 4 summarises the quality of the RNA extracted from the three HAB events using the modified protocol of the *TransZol Up Plus* RNA kit. The quality check of all three RNA samples fulfilled the requirements needed for metatranscriptomic next-generation sequencing stated in Table 3. According to Jiang et al. (2011), the RIN number should be the primary gold standard for ensuring the success of the sequencing.

Besides producing good-quality RNA, the *TransZol Up Plus* RNA kit was also preferred due to the shorter extraction time and the availability of a robust lysis solution (*TransZol Up* solution), which rapidly disrupts the cells without the help of a homogeniser. Therefore, thorough, rapid homogenisation of the sample and completing the RNA isolation process as quickly as possible to prevent further sample degradation is crucial in achieving successful RNA purification.

The modified protocol of the *TransZol Up Plus* RNA kit was optimised for extracting RNA from environmental samples of harmful algal bloom. Because the modifications to

the protocol were focused on eliminating contaminants and impurities from the environmental samples, the optimised RNA isolation technique may be useful for other environmental samples that are highly contaminated, such as soil samples and polluted seawater samples. Besides that, the optimised *TransZol Up Plus* RNA kit can also be applied to water-based environmental samples, like lake and river water samples. However, further optimisation may be needed to be carried out on the modified RNA isolation technique to suit the type of environmental samples.

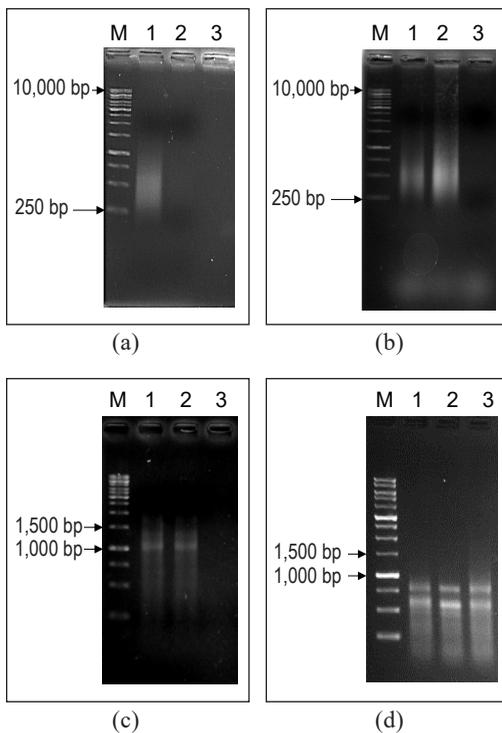
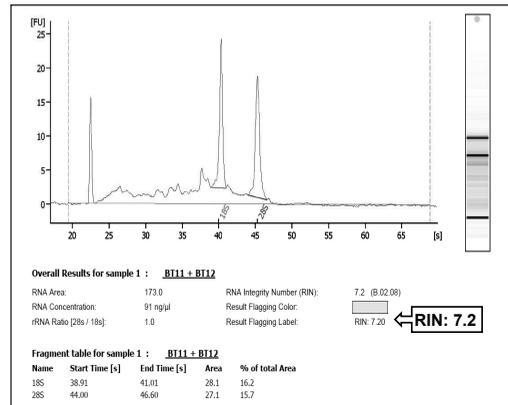
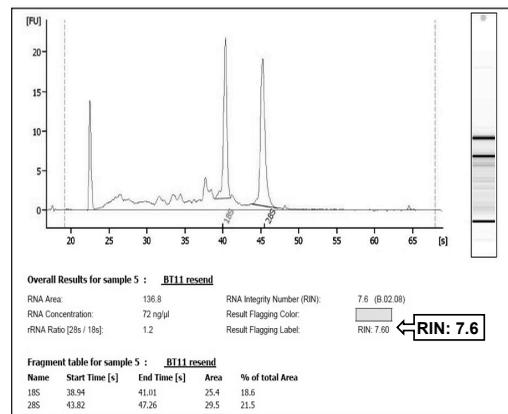


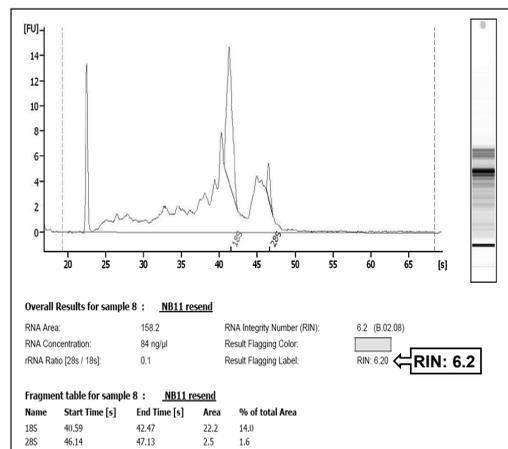
Figure 2. Electrophoresis of total RNA sample on 1.5 % agarose. (a) Phenol-chloroform method, (b) *TransZol Up Plus* RNA kit (Initial attempts), (c) *TransZol Up* RNA extraction kit (Subsequent attempts), (d) *TransZol Up Plus* RNA kit with modifications. Lane M: 1-kb DNA ladder, Lane 1-3: BE1, BE2, BE3.



(a)



(b)



(c)

Figure 3. Profile image of intact 28S rRNA, 18S rRNA, and RIN of RNA sample extracted from three different bloom events, (a) BE1, (b) BE2, and (c) BE3, using the modified protocol of the *TransZol Up Plus* RNA kit.

Table 4

Quality assessment of RNA samples extracted using the modified protocol of the TransZol Up Plus RNA kit

HAB Environmental Sample	Nanodrop			Qubit	Bioanalyzer	
	Conc. (µg/ ul)	A _{260/280}	A _{260/230}	Volume (µl)	Total Amount (µg)	RIN
BE1	91	2.161	2.081	25	1.1	7.2
BE2	72	2.198	2.474	25	1.8	7.6
BE3	84	2.114	2.241	25	2.1	6.2

CONCLUSION

The first step in metatranscriptome sequencing is RNA isolation from an environmental sample. The focus of this study was to establish a suitable protocol for total RNA isolation from seawater environmental samples during a harmful algal bloom occurrence. The modified *TransZol Up Plus* RNA kit proved the most effective in providing reproducible, good-quality, and intact RNA that satisfied the standard criteria for RNA seq-NGS. The extracted RNA is of sufficient quality to produce a library, ensuring a successful RNA-seq experiment. In addition, the total RNA extracted in this study can also be used for other analytical analyses, such as gene expression profiling, cDNA synthesis, and differential expression study.

ACKNOWLEDGEMENT

Malaysian Ministry of Higher Education funded this project through the Fundamental Research Grant Scheme (FRGS/1/2017/WAB09/UMS/03/1).

REFERENCES

- Anderson, D. M., Gilbert, P. M., & Burkholder, J. M. (2002). Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries*, 25, 704-726. <https://doi.org/10.1007/BF02804901>
- Anton, A., Alexander, J., & Estim, A. (2000, February 7-11). Harmful algal blooms in Malaysia: Revisiting Kimanis Bay. In *9th International conference on toxic Phytoplankton* (pp. 7-11). Tasmania, Australia.
- Anton, A., Teoh, P. L., Mohd-Shaleh, S. R., & Mohammad-Noor, N. (2008). First occurrence of *Cochlodinium* blooms in Sabah Malaysia. *Harmful Algae*, 7(3), 331-336. <http://doi.org/10.1016/j.hal.2007.12.013>
- Barsanti, L., & Gualtieri, P. (2014). *Algae: anatomy, biochemistry, and biotechnology*. CRC Press. <https://doi.org/10.1201/b16544>
- Becker, C., Hammerle-Fickinger, A., Riedmaier, I., & Pfaffl, M. W. (2010). mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, 50(4), 237-243. <https://doi.org/10.1016/j.ymeth.2010.01.010>
- Buchmueller, K. L., & Weeks, K. M. (2003). Near native structure in an RNA collapsed state. *Biochemistry*, 42(47), 13869-13878. <https://doi.org/10.1021/bi035476k>

- Chin, G. J. W. L., Teoh, P. L., Kumar, S. V., & Anton, A. (2013). Ribosomal DNA analysis of marine microbes associated with toxin-producing *Pyrodinium bahamense* var. *compressum* (Bohm), a harmful algal bloom species. *Pertanika Journal of Tropical Agricultural Science*, 36(2), 179-188.
- Chomczynski, P., & Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: Twenty-something years on. *Nature Protocols*, 1, 581-585. <https://doi.org/10.1038/nprot.2006.83>
- Cooper, E. D., Bentlage, B., Gibbons, T. R., Bachvaroff, T. R., & Delwiche, C. F. (2014). Metatranscriptome profiling of a harmful algal bloom. *Harmful Algae*, 37, 75-83. <https://doi.org/10.1016/j.hal.2014.04.016>
- Dehghani, M., Rosenblatt, K. P., Li, L., Rakhade, M., & Amato, R. J. (2019). Validation and clinical applications of a comprehensive next generation sequencing system for molecular characterization of solid cancer tissues. *Frontiers in Molecular Biosciences*, 6, 82. <https://doi.org/10.3389/fmolb.2019.00082>
- Die, J. V., & Roman, B. (2012). RNA quality assessment: A view from plant qPCR studies. *Journal of Experimental Botany*, 63(17), 6069-6077. <https://doi.org/10.1093/jxb/ers276>
- Gallagher, J. B., Hoe, C. C., Yusob, M. S. B. M., Gen, C. N., & Mae, G. Y. (2016). Surface chlorophyll patchiness across Sepanggar Bay: Relationships with turbidity and depth. *Transactions on Science and Technology*, 3(2), 421-426.
- Goh, C. S., & Lee, K. T. (2010). A visionary and conceptual macroalgae-based third-generation bioethanol (TGB) biorefinery in Sabah, Malaysia as an underlay for renewable and sustainable development. *Renewable and Sustainable Energy Reviews*, 14(2), 842-848. <https://doi.org/10.1016/j.rser.2009.10.001>
- Gong, W., Browne, J., Hall, N., Schruth, D., Paerl, H., & Marchetti, A. (2017). Molecular insights into a dinoflagellate bloom. *The ISME Journal*, 11, 439-452. <https://doi.org/10.1038/ismej.2016.129>
- Green, M. R., & Sambrook, J. (2020). Precipitation of RNA with ethanol. *Cold Spring Harbor Protocols*, 89-91. <https://doi.org/10.1101/pdb.prot101717>
- Harun, S. N., Mohammad-Noor, N., Ahmad, Z., Chu, K. B., Saad, S., Mohamad Hidayat, N. S., & MuKai, Y. (2015). First report of *Cochlodinium polykrikoides* (Dinophyceae), a harmful algal bloom (HAB) species in the coastal waters of peninsular Malaysia. *Malaysian Journal of Science*, 34(1), 87-92. <https://doi.org/10.22452/mjs.vol34no1.9>
- He, J., & Jiao, Y. (2014) Next-generation sequencing applied to flower development: RNA-Seq. In J. Riechmann & F. Wellmer (Eds.), *Flower Development: Methods and protocols* (pp. 401-411). Humana Press. https://doi.org/10.1007/978-1-4614-9408-9_23
- Islam, M. S., Aryasomayajula, A., & Selvaganapathy, P. R. (2017). A review on macroscale and microscale cell lysis methods. *Micromachones*, 8(3), 83. <https://doi.org/10.3390/mi8030083>
- Jeffries, S., Kiss, M. K., Smith, A. J., & Smith, A. W. (2014). A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples. *BMC Biotechnology*, 14, Article 94. <https://doi.org/10.1186/s12896-014-0094-8>
- Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Gingeras, T. R., & Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, 21, 1543-1551. <https://doi.org/10.1101/gr.121095.111>

- Jipanin, S. J., Muhamad-Shaleh, S. R., Lim, P. T., Leaw, C. P., & Mustapha, S. (2019). The monitoring of harmful algae blooms in Sabah, Malaysia. *Journal of Physics: Conference Series*, 1358, Article 012014. <https://doi.org/10.1088/1742-6596/1358/1/012014>
- Johnson, M. (2012). RNA extraction. *Methods*, 2, 201. <https://doi.org/10.13070/mm.en.2.201>
- Karp, A., Isaac, P. G., & Ingram, D. S. (1998). Isolation of nucleic acids using silica-gel based membranes: Methods based on the use of QIAamp Spin Columns. In A. Karp, P. G. Isaac & D. S. Ingram (Eds.), *Molecular Tools for Screening Biodiversity* (pp. 59-63). Springer. https://doi.org/10.1007/978-94-009-0019-6_14
- Kaser, M., Ruf, M. T., Hauser, J., Marsollier, L., & Pluschke, G. (2009). Optimized method for preparation of DNA from pathogenic and environmental mycobacteria. *Applied and Environmental Microbiology*, 75(2), 414-418. <https://doi.org/10.1128/AEM.01358-08>
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 951-969. <https://doi.org/10.1101/pdb.top084970>
- Law, S. V., Rodrigues, K. F., Jani, J., Anton, A., & Chin, G. J. W. L. (2020). A metagenomic study of bacterial communities associated with the saxitoxin producing dinoflagellate, *Pyrodinium bahamense* var. *compressum*. *Malaysian Journal of Microbiology*, 16(3), 176-183. <https://doi.org/10.21161/mjm.190548>
- Leaw, C. P., Lim, P. T., Ng, B. K., Cheah, M. Y., Ahmad, A., & Usup, G. (2005). Phylogenetic analysis of *Alexandrium* species and *Pyrodinium bahamense* (Dinophyceae) based on theca morphology and nuclear ribosomal gene sequence. *Phycologia*, 44(5), 550-565.
- Liaw, J., & Fung, S., (2000). Sustainable marine aquaculture in Sabah. *Berita IDS*, 15, 1-2.
- Lin, X., Mckinley, J., Resch, C. T., Kaluzny, R., Lauber, R. L., Fredrickson, J., Knight, R., & Konopka, A. (2012). Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *The ISME Journal*, 6, 1665-1676. <https://doi.org/10.1038/ismej.2012.26>
- Lucena-Aguilar, G., Sanchez-Lopez, A. M., Barberan-Aceituno, C., Carrillo-Avila, J. A., Lopez-Guerrero, J. A., & Aguilar-Quesada, R. (2016). DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreservation and Biobanking*, 14(4), 264-270. <https://doi.org/10.1089/bio.2015.0064>
- Mäki, A., Salmi, P., Mikkonen, A., Kremp A., & Tiirola, M. (2017). Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Frontiers in Microbiology*, 8, Article 1848. <https://doi.org/10.3389/fmicb.2017.01848>
- Mohammad-Noor, N., Adam, A., Franco Soler, J. M., Anton, A., & Sitti-Rachanah, M. S. (2010). First record of toxic *Gymnodinium catenatum* off the west coast of Sabah, Malaysia. In K. C. Ho, M. J. Zhou & Y. Z. Qi (Eds.), *Proceedings of 13th International Conference on Harmful Algae* (p. 25). International Society for the Study of Harmful Algae.
- Mutz, K. O., Heilkenbrinker, A., Lonne, M., Walter, J. L., & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology*, 24(1), 22-30. <https://doi.org/10.1016/j.copbio.2012.09.004>

- Palani-Velu, K. D., Chin, G. J. W. L., Rodrigues, K. F., & Anton, A. (2013). Optimization of total RNA isolation from *Cochlodinium polykrikoides*. *Journal of Applied Phycology*, 25, 59-64. <https://doi.org/10.1007/s10811-012-9838-9>
- Roberts, V. A., Vigar, M., Backer, L., Veytsel, G. E., Hilborn, E. D., Hamelin, E. I., Vanden Esschert, K. L., Lively, J. Y., Cope, J. R., Hlavsa, E. I., & Yoder, J. S. (2020). Surveillance for harmful algal bloom events and associated human and animal illnesses - One health harmful algal bloom system, United States, 2016-2018. *Morbidity and Mortality Weekly Report*, 69(50), 1889-1894. <https://doi.org/10.15585/mmwr.mm6950a2>
- Roy, R. N. (1977). Red tide and outbreak of paralytic shellfish poisoning in Sabah. *Medical Journal of Malaysia*, 31(3), 247-251.
- Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols*, (1), pdb-prot4455. <https://doi.org/10.1101/pdb.prot4455>
- Schmid, M. W., Schmidt, A., Klostermeier, U. C., Barann, M., Rosenstiel, P., & Grossniklaus, U. (2012). A powerful method for transcriptional profiling of specific cell types in eukaryotes: Laser-assisted microdissection and RNA sequencing. *PLoS One*, 7(1), Article e29685. <https://doi.org/10.1371/journal.pone.0029685>
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7, Article 3. <https://doi.org/10.1186/1471-2199-7-3>
- Shin, J. H. (2012). Nucleic acid extraction techniques. In Y. W. Tang & C. Stratton (Eds.), *Advanced Techniques in Diagnostic Microbiology* (pp. 209-225). Springer. https://doi.org/10.1007/978-1-4614-3970-7_11
- Tan, K. S., & Ransangan, J. (2016). Effects of environmental conditions and nutrients on the occurrence and distribution of potentially harmful phytoplankton in mesotrophic water. *Sains Malaysiana*, 45(6), 865-877.
- Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: The past and the present. *BioMed Research International*, 2009, Article 574398. <https://doi.org/10.1155/2009/574398>
- Tavares, L., Alves, P. M., Ferreira, R. B., & Santos, C. N. (2011). Comparison of different methods for DNA-free RNA isolation from SK-N-MC neuroblastoma. *BMC Research Notes*, 4, Article 3. <https://doi.org/10.1186/1756-0500-4-3>
- Thompson, K. L., Pine, P. S., Rosenzweig, B. A., Turpaz, Y., & Retief, J. (2007). Characterization of the effect of sample quality on high density oligonucleotide microarray data using progressively degraded rat liver RNA. *BMC Biotechnology*, 7, Article 57. <https://doi.org/10.1186/1472-6750-7-57>
- Usup, G., & Azanza, R. V. (1998). Physiology and bloom dynamics of the tropical dinoflagellate, *Pyrodinium bahamense*. In D. M. Anderson, A D. Cembella & G. M. Hallegraeff (Eds.), *Physiological ecology of harmful algal blooms*. Springer.
- Usup, G., Ahmad, A., & Ismail, N. (1989). *Pyrodinium bahamense* var. *compressum* red tide studies in Sabah, Malaysia. In G. M. Hallegraeff & J. L. Maclean (Eds.), *Biology, Epidemiology and Management of Pyrodinium red tides* (pp. 97-110). Fisheries Department, Ministry of Development, Philippines.

- Usup, G., Ahmad, A., Matsuoka, K., Lim, P. T., & Leaw, C. P. (2012). Biology, ecology, and bloom dynamics of the toxic marine dinoflagellate *Pyrodinium bahamense*. *Harmful Algae*, *14*, 301-312. <https://doi.org/10.1016/j.hal.2011.10.026>
- Wang, Y., Hayatsu, M., & Fujii, T. (2012). Extraction of bacterial RNA from soil: Challenges and solutions. *Microbes and Environments*, *27*(2), 111-121. <https://doi.org/10.1264/jsme2.ME11304>
- Woodson, S. A., & Koculi, E. (2009). Analysis of RNA folding by native polyacrylamide gel electrophoresis. *Methods Enzymology*, *469*, 189-208. [https://doi.org/10.1016/S0076-6879\(09\)69009-1](https://doi.org/10.1016/S0076-6879(09)69009-1)
- Wu, M., & Tinoco, I. (1998). RNA folding causes secondary structure rearrangement. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11555-11560. <https://doi.org/10.1073/pnas.95.20.11555>
- Xu, S., He, C., Kang, Z., Song, S., & Li, C. (2021). *The ecological responses of bacterioplankton during a Phaeocystis globosa bloom in Beibu Gulf, China highlighted by integrated metagenomics and metatranscriptomics*. Research Square. <https://doi.org/10.21203/rs.3.rs-1011939/v1>
- Xu, X., Yu, Z., He, L., Cao, X., Chen, N., & Song, X. (2020). Metabolic analyses by metatranscriptomics highlight plasticity in phosphorus acquisition during monospecific and multispecies algal blooms. *Hydrobiologia*, *847*, 1071-1085. <https://doi.org/10.1007/s10750-019-04169-x>



A Comparative Study of RNA-Seq Aligners Reveals Novoalign's Default Setting as an Optimal Setting for the Alignment of HeLa RNA-Seq Reads

Kristine Sandra Pey Adum and Hasni Arsad*

Integrative Medicine Cluster, Advanced Medicine and Dental Institute, Universiti Sains Malaysia, 13200 USM, Kepala Batas, Penang, Malaysia

ABSTRACT

The introduction of RNA-sequencing (RNA-Seq) technology into biological research has encouraged bioinformatics developers to build various analysis pipelines. The chosen bioinformatics pipeline mostly depends on the research goals and organisms of interest because a single pipeline may not be optimal for all cases. As the first step in most pipelines, alignment has become a crucial step that will affect the downstream analysis. Each alignment tool has its default and parameter settings to maximise the output. However, this poses great challenges for the researchers as they need to determine the alignment tool most compatible with the correct settings to analyse their samples accurately and efficiently. Therefore, in this study, the duplication of real data of the HeLa RNA-seq was used to evaluate the effects of data qualities on four commonly used RNA-Seq tools: HISAT2, Novoalign, TopHat and Subread. Furthermore, these data were also used to evaluate the optimal settings of each aligner for our sample. These tools' performances, precision, recall, F-measure, false discovery rate, error tolerance, parameter stability, runtime and memory requirements were measured. Our results showed significant differences between the settings of each alignment tool tested. Subread and TopHat exhibited the best performance when using optimised parameters setting. In contrast, the most reliable performance was

observed for HISAT2 and Novoalign when the default setting was used. Although HISAT2 was the fastest alignment tool, the highest accuracy was achieved using Novoalign with the default setting.

ARTICLE INFO

Article history:

Received: 29 October 2021

Accepted: 14 February 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.24>

E-mail addresses:

kristinesandra@gmail.com (Kristine Sandra Pey Adum)

hasniarsad@usm.my (Hasni Arsad)

* Corresponding author

Keywords: Alignment, HISAT2, novoalign, RNA-seq, subread, TopHat

INTRODUCTION

Next-generation sequencing (NGS) is a fast-growing technology that can fulfil efficient and highly sensitive sequencing demands. In contrast with the previous sequencing technologies, such as Sanger sequencing, NGS is much cheaper and faster (Križanović et al., 2018). Gaur and Chaturvedi (2017) stated that RNA-sequencing (RNA-Seq) is a powerful technique that enhances the understanding of complex transcriptomes by revealing insights into many biological phenomena, such as the underlying mechanisms and pathways of biological processes. Other applications of NGS also include whole-genome sequencing, followed by genetic variant detection in the whole-genome or the targeted region (Qin, 2019). However, the features and massive volume of NGS reads require the development of a new generation of computational algorithms and analysis pipelines equipped to handle such data (Koboldt, 2020).

Many researchers have developed more than 60 different algorithms for the sequence reads alignment to a reference genome tool, depending on various ranges of capabilities (Fonseca et al., 2012; Keel & Snelling, 2018; Schaarschmidt et al., 2020). As alignment is the first step in the RNA-Seq pipeline, it will drastically affect the downstream analyses. The read alignment in the RNA-Seq experiment can be conducted with or without a reference genome, but most studies would prefer mapping to a reference genome as the results have been proven to be more reliable and more accurate in quantifying lowly-expressed or small transcripts (Wu et al., 2018). The major challenge in handling eukaryotic transcriptomes is the alignment of spliced transcripts reads to the reference genome. Apart from being computationally challenging (Sahlin & Mäkinen, 2021), spliced-alignment read-lengths caused difficulties in detecting isoforms with complicated splicing structures and limiting the quantification of isoform abundance (Zhang et al., 2017).

Accordingly, many spliced aligners have been developed to overcome this problem. Depending on their algorithms, these aligners mapped the reads crossing the splice junction differently. There are two algorithm approaches for the alignment step: hash-tables and Burrow-Wheeler Transform (BWT) algorithms. Hash table-based aligners operate by rapid seeding of alignment candidates. These are then extended or discarded by using more precise alignment algorithms. Then, the reference genome or the reads are split and stored in a hash table to search for the exact match of the seed locations. This low space requirement algorithm builds an index for the positions of sequences rather than sequences themselves.

While the hash table algorithm is praised for its low space consumption, the BWT-based algorithm, on the other hand, loses error tolerance for high-speed retrieval of correct matches. The representation of the data structures by top-down paths in a tree structure are called prefixes/suffixes. Then, a rapid read searching of the substring matching is enabled by primarily beginning at the root. The requirement of vast memory for the uncompressed tree structure is the main drawback of using these algorithms. In order to overcome this

problem, the Ferragina-Manzini index (FM-index) was developed by Ferragina and Manzini (2000) to reduce the memory occupied by the prefix/suffix tree. It is a compressed yet searchable suffix array-like structure based on the Burrows-Wheeler transform (BWT) (Keel & Snelling, 2018).

The selection of a suitable alignment tool for NGS data can be challenging due to the wide range of algorithms available. Therefore, various groups of researchers carried out benchmarking analyses to guide the users in choosing the correct aligners. For instance, a comprehensive benchmarking study of common splice-aware aligners was published by Baruzzo et al. (2017). The authors revealed that the aligners' performances varied by genome complexity. Unfortunately, although many benchmarking analyses had been carried out in guiding the users in choosing the best aligners, the problem is still plaguing the bioinformatics communities, while other solutions have not been derived (Donato et al., 2021; Grytten et al., 2020; Jain et al., 2020; Schilbert et al., 2020; Thankaswamy-Kosalai et al., 2017).

Comprehensive studies on alignment had been carried out, but most were using simulated data. We aimed to evaluate a more realistic setting on real data, so we chose the human cervical cancer cell line (HeLa) dataset for this study. Liu et al. (2019) proposed that HeLa cells present an essential example of human cancer cells that have broadly influenced biological studies. Furthermore, a large number of mutations and chromosomal changes in HeLa cells makes it a complex genome dynamic ecosystem of the tumour genome (Hu et al., 2019). For an aligner to be viable for RNA-Seq, it must be able to (i) align reads across splice junctions, (ii) handle paired-end reads, (iii) handle strand-specific data, and (iv) run efficiently (Baruzzo et al., 2017). Four aligners that satisfy these four requirements are HISAT2 version 2.1.0 (Kim et al., 2015), Novoalign version 4.0 (<http://www.novocraft.com/products/novoalign/>), Subread version 2.0.1 (Liao et al., 2013) and TopHat version 2.1.1 (Trapnell et al., 2009). Based on the algorithms, Novoalign and Subread adopt a hash table algorithm, while HISAT2 and TopHat adopt an FM-index algorithm.

In this study, we aimed to evaluate the effects of reading quality on alignment on four different aligners. Apart from that, we also targeted to compare the default and parameters settings of these aligners to obtain the optimal setting for HeLa RNA-Seq reads.

MATERIALS AND METHODS

Data Sets and Alignment Settings

In this experiment, we used two sets of paired-end, real Illumina sequencing data of human cervical cancer cells line (HeLa) treated with *C. Nutans*. In paired-end sequencing, a DNA fragment was selected and sequenced from both ends, producing high-quality data compared to only single-end sequencing. The raw sequenced data sets contained around 52.26 Mb and 53.89 Mb reads, respectively. In order to examine the performance of the

aligners on the real sequencing data with varying quality, we compared the alignment before and after trimming off the low-quality bases. The trimming of raw data was processed by using the fastp trimming tool (Chen et al., 2018). We processed raw and trimmed reads using FastQC (Andrews, 2010) to evaluate the quality of the bases. The plot of the qualities suggests that the trimmed reads have better quality than the raw reads. The alignments were performed using the default setting of each of the four aligners (Appendix A-D). The human reference genome used in this experiment is the hg38 genome obtained from UCSC (<http://hgdownload.soe.ucsc.edu/downloads.html>), and hg38 is chosen because this genome is the latest and most stable built of human reference genome now. In addition, this genome is the corrected and improvised version of the previous built, hg19.

For the second part of the study, an alignment of each tool was firstly performed using the default parameters using the trimmed data sets. Then, the specific parameter settings suggested by the tool were used to increase the quality of the alignments. In addition, four sets of parameter settings for each aligner were also used.

Evaluation of Precision and Recall

Alignment quality is perceived in the form of alignment precision and recall values. The precision determines which fraction of the aligned reads are being aligned correctly, while the recall value evaluates which fraction of the overall reads is correctly recovered. First, the number of true and false positive alignments was determined to estimate precision and recall values. Then, the mapping of any reader to a correct genomic location was defined as a true positive (TP), while the mapping of any read to an incorrect location was counted as a false positive (FP). Next, false positives were determined, including all the reads aligned to multiple locations. Apart from that, the reads failing to map to any correct position were considered false negative (FN) alignments. Since each read originated from one unique genome location, it should be mappable into a specific location after the alignment step. Thus, there was no such measurement for true negative alignments.

Precision, recall and false discovery rate (FDR) were calculated by using the following Equations 1 to 3:

$$Precision = \frac{TP}{TP+FP} \quad [1]$$

$$Recall = \frac{TP}{TP+FN} \quad [2]$$

$$FDR = \frac{FP}{TP+FN} \quad [3]$$

F-measure that evaluates the trade-off between precision and recall was also calculated in this study using Equation 4:

$$F = \frac{2 * precision * recall}{precision + recall} \quad [4]$$

If the multiple parameters set for one aligner resulted in equal F-measures, then the dimension of comparison would be based on runtime and memory requirements.

Impact of Parameter Choice on Alignment Quality

Four parameter combinations were tested to evaluate the impacts of each change on the alignment performances to investigate the optimised parameter setting further. In addition, the dispersion of F-measures by each alignment tool was used to determine the tool's sensitivity level to the tweaks of parameter values.

Runtime and Memory Requirements

Aligners were installed and ran on the check. If multi-threading was supported, then 12 cores were used. The memory usage was capped at 16 GB. Next, the total CPU time measurement and memory usage were extracted from the reports using the “time” command, especially the memory usage recorded from the maximum memory used during the job execution. The alignment jobs were run on Intel® Core™ i7-8700 CPU @ 3.2GHz x 12 processors.

RESULTS AND DISCUSSIONS

Aligners' Performance on Sequencing Data with Different Qualities

The performance of the alignment on the different data qualities shows a slight difference in the results. For the trimmed data of both replicates, all aligners generally show a higher concordance compared to the untrimmed data, except for TopHat. It indicated that these aligners (Novoalign, HISAT2 and Subread) worked better with high-quality reads. These high-quality reads were obtained after processing our HeLa raw reads using a fastp trimming tool where the low-quality bases and the adapter had been trimmed off. On the other hand, TopHat was not affected by the quality of the aligned reads as the trimmed data showed less concordance compared to the untrimmed data. Nevertheless, the untrimmed data of our sample still show a good concordance but is slightly lower than the trimmed ones—the comparison is shown in Supplementary Data (Table 1A).

In this comparison, it was noticed that the difference among the results was more significant in recall compared to precision. It might be caused by the increasing number of errors that affected the precision values. On the other hand, it might be due to quality concerns which can significantly mislead analytical results and lead to inaccurate conclusions (Zhou et al., 2018). It explained why it is crucial to discard the low-quality bases and adapters that might contaminate the purity of our readings.

Accuracy and Efficiency of the Aligners

The alignment accuracy and efficiency were assessed in terms of the precision and recall values. Precision reveals which portion of the reads was correctly aligned, while recall reveals which portion of the overall reads was being recovered correctly. The aligners studied were built with two different algorithms: the hash table-based and FM-index algorithms. Between the two hash-table-based aligners, Novoalign has a much higher precision value than Subread (Supplementary Data - Table 2A). However, the recall values between these two aligners were equally high (>0.92), except for the Novoalign Tweak 2 parameter setting that obtained an extremely low recall value of 0.79419. A previous study by Donato et al. (2021) compared 17 aligners on simulated and empirical NGS data, and the findings revealed that Novoalign showed the highest accuracy in all alignments. In addition, the study also highlighted that Novoalign could detect a new transcript with greater ease than the other tools tested.

While for FM-index-based alignment, a significant difference between the precision values of HISAT2 and TopHat was observed. HISAT2 showed the highest precision at optimised parameters (labelled as Tweak 2) with the value of 0.80185. TopHat's highest precision was only at 0.70199 when the alignment was carried out at the default setting. Nonetheless, the recall values ranging from 0.91513 to 0.97998 were equally high for both aligners. The highest precision value (0.94773) in HISAT2 was shown in the sample with the tweaked parameters setting labelled HISAT2 Tweak 2. However, the recall value was just average. In contrast, the lowest precision value in HISAT2 was shown in the sample with HISAT Tweak 3 parameters setting, with the value of 0.76961 but with a significantly high recall value of 0.97998. These results showed that in HISAT2, the precision and recall values had a negative correlation. Besides the percentage of mapped reads, the alignment accuracy also depends on the correctness of the reads mapped to the reference genome or transcriptome. Schaarschmidt et al. (2020) revealed that alignment using HISAT2 resulted in high overlapping reads, mainly coming from the soft clipping of the first base of the reads. The failure of TopHat and HISAT2 to tolerate the soft-clipping and mismatches had caused a large fraction of reads to be left unmapped (Sahraeian et al., 2017). However, the setting can be turned off, directly eliminating the observed differences.

Likewise, the precision values of TopHat increased with the decrement of recall values. Amongst the default and tweaked parameters settings of TopHat, the default setting was measured with an outstandingly high value of precision (0.70199). However, on the contrary, the recall value was only 0.91513 and was the lowest among the other TopHat settings. On the other hand, the TopHat with parameter Tweak 3 set had the lowest precision value (0.62911), but the recall was at a seemingly high value of 0.97864. Although this aligner performed well in aligning a read onto the respective genomic locus, similar to our findings, the study by Raplee et al. (2019) also found notable discrepancies and deficiencies

of TopHat in obtaining insufficient genomic alignment for reliable downstream analysis. Furthermore, TopHat prevented the truncation of the reads, which directly led to many unmapped reads (Sahraeian et al., 2017).

Meanwhile, for the F-measure, FM-index-based aligners showed a significant difference between the aligners, as the average of HISAT2 was 0.86355, while the average of TopHat was only 0.76479. Similarly, the F-measures between the two hash-table-based aligners also showed a vast difference. The F-measure values of Novoalign and Subread were 0.84247 and 0.76447, respectively. These results showed that the types of algorithms did not correlate with the F-measure. Overall, between these two FM-index-based aligners, it can be concluded that HISAT2 is a more reliable aligner with reasonable quality alignment.

By observing the F-measures for the overall alignment quality results, it was found that in most of the cases, the F-measures were reduced as the recall values were getting lower. It was notable in three Novoalign cases with almost similar precision values (around 0.72) but different recall values. The case with a high recall value (>0.97) showed a high F-measure, but the case with a low recall value showed an extreme drop in F-measure (0.76). Most of the time, the low recall value caused a reduction in the overall alignment quality in terms of the F-measure.

Performance of Aligners' Optimal Parameter Settings

The optimal parameter for each of the four evaluated aligners was determined by testing all permutations that appeared to have an impact on alignment quality. The optimal parameter sets run along with corresponding performance measures are shown in Table 1.

The recall metric for the optimal parameters was well balanced, and the values ranged between 0.948 (HISAT2) and 0.852 (Subread). While for the precision metric, the highest was shown by Novoalign (0.896) and the lowest shown in Subread (0.680). Novoalign displayed a significantly high value in precision and F-value metrics. FDR value also shows that Novoalign has the lowest value compared to the other aligners. The lower the FDR value defines the expected proportion of false positives among the declared significant results, so the lower the value, the better it will be. FDR is a useful approach to measure the false discoveries within a set of hypothesis tests called significant (Chen et al., 2021).

Table 1
Performance of aligners under different metrics

Metric	Novoalign	HISAT2	Tophat	Subread
Reads aligned	56252385	56049365	59413014	48916204
Recall	92.38	94.77	91.51	85.16
Precision	89.64	80.18	70.20	68.06
F- value	90.99	86.87	79.45	75.66
FDR	1.04	1.98	2.98	3.19

When testing many hypotheses, FDR is often employed to determine significance thresholds and quantify the overall error rate. We observed that Novoalign was the leading aligner with the highest accuracy.

Parameter Stability

Observing the parameters’ effects on each aligner’s performance is crucial and determining the aligner that performs well with the default settings. These evaluations will allow us to assess the robustness of the alignment qualities among parameter variations. While there are enormous spaces in manipulating the parameters of choice, the combinations may not necessarily produce a global optimum output. The main idea is that the parameter variation should allow the users to have a consistent precision and recall value to alter the runtime and memory properties without affecting the overall performance.

Parameter optimisation was performed on a duplicated data sample. Table 2 shows the dispersion values of the F-measure over the chosen parameter space. Low standard deviation, SD, as observed in HISAT2, indicated that the choice of parameters had little impact on the alignment performance. The high SD value indicated a wide alignment quality distribution, as shown by Novoalign. Meanwhile, both Subread and TopHat showed an average number of dispersions of 0.01142 and 0.02054, respectively. Different parameters had little impact on the alignment performance of HISAT2, whereas the alignment performance of Novoalign was widely affected. These results reflected that Novoalign was highly sensitive in terms of the choice of parameter settings. Hence, a precise setting must be carefully chosen when using Novoalign, as little change can cause a huge difference in the results.

Precision and recall values significantly affected the Novoalign aligner, as both of these values were widely distributed. Unlike Novoalign, the values of precision and recall in HISAT2 were consistent, with both lowly distributed. These results were not correlated with the Subread results. In Subread, there was a significantly high difference between the precision and recall values. In terms of precision, it was noticed that the values were consistent within each of the tools, regardless of the settings, except for the Novoalign default setting. The Novoalign default setting showed extremely high precision compared to the other parameter settings. However, for the recall values, the default setting of Novoalign was not as outstanding as the precision value since higher recall values were shown in the tweaked parameter settings. These results illustrated the importance of evaluating both the precision and recall values. Furthermore, the results

Table 2
Dispersion of F-measure across all parameter settings tested for each aligner

Aligner	Dispersion of F-measure
Novoalign	0.05433
HISAT2	0.00326
Tophat	0.02054
Subread	0.01142

were represented by the precision-recall analysis of HISAT2 and Novoalign, with the two aligners determined with extremely low and high F-measure dispersion, respectively (Figure 1).

Remarkably, the high dispersion value of Novoalign also resulted in a dramatic drop in the recall, especially if the value of A in the alignment scoring threshold was set at the highest score acceptable for alignment, which was 20. These results were out of the expectation, as one would expect many true positives output from this alignment scoring setting.

In contrast, the false negative value was extremely high compared to the other parameters. A low accuracy discovered in Novoalign might be due to its over-mapping at both ends of short reads (Nodehi et al., 2021; Shang et al., 2014). Thus, although dispersion does not state the alignment software’s overall performance, it indicates whether the optimal performance can be achieved without an in-depth understanding of algorithmic details.

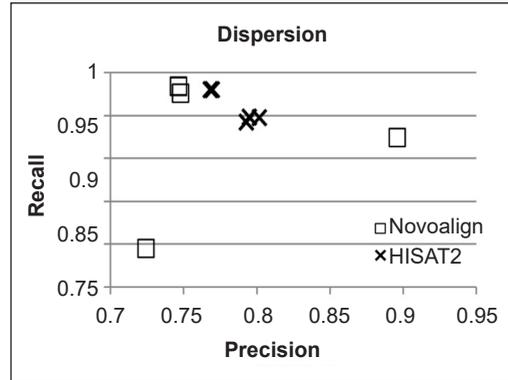


Figure 1. Influence of parameter selection. Precision (x-axis) and recall (y-axis) are shown for Novoalign (squares) and HISAT2 (X signs) aligners

Runtime and Memory Requirements

The runtime evaluation of each aligner was based on two main steps in the alignment process. The first is the indexing and the second one is the alignment. By comparing the indexing runtime for each of the aligners, it was found that the relationship between the indexing time correlated with the types of algorithms used. For example, HISAT2 and TopHat, FM-index-based aligners require a longer time (more than 60 mins) to build a genome index. In contrast, hash-table-based aligners, like Subread and Novoalign, can build genome index in less time (< 15mins).

Most of these four aligners were designed with a trade-off between the indexing and alignment runtime (Table 3). For example, Subread and Novoalign were able to build an index of the genome within a short duration. Still, they required plenty of alignment time at the default setting, with 47.30 mins and 244.8 mins recorded, respectively. Conversely,

Table 3
Indexing and alignment runtime of the aligners

Aligner	Algorithm	Indexing runtime (mins)	Alignment runtime (mins)
HISAT2	FM-index based	61.13	12.02
TopHat		78.39	602.19
Subread	Hash-table based	13.30	47.3
Novoalign		10.49	244.8

HISAT2 required a short duration to build an index but took 12.02 mins of alignment time when using the default setting. It was approximately five times faster than the indexing runtime (Figure 2).

On the other hand, TopHat required an extremely long time for indexing and aligning compared to other tools. The TopHat’s default setting required 602.19 mins to complete the alignment. However, TopHat required the least memory compared to the other three tools to compensate for the lengthy runtime. TopHat only required 4.0 Gb RAM, while Novoalign, Subread and HISAT2 required 8.0 Gb, 10.0 Gb and 6.7 Gb of RAM, respectively.

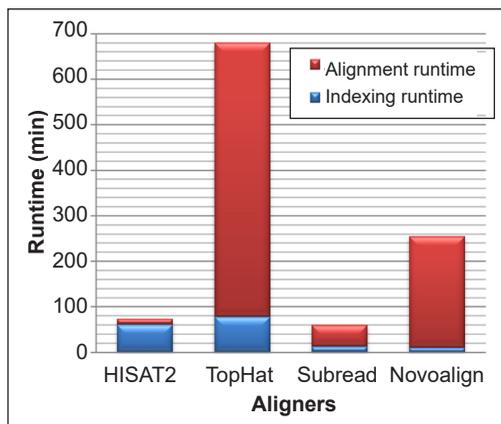


Figure 2. The measurements of indexing and alignment runtime of each aligner on the default setting

By observing the runtime for the default and optimised parameters settings of each aligner, we found that the alignment runtime of each aligner barely depended on the types of algorithms but more on the individual settings. Although the runtime differs significantly among the hash table-based aligners, the runtime among Subread’s optimised parameters and default settings was more consistent than Novoalign. Subread generally required 1 hour of runtime and even less for the default setting. In contrast to Subread, the default setting of Novoalign required more than 2 hours of runtime to obtain a comparable precision and recall. In addition, Novoalign had the longest computational time of reads mapping, probably due to its predispositions toward other parameters (Donato et al., 2021).

As for the FM-index-based approach, the alignment runtime of the aligners varied vastly. HISAT2 outperformed TopHat in terms of the runtime. The average runtime of HISAT2 was only 12 mins and 43 sec, while TopHat needed at least 3.45 hours of runtime. In terms of memory consumption, HISAT2 required 6.7 Gb RAM, while TopHat required the least memory at only 4.0 Gb RAM. These results were consistent with the previous findings by Keel and Snelling (2018), who found that HISAT2 was significantly faster and used less memory through simulated data sets. Despite the very low memory consumption, TopHat could still achieve a reasonable alignment quality, thus supporting the widespread use of TopHat within many of the RNA-seq mapping approaches and as the most cited aligner. On the other hand, no significant correlation was observed between the F-measure with the runtime and memory requirements.

HISAT2 performed extremely fast alignment with comparable accuracy to the other aligners. In contrast, the alignment runtime for Subread was the second-fastest but achieved poor alignment quality. The runtime for Novoalign was acceptable and at an average rate,

while the alignment with TopHat was the slowest. The alignment with TopHat was thus considered inefficient, especially when working with multiple data sets.

Error Tolerance

This section aims to analyse each aligner's sensitivity in response to a specific number of errors allowed in the reads. Goodwin et al. (2016) believed that NGS platforms provide a massive amount of data, while each platform is associated with error rates ranging from 0.1 to 15%. Therefore, a good alignment algorithm used in mapping sequence data must be able to compensate for these inevitable raw data errors (Keel & Snelling, 2018). A previous study by Sun et al. (2017) found that alignment is a critical step for intermediate indel detection. Therefore, each read was measured to determine the number of mismatches or indels for this purpose. The highest number of mismatches and indels in a correctly aligned read was 10 and 5, respectively, for all the aligners tested, except for Novoalign. Unlike other aligners, Novoalign first searches the candidate alignment positions from the reference genome using the Needleman-Wunsch algorithm based on the alignment score. Due to this alignment-score-based search algorithm, the users cannot define the number of allowed mismatches in each alignment, but the users can still set up a threshold of an alignment score. The mapping quality scores define the accuracy of alignment, meaning that the higher the alignment quality score, the more accurate an alignment is. Thus, the alignment score threshold is from 30 to 180 for Novoalign.

Figure 3 shows the impacts of errors on the alignment quality for each aligner by using its default setting. Generally, it was noticed that the precision of the alignments was barely affected by the number of errors. As illustrated in Figures 3(c) and 3(d), the precision values of HISAT2 and TopHat showed a flat line. Sun et al. (2017) believed that most variant calling programs would miss the intermediate indels from these aligners, except when the soft-clipped reads were sufficiently triggered. Furthermore, the study also discovered that the TopHat family RNA-seq mapping programs do not align the reads with intermediate indels, or the reads were minimally aligned when HISAT2 was used.

In contrast with precision readings, the recall values obtained showed that more significant changes were recorded with increasing errors. Interestingly, we also determined a drastic drop in Subread once the mismatches and indels were set at 5 and 10, respectively, as shown in Figure 3(b). It was possibly due to the general design of the alignment algorithm itself, as the algorithm is robust enough to detect a small number of single-based mismatches, depending on the parameter setting. As a result, most of the algorithm's recall stayed relatively constant while still being within the tolerated range of mismatches but dropped significantly as soon as this range was exceeded, as shown in the Subread aligner. In addition, the other aligners could tolerate up to five mismatches. For Novoalign, as illustrated in Figure 3(a), the recall rate gradually decreased as the alignment score decreased.

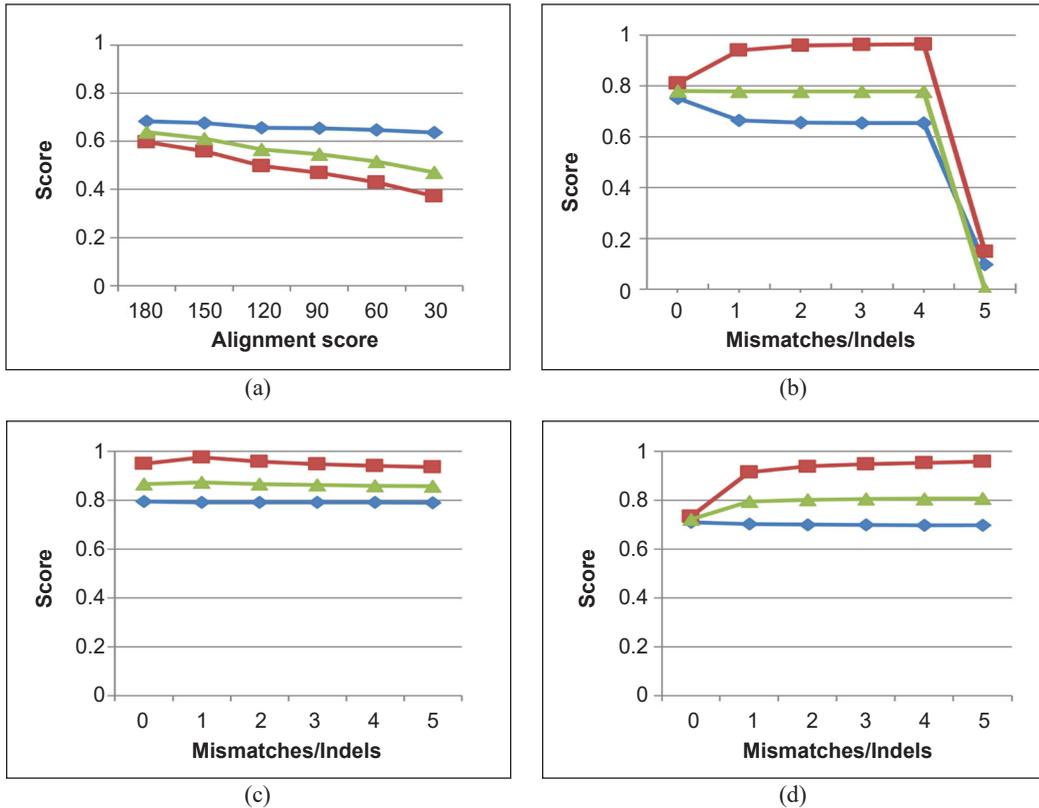


Figure 3. The impacts of errors on the alignment quality for each of the aligners tested: (a) Novoalign; (b) Subread; (c) HISAT2; and (d) TopHat. The dependencies of alignment precision (blue lines), recall (red lines) and F-measure (green lines) on each of the alignment algorithms based on the default setting were analysed in this study.

Some aligners tolerate indels as these tools are designed to handle gapped alignments. As a result, most algorithms' recall values were considerably impaired by indels' presence. A remarkable tolerance to indels was shown by HISAT2 and TopHat, with a near-constant performance observed, even as the indel counts increase. Even though the baseline recall of Novoalign was already rather low at 0.6, the other aligners were shown to tolerate indels, with significantly higher values of baseline recall recorded. Subread was found to be the most vulnerable to indels among these aligners. The main limitation in the gapped alignment-based indels detection method is the need for indels that are entirely contained within a read and correctly detected during the initial read mapping step (Donato et al., 2021; Li et al., 2009). In the case of small indel detection, the supporting reads are frequently presented with too few bases that can match with the reference genome, or the reads may contain only one end that can map correctly to the reference genome. However, the remaining bases following the indel may be trimmed or soft-clipped by the NGS aligner (Donato et al., 2021; Landman et al., 2014).

As our study was using RNA-Seq data of HeLa cells, we needed to have the tools that could tolerate indels and mismatches. It was because, according to Bottomley et al. (1969), Liu et al. (2019), and Rutledge (2014), HeLa contained a huge number of genomic variants. Furthermore, it was reported that the variation in the karyotype of HeLa had been stated inconsistently and proven that such changes had affected the phenotype of the cells (Fasterius & Al-Khalili Szigyarto, 2018). Indeed, HeLa has been purposely established to be genetically unstable (Fasterius & Al-Khalili Szigyarto, 2018; Yoo et al., 2017). Therefore, by allowing the mismatches in the alignment, we still managed to handle these biological characteristics of HeLa whilst obtaining the maximum alignment precision.

CONCLUSION

In this study, we concluded that the alignment procedure's accuracy depended on two aspects. The first was the quality of the reads, and the second was the parameter settings. A high-quality read is free from any adapter contamination and low-quality bases. The alignments of the high-quality reads show significant increment in all aligners tested in this study except for TopHat data 1. It indicated that the TopHat alignment quality was not affected by the quality of the reads. While for the parameter settings, the number of mismatches and indels allowed displayed a great impact on alignment accuracy.

Apart from the accuracy, the notable difference between these aligners was the runtime. After comparing each aligner's runtime and alignment qualities, we can summarise that the fast runtime did not guarantee a high alignment quality. So, the choice of the aligner depends on the target of the researchers. Nevertheless, our study found that HISAT2 performed extremely fast alignment with comparable accuracy to the other aligners.

In conclusion, the default setting of Novoalign was the most reliable setting that suited our data sample. In addition, our study provides a systematic comparison between the commonly used alignment programmes in RNA-Seq studies. Furthermore, the approach from this study can be applied to future research that deals with human cancer cell datasets generated from different platforms. Moreover, this study can also be utilised as guidance in selecting the best alignment tool for various sources of samples.

In this study, we focused mainly on comparing the aligners from two specific angles, using real data of HeLa cells reads of varying qualities. Thus, there were a few limitations in this study. Firstly, rather than the data from real cases of cancer, we only used a sample from a cell line. Secondly, our sequencing data sets were only sequenced from the Illumina sequencer. Third, we were only using pair-end data in this study without using the single-pair end to compare the results. Lastly, while there are many available alignment tools, we only compared a small number of alignment tools in this study. Although this article has these limitations, our study applies to the pair-end whole genome real sequencing data and data generated from other sources. The main challenge we faced was to set up a fair

comparison of the chosen parameters and evaluation metrics among these different aligners. Regardless of this challenge, we managed to achieve the objectives of this study, and we expect that this study can guide other researchers in choosing the optimised settings of the aligners of interest.

ACKNOWLEDGEMENTS

The authors want to thank Universiti Sains Malaysia, Malaysia, for the support.

REFERENCES

- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Baruzzo, G., Hayer, K. E., Kim, E. J., Di Camillo, B., Fitzgerald, G. A., & Grant, G. R. (2017). Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods*, *14*(2), 135-139. <https://doi.org/10.1038/nmeth.4106>
- Bottomley, R. H., Trainer, A. L., & Griffin, M. J. (1969). Enzymatic and chromosomal characterization of HeLa variants. *The Journal of Cell Biology*, *41*(3), 806-815. <https://doi.org/10.1083/jcb.41.3.806>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen, X., Robinson, D. G., & Storey, J. D. (2021). The functional false discovery rate with applications to genomics. *Biostatistics*, *22*(1), 68-81. <https://doi.org/10.1093/biostatistics/kxz010>
- Donato, L., Scimone, C., Rinaldi, C., D'Angelo, R., & Sidoti, A. (2021). New evaluation methods of read mapping by 17 aligners on simulated and empirical NGS data: An updated comparison of DNA- and RNA-seq data from Illumina and Ion Torrent technologies. *Neural Computing and Applications*, *33*(22), 15669-15692. <https://doi.org/10.1007/s00521-021-06188-z>
- Fasterius, E., & Al-Khalili Szgyarto, C. (2018). Analysis of public RNA-sequencing data reveals biological consequences of genetic heterogeneity in cell line populations. *Scientific Reports*, *8*(1), 1-11. <https://doi.org/10.1038/s41598-018-29506-3>
- Ferragina, P., & Manzini, G. (2000). Opportunistic data structures with applications. In *Proceedings 41st Annual Symposium on Foundations of Computer Science* (pp. 390-398). IEEE Publishing. <https://doi.org/10.1109/sfcs.2000.892127>
- Fonseca, N. A., Rung, J., Brazma, A., & Marioni, J. C. (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics*, *28*(24), 3169-3177. <https://doi.org/10.1093/bioinformatics/bts605>
- Gaur, P., & Chaturvedi, A. (2017). *A survey of bioinformatics-based tools in RNA-sequencing (RNA-seq) data analysis*. In *Translational Bioinformatics and its Application* (pp. 223-248). Springer. https://doi.org/10.1007/978-94-024-1045-7_10
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>

- Grytten, I., Rand, K. D., Nederbragt, A. J., & Sandve, G. K. (2020). Assessing graph-based read mappers against a novel baseline approach highlights strengths and weaknesses of the current generation of methods. *BMC Genomics*, *21*, Article 282. <https://doi.org/10.1186/s12864-020-6685-y>
- Hu, W. E., Zhang, X., Guo, Q. F., Yang, J. W., Yang, Y., Wei, S. C., & Su, X. D. (2019). HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PLoS One*, *14*(12), Article e0225466. <https://doi.org/10.1371/journal.pone.0225466>
- Jain, C., Rhie, A., Zhang, H., Chu, C., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Weighted minimizer sampling improves long read mapping. *Bioinformatics*, *36*, I111-I118. <https://doi.org/10.1093/BIOINFORMATICS/BTAA435>
- Keel, B. N., & Snelling, W. M. (2018). Comparison of Burrows-Wheeler transform-based mapping algorithms used in high-throughput whole-genome sequencing: Application to illumina data for livestock genomes 1. *Frontiers in Genetics*, *9*, 1-6. <https://doi.org/10.3389/fgene.2018.00035>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357-360. <https://doi.org/10.1038/nmeth.3317>
- Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, *12*(1), 1-13. <https://doi.org/10.1186/s13073-020-00791-w>
- Križanović, K., Echchiki, A., Roux, J., & Šikić, M. (2018). Evaluation of tools for long read RNA-seq splice-aware alignment. *Bioinformatics*, *34*(5), 748-754. <https://doi.org/10.1093/bioinformatics/btx668>
- Landman, S. R., Hwang, T. H., Silverstein, K. A. T., Li, Y., Dehm, S. M., Steinbach, M., & Kumar, V. (2014). SHEAR: Sample heterogeneity estimation and assembly by reference. *BMC Genomics*, *15*(1), 1-12. <https://doi.org/10.1186/1471-2164-15-84>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liao, Y., Smyth, G. K., & Shi, W. (2013). The subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108-e108. <https://doi.org/10.1093/nar/gkt214>
- Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E. G., Van Drogen, A., Borel, C., Frank, M., Germain, P. L., Bludau, I., Mehnert, M., Seifert, M., Emmenlauer, M., Sorg, I., Bezrukov, F., Bena, F. S., Zhou, H., Dehio, C., Testa, G., & Aebersold, R. (2019). Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nature Biotechnology*, *37*(3), 314-322. <https://doi.org/10.1038/s41587-019-0037-y>
- Nodehi, H. M., Tabatabaiefar, M. A., & Sehhati, M. (2021). Selection of optimal bioinformatic tools and proper reference for reducing the alignment error in targeted sequencing data. *Journal of Medical Signals and Sensors*, *11*(1), 37-44. <https://doi.org/10.4103/jmss.JMSS-7-20>
- Qin, D. (2019). Next-generation sequencing and its clinical application. *Cancer Biology and Medicine*, *16*(1), 4-10. <https://doi.org/10.20892/j.issn.2095-3941.2018.0055>
- Raplee, I. D., Evsikov, A. V., & De Evsikova, C. M. (2019). Aligning the aligners: Comparison of rna sequencing data alignment and gene expression quantification tools for clinical breast cancer research. *Journal of Personalized Medicine*, *9*(2), Article 18. <https://doi.org/10.3390/jpm9020018>

- Rutledge, S. (2014). What HeLa cells are you using? *The Winnower*, 9, 1-9. <https://doi.org/10.15200/winn.143896.65158>
- Sahlin, K., & Mäkinen, V. (2021). Accurate spliced alignment of long RNA sequencing reads. *Bioinformatics*, 37(24), 4643-4651. <https://doi.org/10.1093/bioinformatics/btab540>
- Sahraeian, S. M. E., Mohiyuddin, M., Sebra, R., Tilgner, H., Afshar, P. T., Au, K. F., Bani Asadi, N., Gerstein, M. B., Wong, W. H., Snyder, M. P., Schadt, E., & Lam, H. Y. K. (2017). Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nature Communications*, 8(1), 1-14. <https://doi.org/10.1038/s41467-017-00050-4>
- Schaarschmidt, S., Fischer, A., Zuther, E., & Hincha, D. K. (2020). Evaluation of seven different RNA-seq alignment tools based on experimental data from the model plant *Arabidopsis thaliana*. *International Journal of Molecular Sciences*, 21(5), Article 1720. <https://doi.org/10.3390/ijms21051720>
- Schilbert, H. M., Rempel, A., & Pucker, B. (2020). Comparison of read mapping and variant calling tools for the analysis of plant NGS data. *Plants*, 9(4), Article 439. <https://doi.org/10.3390/plants9040439>
- Shang, J., Zhu, F., Vongsangnak, W., Tang, Y., Zhang, W., & Shen, B. (2014). Evaluation and comparison of multiple aligners for next-generation sequencing data analysis. *BioMed Research International*, 2014, Article 309650. <https://doi.org/10.1155/2014/309650>
- Sun, Z., Bhagwate, A., Prodduturi, N., Yang, P., & Kocher, J. P. A. (2017). Indel detection from RNA-seq data: Tool evaluation and strategies for accurate detection of actionable mutations. *Briefings in Bioinformatics*, 18(6), 973-983. <https://doi.org/10.1093/bib/bbw069>
- Thankaswamy-Kosalai, S., Sen, P., & Nookaew, I. (2017). Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics. *Genomics*, 109(3-4), 186-191. <https://doi.org/10.1016/j.ygeno.2017.03.001>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics*, 25(9), 1105-1111. <https://doi.org/10.1093/bioinformatics/btp120>
- Wu, D. C., Yao, J., Ho, K. S., Lambowitz, A. M., & Wilke, C. O. (2018). Limitation of alignment-free tools in total RNA-seq quantification. *BMC Genomics*, 19(1), 1-14. <https://doi.org/10.1101/246967>
- Yoo, Y. S., Han, H. G., & Jeon, Y. J. (2017). Unfolded protein response of the endoplasmic reticulum in tumor progression and immunogenicity. *Oxidative Medicine and Cellular Longevity*, 2017, Article 2969271. <https://doi.org/10.1155/2017/2969271>
- Zhang, C., Zhang, B., Lin, L. L., & Zhao, S. (2017). Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics*, 18(1), 1-11. <https://doi.org/10.1186/s12864-017-4002-1>
- Zhou, Q., Su, X., Jing, G., Chen, S., & Ning, K. (2018). RNA-QC-chain: Comprehensive and fast quality control for RNA-Seq data. *BMC Genomics*, 19(1), 1-10. <https://doi.org/10.1186/s12864-018-4503-6>

SUPPLEMENTARY DATA

Table 1A

Percentage of reads uniquely aligned in two data sets by four aligners under default setting

Aligners	Metrics	Dataset 1		Dataset 2	
		Untrimmed data	Trimmed data	Untrimmed data	Trimmed data
Novoalign	Precision	0.8678	0.8964	0.8262	0.8618
	Recall	0.98886	0.9238	0.9526	0.8693
	F	0.9242	0.9099	0.8849	0.8655
	FDR	0.1323	0.1036	0.1738	0.1382
HISAT2	Precision	0.7554	0.7947	0.7851	0.8018
	Recall	0.7941	0.9484	0.9816	0.9477
	F	0.7743	0.8648	0.8724	0.8687
	FDR	0.2446	0.2053	0.2149	0.1982
Tophat	Precision	0.7087	0.6286	0.6348	0.6291
	Recall	0.8620	0.8898	0.9451	0.9786
	F	0.7779	0.7367	0.7608	0.7659
	FDR	0.2913	0.314	0.3652	0.3709
Subread	Precision	0.6028	0.6806	0.5794	0.6330
	Recall	0.8689	0.8516	0.8648	0.9401
	F	0.7118	0.7566	0.6940	0.7566
	FDR	0.3972	0.3194	0.4205	0.3670

Table 2A

The alignment summaries of default and four sets of parameters settings by using Novoalign, HISAT2, TopHat and Subread aligners

Tools	Sample	Total reads	Precision	Recall	F	FDR	Runtime	Memory
Novoalign	NOVOA default	56252385	0.89636	0.92376	0.90985	0.10364	138m 39s	8.0 Gb
	NOVOA tweak1	68796275	0.74681	0.98410	0.84919	0.25319	270m 04s	8.0 Gb
	NOVOA tweak2	64061997	0.72463	0.79419	0.75782	0.27537	57m 59s	8.0 Gb
	NOVOA tweak3	68769479	0.74690	0.98297	0.84883	0.25310	272m 22s	8.0 Gb
	NOVOA tweak4	68497483	0.74798	0.97529	0.84664	0.25202	244m 08s	8.0 Gb
HISAT2	HISAT2 default	56525790	0.79269	0.94209	0.86095	0.20731	11m 08s	6.7 Gb
	HISAT2 tweak1	57616454	0.76800	0.97996	0.86113	0.23200	14m 38s	6.7 Gb
	HISAT2 tweak2	56049365	0.80185	0.94773	0.86870	0.19816	11m 44s	6.7 Gb
	HISAT2 tweak3	57530440	0.76961	0.97998	0.86215	0.23039	13m 23s	6.7 Gb
	HISAT2 tweak4	56519361	0.79474	0.94844	0.86481	0.20526	12m 02s	6.7 Gb
TopHat	TOPHAT default	59413014	0.70199	0.91513	0.79452	0.29801	600m 30s	4.0 Gb
	TOPHAT tweak1	70379105	0.62912	0.97838	0.76580	0.37089	274m 14s	4.0 Gb
	TOPHAT tweak2	67973298	0.63477	0.94999	0.76103	0.36523	207m 53s	4.0 Gb
	TOPHAT tweak3	70390636	0.62911	0.97864	0.76588	0.37089	284m 12s	4.0 Gb
	TOPHAT tweak4	62577636	0.62861	0.88975	0.73673	0.37139	602m 19s	4.0 Gb
Subread	Subread default	48916204	0.65845	0.95194	0.77845	0.34155	45m 45s	10.0 Gb
	Subread tweak1	48916204	0.64358	0.97506	0.77538	0.35642	49m 30s	10.0 Gb
	Subread tweak2	48916204	0.61826	0.97060	0.75537	0.38174	69m 51s	10.0 Gb
	Subread tweak3	48916204	0.63300	0.94011	0.75658	0.36700	68m 46s	10.0 Gb
	Subread tweak4	48916204	0.68062	0.85162	0.75658	0.31938	47m 30s	10.0 Gb

APPENDICES

Aligners Command and Parameters Setting

Appendix A

Alignment with HISAT2

Read alignment was performed using command as follows:

```
hisat2 --threads 16 --time --reorder --known-splicesite-infile <output index
path>/<genome name>.splicesites.txt --novel-splicesite-outfile splicesites.novel.txt -
-novel-splicesite-infile splicesites.novel.txt -f -x <index name> -1 <read file 1> -2
<read file 2> -S <output sam file>
-N <NUM_MISMATCH>
-L <SEED_LENGTH>
-i S,1,<SEED_INTERVAL>
-D <SEED_EXTENSION>
-R <RE_SEED>
--pen-noncansplice<PENALITY_NONCANONICAL>
--mp <MAX_MISMATCH_PENALITY>,<MIN_MISMATCH_PENALITY>
--sp <MAX_SOFTCLIPPING_PENALITY>,<MIN_SOFTCLIPPING_PENALITY>
```

Sample	-N <NUM_MISMATCH> -L <SEED_LENGTH> -i S,1,<SEED_INTERVAL> -D <SEED_EXTENSION> -R <RE_SEED> --pen noncansplice<PENALITY_ NONCANONICAL> --mp <MAX_MISMATCH_PENALITY>,<MIN_ MISMATCH_PENALITY> --sp<MAX_SOFTCLIPPING_PENALITY>,<MIN_SOFTCLIPPING_ PENALITY>
default	0-22-1.15-15-2-3-6,2-2,1
Tweak 1	0-20-0.5-25-5-20-1,0-3,0
Tweak 2	1-20-0.5-25-5-20-3,0-2,1
Tweak 3	0-22-1.5-15-2-20-1,0-2,1
Tweak 4	OFF-OFF- OFF-OFF- OFF-OFF- OFF-OFF

Appendix B

Alignment with Novoalign

Read alignment was performed using command as follows:

```
novoalign -d <output index file> -f <read file 1> <read file 2> -F FA -o SAM -r All 10
-t <A_SCORE>,<B_SCORE> -h -1 -1 -i PE <FRAGMENT_LENGTH_MEAN>,<FRAGMENT_
LENGTH_SD> -v 0
70 70 “[>][[^:]*)” > <output sam file> 2>alignment.log
```

Sample	REPEAT_FILTER -t <A_SCORE>,<B_SCORE>
default	Default-Default-default
tweak 1	OFF-10-4.5
tweak 2	OFF-20-2
tweak 3	OFF-12-4.5
tweak 4	default-20-4.5

Appendix C

Alignment with Subread

Read alignment was performed using command as follows:

```
subjunc -i <index> -r <read file 1> -R <read file 2> -T 16
--SAMoutput -o <output alignment> --allJunctions
-d <MIN_FRAGMENT_LENGTH>
-I <INDEL>
-m <NUM_HIT_SUBREADS>
-M <MISMATCHES>
-n <NUM_EXTRACTED_SUBREADS>
-p <NUM_HIT_PAIR_SUBREADS>
--complexIndels
```

Sample	-d <MIN_FRAGMENT_LENGTH> -I <INDEL> -m <NUM_HIT_SUBREADS> -M <MISMATCHES> -n <NUM_EXTRACTED_SUBREADS> -p <NUM_HIT_PAIR_SUBREADS> --complexIndels
default	50-5-3-3-10-1-off
tweak 1	0-10-1-20-15-1-off
tweak 2	0-10-1-20-5-1-on
tweak 3	0-10-3-3-10-3-on
tweak 4	0-5-3-3-10-1-off

Appendix D

Alignment with TopHat

Read alignment was performed using command as follows:

```
tophat2 --output-dir <output path> <index> <reads file 1> <reads file 2> --num-threads 16 --GTF <gtf file>
--mate-inner-dist <INNER_MATE_MEAN>
--mate-std-dev <INNER_MATE_SD>
--b2-very-sensitive
--read-mismatches <NUM_MISMATCHES>
--read-gap-length <NUM_GAP_LENGTH>
--read-edit-dist <NUM_EDIT_DIST>
--read-realign-edit-dist <NUM_REALIGN_EDIT_DIST>
--max-insertion-length <NUM_INSERTION_LENGTH>
--max-deletion-length <NUM_DELETION_LENGTH>
--max-multihits <NUM_MULTIHITS>
```

Sample	B2_VERY_SENSITIVE - NUM_MISMATCHES - NUM_GAP_LENGTH - NUM_EDIT_DIST - NUM_REALIGN_EDIT_DIST - NUM_INSERTION_LENGTH - NUM_DELETION_LENGTH - NUM_MULTIHITS
default	On-2-2-2-3-3-3-20
tweak 1	On-18-25-25-26-24-24-100
tweak 2	On-7-6-7-8-4-4-100
tweak 3	On-25-25-25-26-24-24-100
tweak 4	OFF-OFF- OFF-OFF- OFF-OFF- OFF-OFF



The Kinetic Evaluation and DFT Study of Cis-[Pt(Asc)(NH₃)₂] Complex as an Inhibitor to Type 2 Diabetic Human Amylase

Khalid Farouk Al-Rawi¹, Khaldoon Taher Maher², Othman Ibrahim Alajrawy^{3*} and Firas Taher Maher⁴

¹Department of Chemistry, College of Science, University of Anbar, 00964 Iraq

²Al-Amal National Hospital for Cancer Management, Baghdad 00964 Iraq

³Department of Applied Chemistry, College of Applied Science, University of Fallujah, 00964 Iraq

⁴Department of Chemistry, College of Science, Tikrit University, Tikrit 00964 Iraq

ABSTRACT

Several metal complexes and organic compounds and extracted herbs that might be involved in the bio-mechanism of the type 2 diabetes mellitus treatments. This research aims to synthesize a new platinum (II) complex and study its kinetics as an inhibitor for freshly purified amylase from type 2 human diabetics. The amylase enzyme was precipitated from diabetic patients. The complex cis- [Pt(Asc)(NH₃)₂] was synthesized and characterized experimentally and theoretically by DFT calculations to conclude the structure. Both calculations confirmed the square planar geometry for the prepared complex. The results showed that the complex is more stable and polar than the L-ascorbic acid derivative. Therefore, we suggested that the synthesized Pt(II) complex is appropriate to be examined as an inhibitor for the amylase enzyme. Several concentrations from the Pt(II) complex were prepared for kinetic purposes. Kinetic results have shown that the newly prepared

complex has a remarkable inhibition effect on the amylase enzyme. Kinetic parameters were fitted using the Lineweaver–Burk plot. The inhibition reaction was confirmed as a non-competitive inhibitor. Also, an inorganic compound derived from vitamin C was prepared and diagnosed by several spectroscopic methods, and a comparison between the experimental and theoretical data was conducted. The DFT study of the prepared complex gave a useful explanation for the complex and its stability. Thus, an

ARTICLE INFO

Article history:

Received: 10 November 2021

Accepted: 05 April 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.25>

E-mail addresses:

sc.kfwi72@uoanbar.edu.iq (Khalid Farouk Al-Rawi)

khaldontaher@gmail.com (Khaldoon Taher Maher)

othman-ibrahimasc@uofallujah.edu.iq (Othman Ibrahim Alajrawy)

Firastaher3@gmail.com (Firas Taher Maher)

* Corresponding author

inhibitory effect on the activity of the amylase enzyme was clearly shown by the newly prepared Pt(II) complex. It can be concluded that Pt(II) complex could be used as an amylase inhibitor.

Keywords: Amylase, cisplatin, inhibition complex, L-ascorbic acid, platinum complexes, T2DM, theoretical calculation

INTRODUCTION

Diabetes mellitus is a chronic disease when the pancreas cannot generate enough insulin (Ard et al., 2020). It also occurs when the body is unable to successfully consume the produced insulin (Halim & Halim, 2019). It is well known that insulin is a regulatory hormone for the sugar level in the blood. The symptom of hyperglycemia or hypoglycemia is uncontrolled diabetes which can cause severe impairment to some organs, such as nerves and vessels of the blood (Egan & Dinneen, 2019).

The digestion of the carbohydrates into small sugar molecules such as glucose starts with the amylase enzyme (Association, 2014). A human amylase is an enzyme that has the following commission number E.C. 3.2.1.1 (EFSA Panel on Food Contact Materials et al., 2020). The main function of amylase is to digest carbohydrates molecules into smaller subunits (Kumar & Chakravarty, 2018). There are two main types of the amylase enzyme, salivary and pancreatic, found in some fungi, bacteria, and plant seeds (Hong et al., 2019; López et al., 2018; Whitcomb & Lowe, 2007). The amylase enzyme produced from saliva is used to hydrolyze internal α -1,4-glycosidic linkages in starch in low molecular weight products, such as maltose, glucose, and maltotriose units (Yu, 2019). The pancreatic amylase is associated with the digestion of starch. This enzyme is found in the blood, helping in digesting the dead white blood cells (Chen et al., 2019). The amylase enzyme can be used to produce corn syrup in high-fructose concentration and alcohol. It is also used to facilitate the digestion of animal food, as well as in the yeast industry (Burhan et al., 2003).

However, the more digested carbohydrate in the body and sugar absorbed, the more chances for hyperglycemia and diabetes mellitus risks. Therefore, amylase inhibition would be a good way to decrease sugar production in the body (Oboh et al., 2014). Several studies have suggested various metal complexes, synthesized organic compounds, and plant extracts involved in the amylase inhibition mechanisms (Bergamini et al., 2008). We have coordinately synthesized a new platinum (II) complex using L-ascorbic acid as a ligand. This complex was synthesized, characterized by different spectroscopic techniques, and finally examined by computational studies utilizing density functional theory (DFT). It is suggested that the platinum complex may bind to an active site of the enzyme or another site which affects the binding of substrate to the enzyme's active site, leading to amylase inhibition (Mahmood, 2016).

The metal ascorbates complexes had previously shown beneficial biological applications. First, they were used for treating symptoms related to vitamin C and metal ion deficiencies. Secondly, metal ascorbates had been previously utilized to develop some therapeutic agents used as potential anticancer, antibacterial, antioxidant, and antihypoxic (Bradshaw et al., 2011; Romero-Canelon & Sadler, 2013). In addition, metal ascorbates were also used as synthetic models for some metal-containing biological systems. Studies have investigated the bioactivity of the metal ascorbate complexes such as *Cis*-diamineplatinum (II) ascorbate complexes. These studies confirmed the activity of these complexes in the anticancer prodrugs (Hollis et al., 1987). *Cis*-diamineplatinum (II) ascorbate complexes had revealed antitumor activity (Hollis et al., 1987). Similar complexes had been identified as the best-found metal involving the drug. [Pt(phen)₂]⁺² had shown a remarkable inhibiting effect on tumor cells (Zümreoglu-Karan, 2006). Another well-identified complex is phenanthroline ascorbates which had been revealed as an anticancer drug. Other researchers synthesized complexes as anticancer (Czarnomysy et al., 2018), antineoplastic activities (Yalçın, 1995), and antiproliferative activity (Bergamini et al., 2008).

MATERIALS AND METHODS

Subjects

Patients. One hundred diabetic patients and 100 healthy individuals were recruited as a control group throughout the study from December 2019 to the end of March 2020 in Tikrit Teaching Hospital, Tikrit, Iraq. The research was registered by the Research Ethics Committee of the Iraqi Ministry of Health, Iraq. Ten milliliters of blood were collected into a test tube without anticoagulants. Blood samples were left for 20–30 minutes at 37 °C. The serum was obtained by blood centrifugation at 4000 rpm (1814 × g) for 10 min, and then the serum was used for the kinetic assays. Amylase was estimated using amylase Kits from BIOLABO Company-BIOLABO SAS-AMYLASE (NPG3: Liquid reagent) REF.LP99553, the French manufacturer of Reagents for Medical Biology, France, uses the colorimetric method.

Purification of Amylase

The following procedure was used to purify amylase: (NH₄)₂SO₄ (60%), 12 g was added to precipitate amylase (Duong-Ly & Gabelli, 2014). The precipitate was dissolved in a Tris-HCl buffer pH 7.0 and dialyzed out of salt using a buffer solution pH 6.0. Next, an Ion-exchange (I.E) type Diethylaminoethyl-Sephacel (DEAE-Sephacel) was used for further purification and to make the enzyme more concentrated with a flow rate of 2 ml/4 min (Albuquerque et al., 2016). Then the product was loaded to the gel filtration chromatography (G.L) column type Sephadex G50, pH: 6, with a flow rate of 2 ml /min on

column 30×2.5 cm to confirm the molecular weight of the enzyme (Zhao et al., 2018). All materials were purchased from BDH and Fluka Companies. The electrophoresis technique was finally used to confirm the purity and the molecular weight of the amylase enzyme using polyacrylamide gel on SDS-PAGE. The electrophoresis adopted in this study was carried out with a current of 30 mA, and the gel was stained with a Coomassie brilliant blue R – 250 (Hames, 1998).

Preparation of the Compounds

Preparation of the L-ascorbic acid derivative and the cis-[Pt(Asc)(NH₃)₂] complex was carried out according to the methods described in previous studies (Salomon, 1963).

Effect of the New Inhibitors

The Pt(II) compound was measured at different concentrations ranging from 5×10^{-1} to 5×10^{-8} g/25 ml. In addition, the type of inhibition was studied with a sequential concentration of the substrate 0.5–5 mM and a single concentration of Pt(II) complex 0.005 g/25 ml. Finally, a Lineweaver–Burk plot was applied to obtain the *K_m*, the *V_{max}*, and the inhibition type.

Statistical Analysis

The biochemical data were analyzed statistically using the Graph Pad Prism software 7.04 (USA). ANOVA was performed to estimate the mean \pm standard deviation (SD), and statistical significance was considered whenever the P value was equal to or less than 0.05.

Theoretical Calculations

The density functional theory (DFT) calculations were done using the package of Gaussian 09W software with different basis sets DFT-B3LYP/6-31G for the (Asc) derivative, whereas LanL2DZ was used for the complex preparations.

RESULTS AND DISCUSSION

Biochemical Study

This study showed a significant increase in the level of amylase enzyme reaching 270 ± 5.32 U/L for patients, while the amylase activity in the control group was 105 ± 3.24 U/L. The enzyme leakage caused this increase from the pancreatic channel to the blood instead of the intestine (Tilley & Smith Jr, 2015). The amylase enzyme levels in chronic pancreatitis were lower than in acute pancreatitis but higher than normal. The normal level of acute pancreatitis was temporary and narrow for 8–72 hours, whereas the level of the enzyme was the highest in 24–30 hours from the onset of the disease. It happened in case of levels rise to 550 units and sometimes 2,000 units and continue to increase. A few days

later, the levels decreased due to the enzyme's ability to leak through the kidneys into the bloodstream. Blood samples were used to detect the efficacy of the amylase enzyme. The samples were collected from the patients in a short period to obtain a reliable diagnostic tool (Varley et al., 1988).

Purification of Amylase

Table 1 shows the purification of the amylase enzyme using several steps. Firstly, (NH₄)₂SO₄ salt 60% was used to precipitate the amylase to obtain 1.21 folds of purification, yielding 83%, and specific activity (S.A) was 10.74 IU/mg. Secondly, dialysis gave 1.99 folds of the purification process. The final yield was about 71%, and the S.A was 21.37 IU/mg. Next, ion exchange chromatography was applied to indicate one band of the purified enzyme with 4.25 folds. The final yield percentage of this step was 52%, and S.A was 37.57 IU/mg. Finally, gel filtration chromatography was applied to obtain 4.31 folds of purification, and the final yield was 35.5%, whereas the S.A was 38.18 IU/mg.

Table 1

Partial purification of amylase from the serum of type 2 diabetic patients using ammonium sulfate, ion-exchange, and gel filtration chromatography

Yield %	Folds of Purification (time)	Specific activity (IU/mg)	Total protein (mg)	Protein conc. (mg/mL)	Total activity (IU)	Activity (IU/mL)	Elute mL	Step
100	1	8.84	266.8	13.34	2360	118	20	Crude portion
83	1.21	10.74	145.92	9.12	1568	98	16	Ammonium Sulfate
71	1.99	21.37	43.23	3.93	924	84	11	Dialysis
52	4.25	37.57	8.25	1.65	310	62	5	Ionic Exchange
35.5	4.31	38.18	4.3	1.1	168	42	4	Gel filtration

Electrophoresis of the purified amylase from human red blood cells showed one protein band appearing on the SDS-PAGE at the exact molecular weight, (45.20) kDa, compared to standards (Markers) by fitting the results of log M.Wt. against relative mobility as shown in Figure 1. It represents the amylase enzyme obtained after the ion exchange and gel filtration chromatography.

Marker lane refers to proteins marker with different molecular weights (commercial). Extracted lane refers to the enzyme after extractions. +AS lane refers to the enzyme after extracting by ammonium sulfate. The lane labeled with +IXE refers to amylase enzyme after ion exchange column with impurities; Wash 1 and 2 lanes: the fractionated volumes after washing IXE column. L lane: Load of the extract to the column, +G.F Lane shows the purified amylase after the gel filtration step.

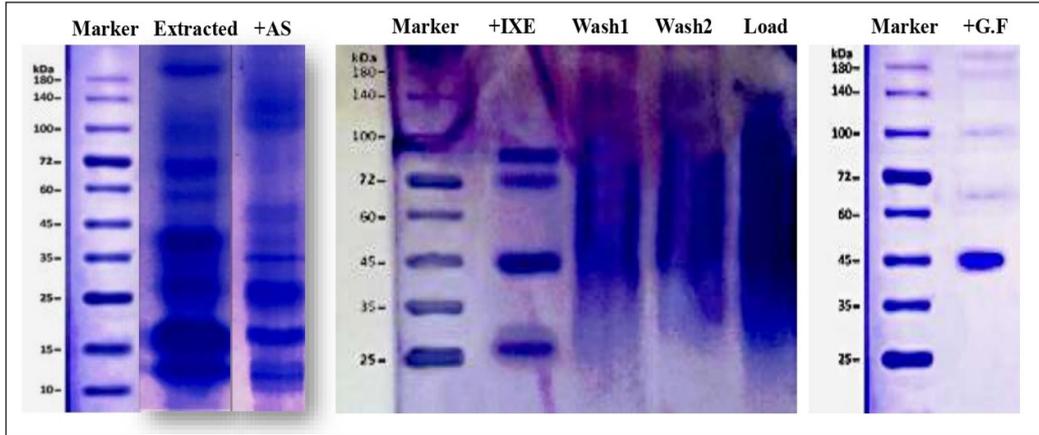


Figure 1. Electrophoresis of the purified amylase enzyme from human red blood cells with polyacrylamide gel

Kinetic Study of a Partially Purified Amylase Enzyme

The results showed an increase in the amylase activity with an increase in the substrate concentration. The V_{max} of amylase enzymes was calculated at the optimal substrate concentration of 6 mmol using the Michaelis-Menten model showing a hyperbolic curve (Figure 2). The velocity of the amylase enzyme tested by the Lineweaver–Burk model showed that K_m was 10.914 mM and V_{max} was 34.25 IU/L (Figure 3). The maximum temperature at which amylase showed bioactivity was 60 °C (Figure 4). Regarding the effect of the pH on amylase activity, the optimum pH was 7.3. We noticed that when pH

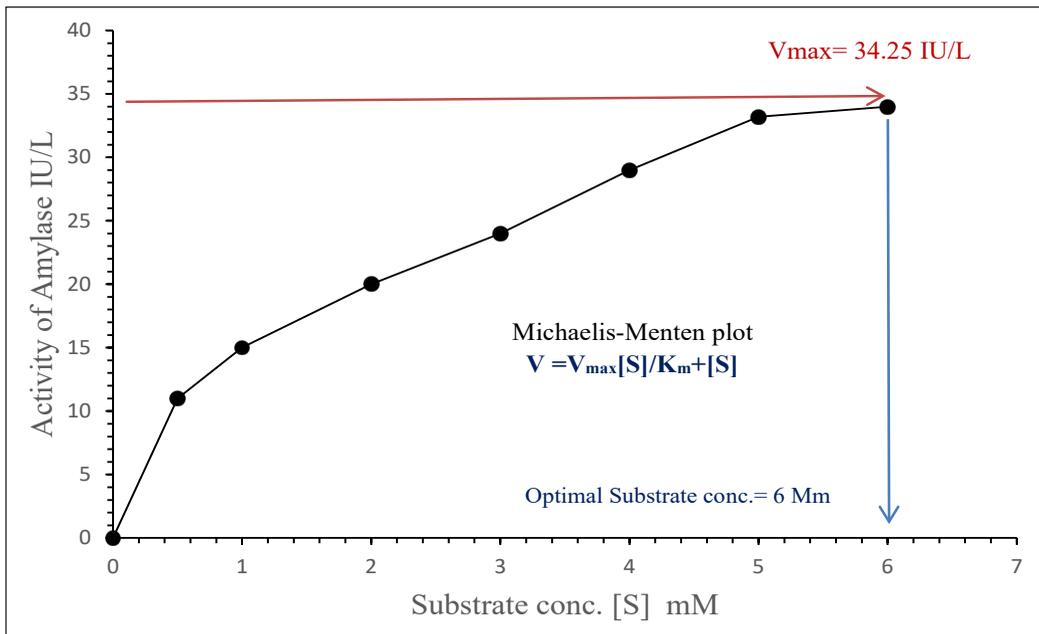


Figure 2. The relationship between the amylase activity and the substrate concentration

increases, the enzyme's activity increases until it reaches the maximum pH, whereby the enzyme could be denatured after this point (Figure 5). However, the activity was increased with time to the maximum of 37 °C after 40 minutes. We noticed that when the temperature increases, the enzyme's activity still increases until it reaches the maximum temperature for the same reason (Figure 6). These results confirmed previous studies about the maximum temperature and the best pH where amylase still shows bioactivity (Al-Qodah, 2007; Adnan, 2010; Mohhmod, 2010; Dhote et al., 2014).

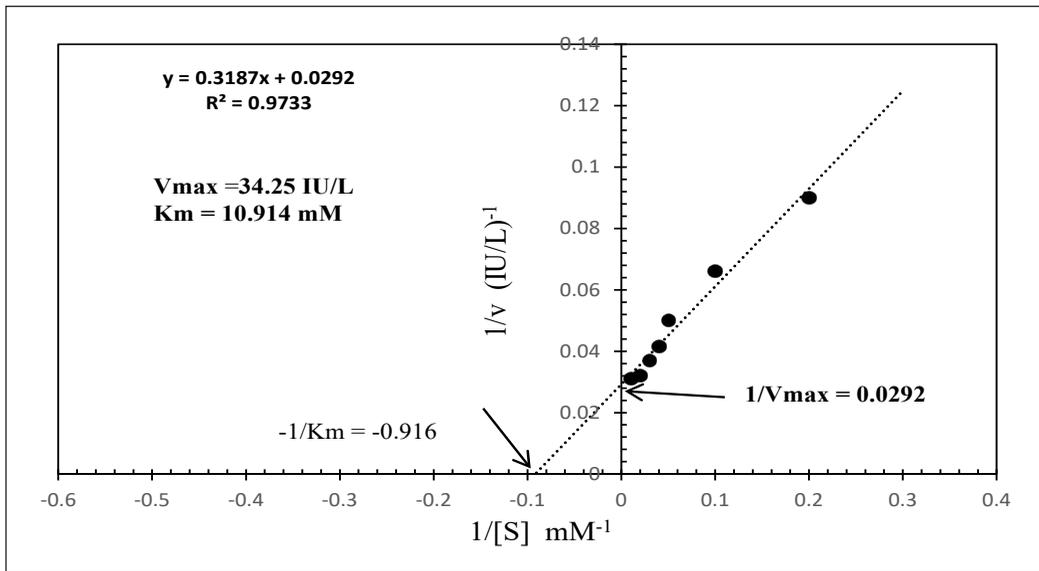


Figure 3. Lineweaver–Burk plotting $1/v$ versus $1/[S]$ of purified amylase

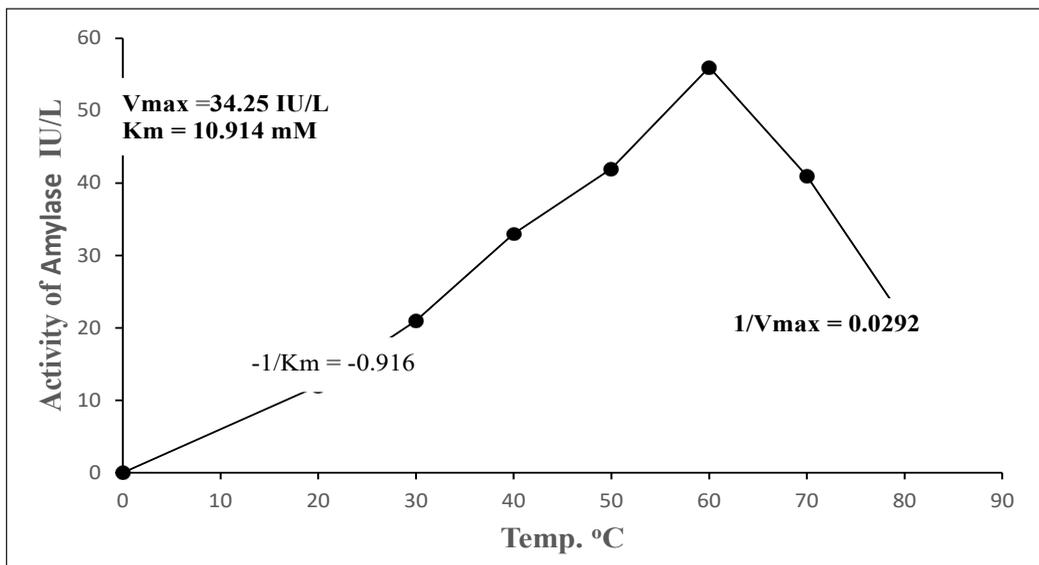


Figure 4. Effect of temperature on amylase activity

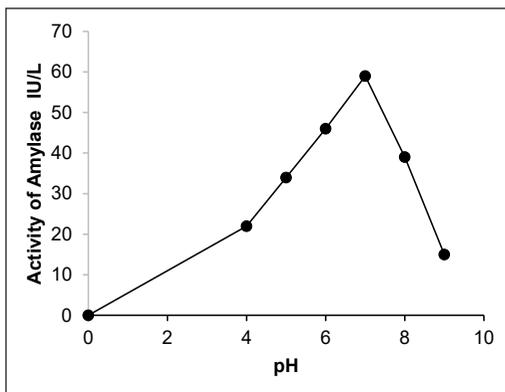


Figure 5. Effect of pH on amylase activity

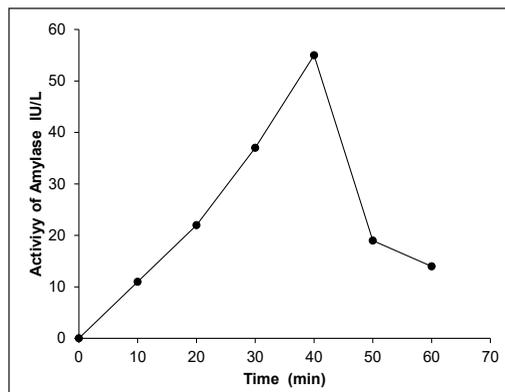


Figure 6. Incubation time of substrate with purified amylase at 25 °C and pH 7

Preparation of 5,6-O-isopropylidene-L-Ascorbic Acid

The L-ascorbic acid molecule has four hydroxyl groups; these groups are active for classical esterification (Figure 7). The synthesis of the derivative occurred in 2 and 3-positions. The first conversion into its 5,6-isopropylidene derivative happened because the carbon-6 hydroxyl group (a primary hydroxyl group) is the most reactive. The acetal is stable in alkaline conditions but readily hydrolyzed by dilute acid (Maher et al., 2017). Hence, it is very useful as a blocking agent. It was used in this work to protect the hydroxyl group at C-5 and C-6, whereas leaving the hydroxyl groups at C-2 and C-3 free for the required chemical modification. According to the literature, the L-ascorbic acid derivative was prepared from the reaction of L-ascorbic acid with acetone in acidic media (Salomon, 1963).

The FTIR spectra of the L-ascorbic acid derivative are depicted in Figure 8. The stretching bands at 3244 cm^{-1} are assigned for $\nu(\text{O-H})$, 2995 cm^{-1} for $\nu(\text{C-H aliphatic})$, 1755 cm^{-1} for $\nu(\text{C=O})$ lactone, 1664 cm^{-1} for $\nu(\text{C=C})$, and 1141 cm^{-1} for $\nu(\text{C-O})$. The calculated spectrum was performed to make a comparison with the experimental one. The obtained data are acceptable with the experimental data; the stretching of the $\nu(\text{O-H})$ appeared at 3455 , the stretching of the $\nu(\text{C-H aliphatic})$ at 3135 , the stretching of the $\nu(\text{C=O})$ lactone

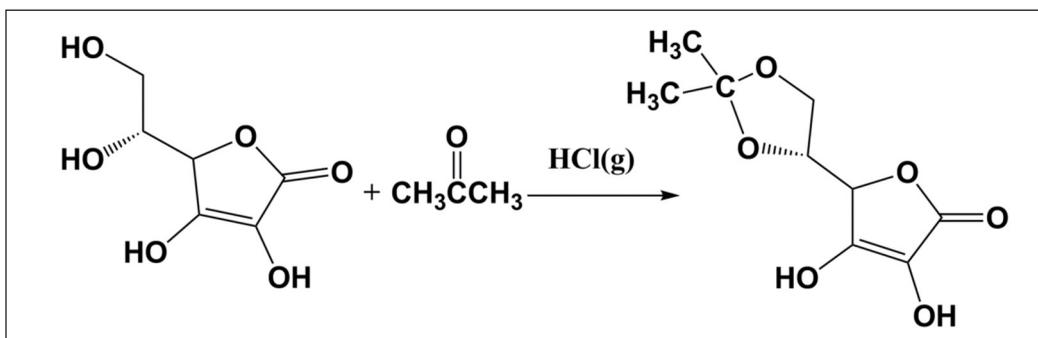


Figure 7. The preparation of L-ascorbic acid derivative

at (1888), the stretching of the $\nu(\text{C}=\text{C})$ at 1789 and finally the stretching of the $\nu(\text{C}-\text{O})$ at 1172 (Table 2).

The ¹H-NMR spectrum of L-ascorbic acid derivative Figure 9, showed a singlet signal at 11.27 ppm, which belongs to the proton of the (OH) hydroxyl group at C-2. The singlet signal at 8.46 ppm belongs to the proton of the (OH) hydroxyl group at C-3, and the doublet signal at 4.71 ppm belongs to the proton of the (H-C4). The quartet signal at

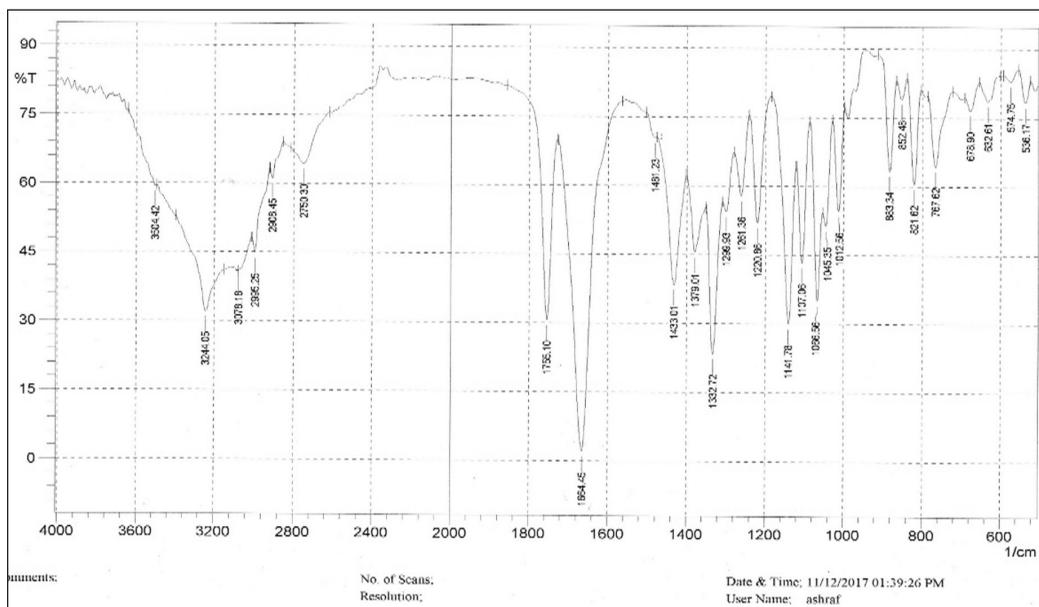


Figure 8. The experimental FTIR spectrum of L-ascorbic acid derivative

Table 2

Observed and calculated vibrational frequencies cm^{-1} for L-ascorbic derivative ligand and Pt (II) complex

Compound	Obs.	Calc.	Assignment
L-ascorbic derivative Ligand	3244	3720	$\nu(\text{O}-\text{H})$
	2995	3135	$\nu(\text{C}-\text{H})$
	1755	1888	$\nu(\text{C}=\text{O})$
	1644	1789	$\nu(\text{C}=\text{C})$
	1144	1172	$\nu(\text{C}-\text{O})$
	-	-	$\nu(\text{O}-\text{H})$
	3251	3591	$\nu(\text{N}-\text{H})$
Pt-ASCA complex	2981	3127	$\nu(\text{C}-\text{H})$
	1743	1774	$\nu(\text{C}=\text{O})$
	1612	1664	$\nu(\text{C}=\text{C})$
	1139	1334	$\nu(\text{C}-\text{O})$
	574	571	$\nu(\text{M}-\text{O})$
468	472	$\nu(\text{M}-\text{N})$	

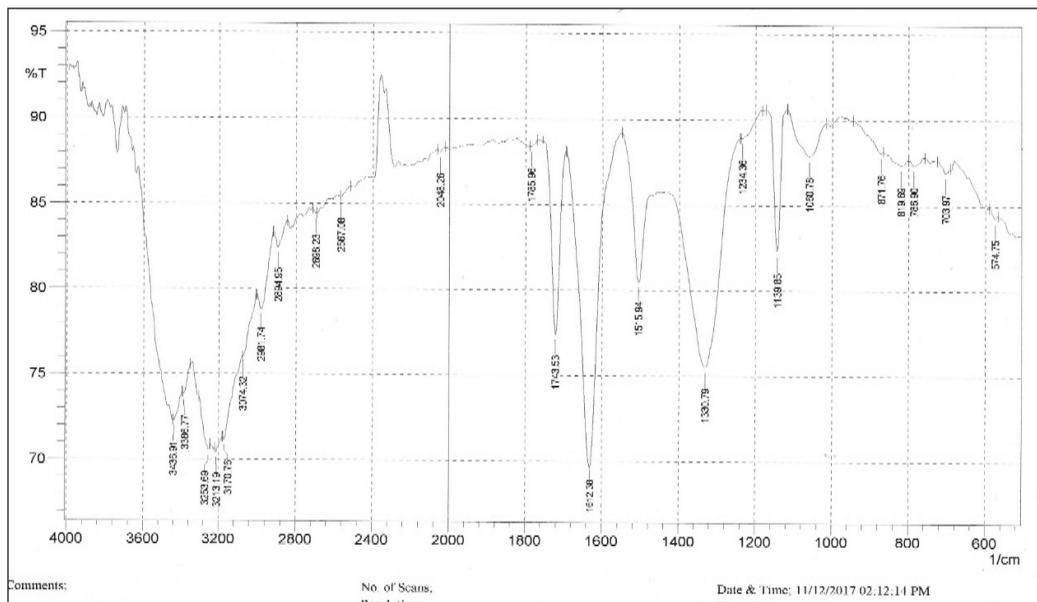


Figure 9. The ¹H-NMR spectrum of L-ascorbic acid derivative

4.26 ppm belongs to the proton of the (H-C5), and the doublet signal at 4.11 ppm belongs to the proton of the (H-C6). The doublet signal at 3.89 ppm belongs to the proton of the (H-C6), and the singlet signal at 1.26 ppm belongs to the proton of the aliphatic 2(CH₃) (Silverstein et al., 2005; Mistry, 2009). These data have been compared with the calculated data, which showed an acceptable difference between the two results. The deviation between the obtained data may be attributed to the different ways to get the data.

Synthesis of the Complex Cis-[Pt(Asc)(NH₃)₂]

The complex was prepared by the reaction of the ligand with cis-[Pt(Cl)₂(NH₃)₂] solution, as shown in Figure 10. The FTIR spectrum of the cis-[Pt(Asc)(NH₃)₂] complex is depicted in Figure 11. The spectrum showed stretching bands at 3213, 3253 cm⁻¹ assigned for ν(NH₃),

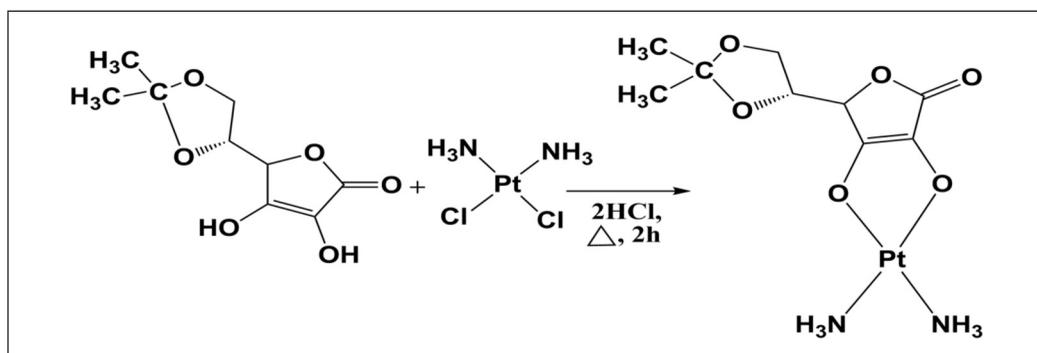


Figure 10. The reaction of the L-ascorbic acid derivative ligand with cis-[Pt(Cl)₂(NH₃)₂]

2981 cm⁻¹ for ν(C-H aliphatic), 1743 cm⁻¹ for ν(C=O) lactone, 1612 cm⁻¹ for ν(C=C), 1139 cm⁻¹ for ν(C-O). The ¹H-NMR spectrum of the cis-[Pt(Asc)(NH₃)₂] showed a singlet signal at 4.35 ppm, which belongs to the protons of the (NH₃) (Figure 12). The doublet signal at 4.21 ppm belongs to the proton of the (H-C4), and a quartet signal at 3.9 ppm belongs to the proton of the (H-C5). The doublet signal at (3.78-3.7 ppm) belongs to the protons of the (2H-C6), and the singlet signal at 1.25 ppm belongs to the protons of the aliphatic 2(CH₃) (Silverstein et al., 2005; Mistry, 2009).

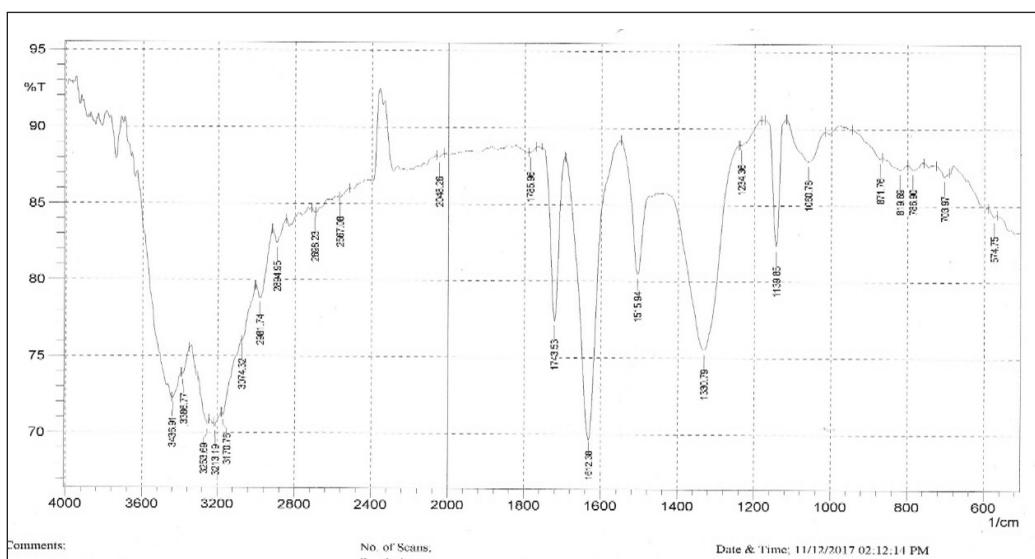


Figure 11. The experimental FTIR spectrum of the complex cis-[Pt(Asc)(NH₃)₂]

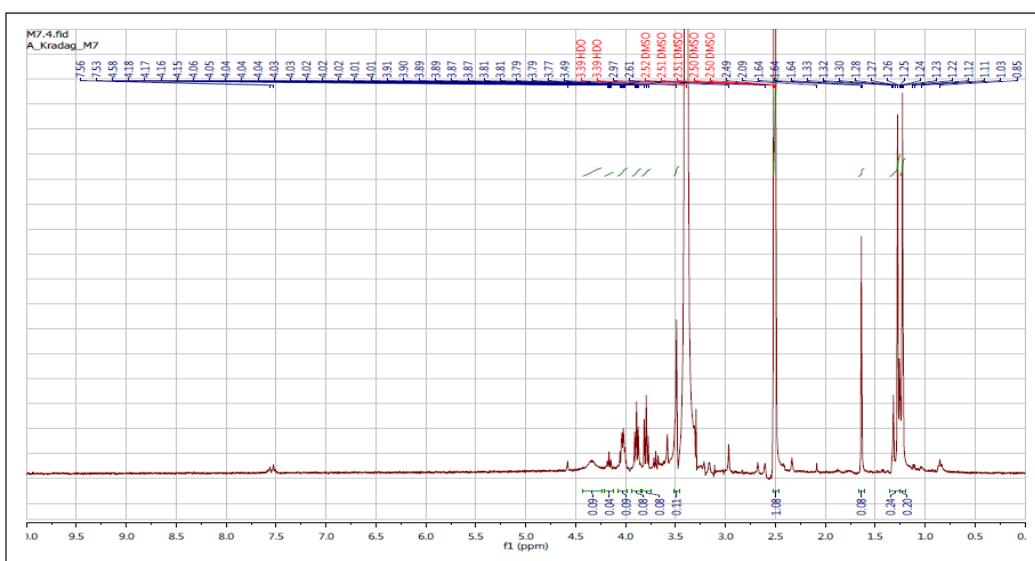


Figure 12. The ¹H-NMR spectrum of the complex cis-[Pt(Asc)(NH₃)₂]

DFT Calculations

The molecular geometries of the L-ascorbic acid derivative, the cisplatin, and the prepared Pt(II) complex were fully optimized using G 09W (Chattaraj & Poddar, 1999; Caruso et al., 2012; Sabounchei et al., 2015; Ali et al., 2019). The optimized structures and natural bond charges are presented in Figure 13. Structural data are given in Table 3. The angles around the Pt(II) atom ranged from 81.82° to 105.83°, and the two dihedral angles are 172.43° to 169.53°. These angles deviated from the perfect square planar geometry, suggesting the distorted structure for the Pt(II) complex, whereby the O-Pt-O bond angle was 87.71°. Compared with the literature on platinum(II) complexes, the bond angle was 84.36° in [Pt(N, N-dimethyl-N'-(2H-1,2,4-triazole-3-yl) formamidine)ox].2H₂O (Soliman et al., 2016), and the N-Pt-N bond angle was 105.83. The Pt-O bond lengths found 2.020-2.042 Å

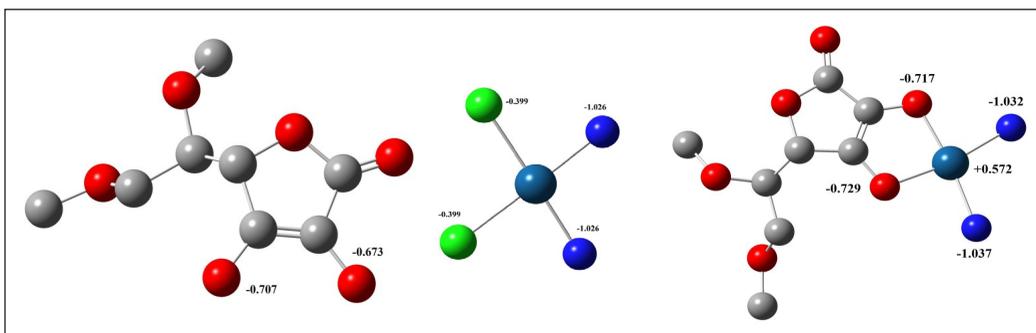


Figure 13. The optimized structures and natural bond charges of the studied compounds. The hydrogen atoms have been omitted for simplicity

Table 3

Bond lengths (Å), bond angles (°), and dihedral angles (°) of optimized cisplatin and Pt(II) complex using the DFT/B3LYP/Lan2DZ basis set

Cisplatin °A	Bond Lengths	Bond Angles °
Pt(1)-N(2)	2.110	N(2)-Pt(1)-N(5)
Pt(1)-N(5)	2.110	Cl(10)-Pt(1)-Cl(11)
Pt(1)-Cl(10)	2.410	N(2)-Pt(1)-Cl(10)
Pt(1)-Cl(11)	2.410	N(5)-Pt(1)-Cl(11)
		N(2)-Pt(1)-Cl(11)
		N(5)-Pt(1)-Cl(10)
Pt-ASCA complex		
Pt(25)-O(1)	2.042	O(1)-Pt(25)-O(10)
Pt(25)-O(10)	2.020	N(26)-Pt(25)-N(29)
Pt(25)-N(26)	2.100	N(26)-Pt(25)-O(10)
Pt(25)-N(29)	2.115	N(29)-Pt(25)-O(1)
		N(29)-Pt(25)-O(10)
		N(26)-Pt(25)-O(1)

are consistent with the range of reported values of 2.016 (Bakalova et al., 2009; Mansour, 2013; Soliman et al., 2016). The charge densities distributed on L-ascorbic acid derivative ligand oxygen atoms explain their expected donor properties (Chattaraj & Poddar, 1999). Platinum has a positive charge of +0.572 and acts as the acceptor. The back donation of electrons from Pt(II) to the L-ascorbic acid derivative ligand is assumed from the increase of the negative charges on the oxygen atoms. It is an MLCT from the Pt(II) to the π^* orbitals of the L-ascorbic acid derivative ligand. The Pt(II) complex 15.788 Debye is more polarized than the L-ascorbic acid derivative ligand 6.489 Debye, whereas, for cisplatin 11.719, the electronic energy of the Pt(II) complex is (-994.402 a.u.). It confirmed that the complex is more stable than the ligand (-763.372 a.u.) and the cisplatin 262.298 a.u. (Sabounchei et al., 2015). HOMO and LUMO energies of the L-ascorbic acid derivative, the cisplatin, and the Pt(II) complex are given in Table 4. The hardness is ($\eta=(I-A)/2$), A is the electron affinity, I is the ionization energy, and $(I-A)$ is the difference between the HOMO and the LUMO orbitals (Sabounchei et al., 2015). The higher the value of the gap between the HOMO and the LUMO, the harder the molecule. The η values and ΔE are given in Table 4. The transition is easier in the Pt(II) complex than in the ligand and the cisplatin; the ΔE of the complex is 0.133, the ligand is 0.205, and the cisplatin is 0.161 (McGuire et al., 1984; Bakalova et al., 2009; Mansour, 2013; Elghalban et al., 2014; Sabounchei et al., 2015; Soliman et al., 2016; Ali et al., 2019) and hence the complex is softer ($\eta=0.006$) than in the ligand 0.102 and the cisplatin 0.080 (Chattaraj & Poddar, 1999). The electronic energy and the dipole moments are tabulated in Table 4. The plots of the HOMO and the LUMO for the complex and the L-ascorbic acid derivative ligand are presented in Figure 14.

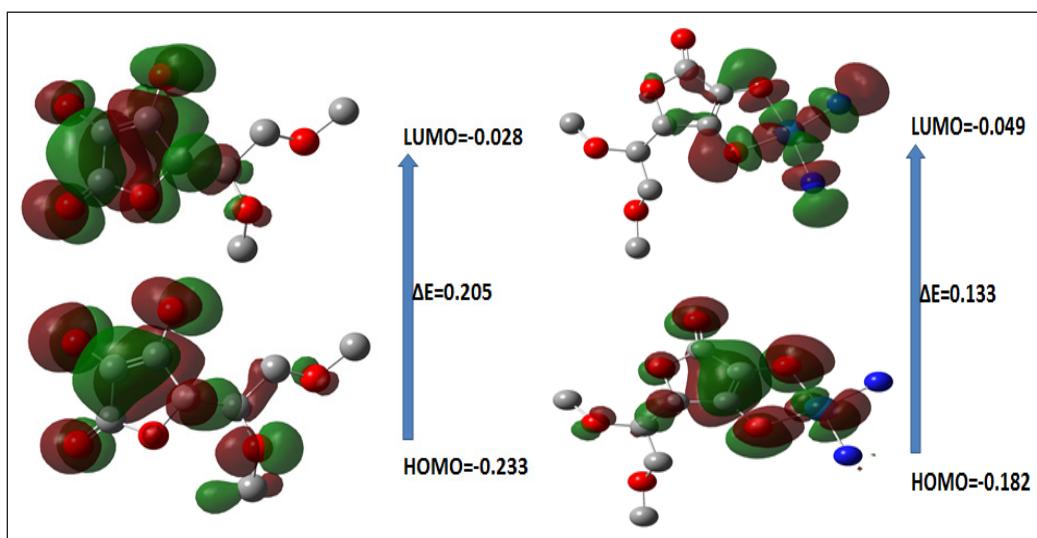


Figure 14. The isodensity surface plots of the HOMO and the LUMO for the Pt(II) complex (right) and the L-ascorbic acid derivative ligand (left). The hydrogen atoms have been omitted for simplicity.

Table 4

The calculated quantum chemical parameters of the L-ascorbic derivative ligand, cisplatin, and Pt(II) complex

Compound	HOMO	LUMO	η	ΔE	Total energy a.u	D.M Debye
L-ascorbic derivative Ligand	-0.233	-0.028	0.102	0.205	-763.37	6.45
Cisplatin	-0.230	-0.069	0.080	0.161	-262.29	11.71
Pt-ASCA complex	-0.182	-0.049	0.066	0.133	-994.40	15.78

The electronic energy of the complex is (-994.402 a.u.), indicating that the complex is more stable than the ligand (-763.372 a.u.) and the cisplatin (-262.29 a.u.) (Chattaraj & Poddar, 1999). Analysis of the natural bond orbital for Pt(II) complex showed the electronic configuration of Pt to be [core] $6s^{0.54} 5d^{8.62} 6p^{0.08} 6d^{0.01} 7p^{0.18}$, core = 67.993, valence = 9.418 and Rydberg = 0.016, which give total electrons = 77.428 and the charge on Pt is +0.571e (Bakalova et al., 2009). The 5d orbitals are d_{xz} 1.814; d_{xy} 1.691; d_{yz} 1.930; $d_{x^2-y^2}$ 1.296 and dz^2 1.884. The 5d-electron is (8.615) Table 5. The molecular electrostatic potential (MEP) of the L-ascorbic acid derivative ligand, the Pt(II) complex, and cisplatin are shown in Figure 15. The red regions represent (electrophilic reactivity) and the blue regions (nucleophilic reactivity), respectively. The oxygen atoms (red negative) in the L-ascorbic derivative ligand confirmed the reactive sites for the electrophilic attack (McGuire et al.,

Table 5

The electronic configuration, populations of 4d-orbitals and platinum, oxygen, and chloride charge in the cisplatin and Pt (II)-L-ascorbic acid complexes

Compound	Electronic arrangement	d_{xy}	d_{xz}	d_{yz}	$d_{x^2-y^2}$	dz^2	Rydberg electrons	Charge Pt	Charge O and Cl
Cisplatin	[core] $6s^{0.57} 5d^{8.80} 6p^{0.22} 6d^{0.01} 7p^{0.16}$	0.963	1.986	1.990	1.980	1.874	0.019	+0.251	-0.399, Cl -0.399 Cl
Pt(II)-ASCD complex	[core] $6s^{0.54} 5d^{8.62} 6p^{0.08} 6d^{0.01} 7p^{0.18}$	1.691	1.814	1.930	1.296	1.884	0.016	+0.571	-0.927 O -0.917 O

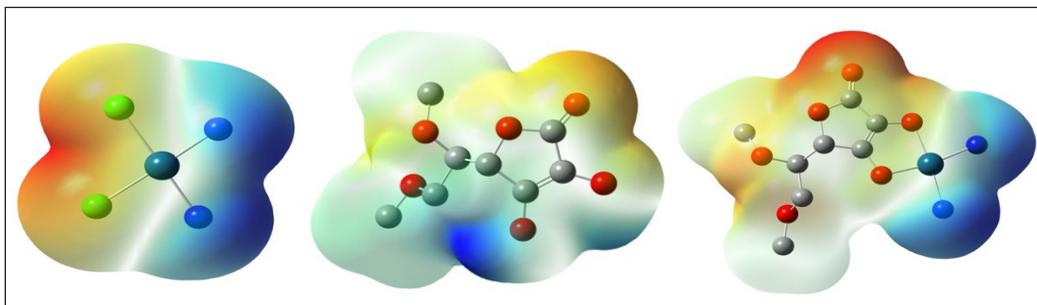


Figure 15. The MEP (molecular electrostatic potential) of the L-ascorbic acid derivative ligand, the cisplatin, and the Pt(II) complex. The hydrogen atoms have been omitted for simplicity.

1984; Bakalova et al., 2009; Mansour, 2013; Elghalban et al., 2014; Sabounchei et al., 2015; Soliman et al., 2016; Ali et al., 2019).

Effect of the New Inhibitors

The effect of the prepared Pt(II) complex on the purified amylase enzyme activity is illustrated in Table 6. The newly prepared Pt(II) complex’s inhibitory effect on amylase enzyme was examined *in vitro*. The results showed that the inhibitory effect of the amylase enzyme increases with the increase of complex concentrations. The activity (without inhibitor) was 78.68± 6.62U/L. The activity with the complex (inhibitors) ranged from 23.81±2.29 to 76.45±5.76 U/L. At the concentrations of 5 × 10⁻⁶ mg/25 ml and 5 × 10⁻⁷ mg/25 ml, it was found that the activity

Table 6
The inhibitory effect of the compound on amylase activity

Concentrations	Amylase Activity U/L Compounds
Normal/without any concentrations of complex	78.68±6.62
0.5 mg/25ml	23.81±2.29
0.5 × 10 ⁻¹ mg/25ml	34.13±2.98
0.5 × 10 ⁻² mg/25ml	46.21±3.38
0.5 × 10 ⁻³ mg/25ml	52.13±4.11
0.5 × 10 ⁻⁴ mg/25ml	60.24±4.81
0.5 × 10 ⁻⁵ mg/25ml	66.40±5.27
0.5 × 10 ⁻⁶ mg/25ml	76.45±5.76

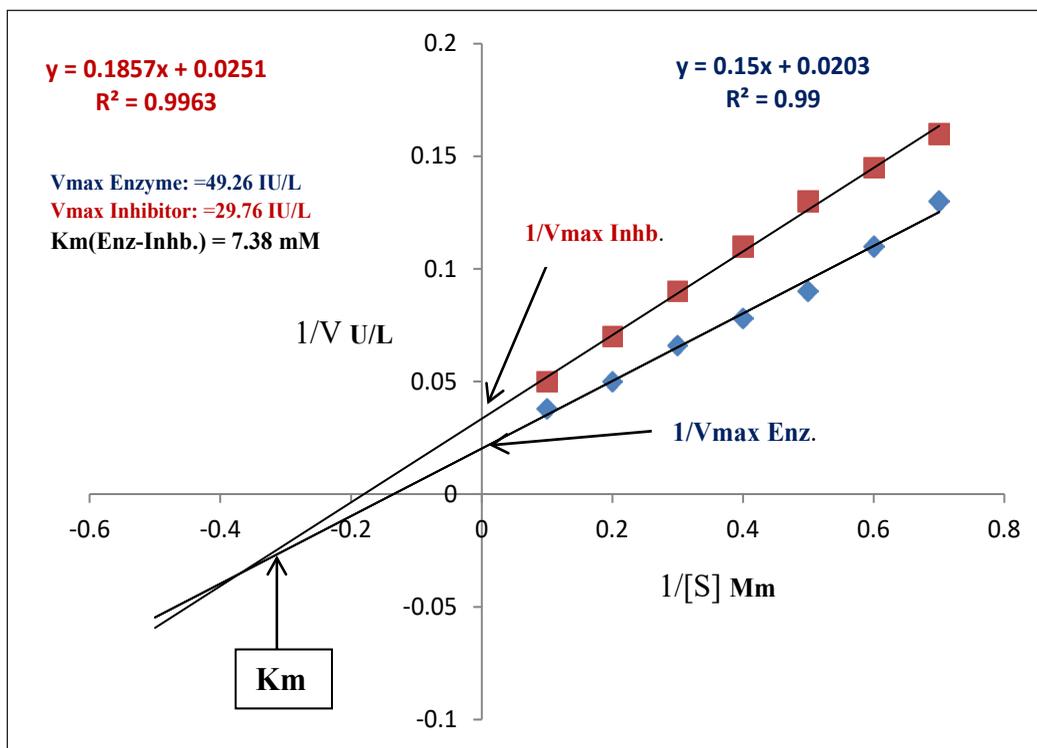


Figure 16. Lineweaver–Burk plot fitting showed that the prepared complex behaves as non-competitive inhibitors. The blue and the red points refer to the inhibitor’s absence and presence, respectively

of the enzyme reached the normal value because of the multi dilutions of the concentration of the complex. These results are in agreement with the literature data (Kalita et al., 2018; Oboh et al., 2015). The active site of the amylase enzyme contains amino acids, aspartic acids, glutamic acids, and methionine, which can bind the platinum found in the structure of the complex (inhibitor), thus impeding the ground of the active site with the foundation and leading to inhibition of the amylase activity (Yang et al., 2012). The type of inhibition was shown as non-competitive (K_m value was fixed: 7.38 mM, and the V_{max} was decreased (V_{max} without inhibitor 49.26 IU/L) (V_{max} with inhibitor 29.76 IU/L) as in Figure 16.

CONCLUSION

We have successfully purified amylase from the serum of patients with T2DM. Several spectroscopic methods prepared and diagnosed inorganic complex derived from vitamin C. Also, a comparison between the experimental and theoretical data was performed. The DFT study of the prepared complex gave a useful explanation of the complex and its stability. Furthermore, the Pt(II) complex showed an inhibitory effect on the amylase activity, which needs further investigation in vivo as a medical drug.

ACKNOWLEDGMENT

Our deep grateful thanks to Department of Chemistry, College of Science. Anbar University for cooperating with us to complete the experimental and to Department of Applied Chemistry, College of Applied Sciences. University of Fallujah for cooperating with us to complete the calculated part.

REFERENCES

- Adnan, F. (2010). Kinetics and thermodynamic studies of alpha amylase from *Bacillus licheniformis* mutant. *Pakistan Journal of Botany*, 42(5), 3507-3516.
- Albuquerque, T. L. D., Peirce, S., Rueda, N., Marzocchella, A., Gonçalves, L. R. B., Rocha, M. V. P., & Fernandez-Lafuente, R. (2016). Ion exchange of β -galactosidase: the effect of the immobilization pH on enzyme stability. *Process Biochemistry*, 51(7), 875–880. <https://doi.org/10.1016/j.procbio.2016.03.014>
- Al-Qodah, Z. (2007). Determination of kinetic parameters of α -amylase producing thermophile *Bacillus sphaericus*. *African Journal of Biotechnology*, 6(6), 699-706.
- Ali, R., Bulat, K. H. K., Azmi, A. A., & Anuar, S. T. (2019). Theoretical approach of dft b3LYP/6-31G (d, p) on evaluating the performance of tert-butylhydroquinone and free fatty acids in inhibiting the oxidation of palm olein. *Journal of Oil Palm Research*, 31(1), 122-129. <https://doi.org/10.21894/JOPR.2019.0005>
- Ard, D., Tettey, N. S., & Feresu, S. (2020). The influence of family history of type 2 diabetes mellitus on positive health behavior changes among African Americans. *International Journal of Chronic Diseases*, 2022, Article 8016542. <https://doi.org/10.1155/2020/8016542>

- Association, A. D. (2014). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 37(Supplement 1), S81-S90. <https://doi.org/10.2337/dc14-S081>
- Bakalova, A., Varbanov, H., Stanchev, S., Ivanov, D., & Jensen, F. (2009). DFT study of the structure and spectral behavior of new pt (II) complexes with 5-methyl-5 (4-pyridyl) hydantoin. *International Journal of Quantum Chemistry*, 109(4), 826-836. <https://doi.org/10.1002/qua.21890>
- Bergamini, P., Marchesi, E., Bertolasi, V., Fogagnolo, M., Scarpantonio, L., Manfredini, S., Vertuani, S., & Canella, A. (2008). New coordination modes of L-ascorbic acid and dehydro-L-ascorbic acid as dianionic chelating ligand for platinum. *European Journal of Inorganic Chemistry*, 2008(4), 529-537. <https://doi.org/10.1002/ejic.200700808>
- Bradshaw, M. P., Barril, C., Clark, A. C., Prenzler, P. D., & Scollary, G. R. (2011). Ascorbic acid: A review of its chemistry and reactivity in relation to a wine environment. *Critical Reviews in Food Science and Nutrition*, 51(6), 479-498. <https://doi.org/10.1080/10408391003690559>
- Burhan, A., Nisa, U., Gökhan, C., Ömer, C., Ashabil, A., & Osman, G. (2003). Enzymatic properties of a novel thermostable, thermophilic, alkaline and chelator resistant amylase from an alkaliphilic *Bacillus* sp. isolate ANT-6. *Process Biochemistry*, 38(10), 1397-1403. [https://doi.org/10.1016/S0032-9592\(03\)00037-2](https://doi.org/10.1016/S0032-9592(03)00037-2)
- Caruso, F., Rossi, M., Benson, A., Opazo, C., Freedman, D., Monti, E., Gariboldi, M. B., Shaulky, J., Marchetti, F., & Pettinari, R. (2012). Ruthenium-arene complexes of curcumin: X-ray and density functional theory structure, synthesis, and spectroscopic characterization, in vitro antitumor activity, and DNA docking studies of (p-cymene) Ru (curcuminato) chloro. *Journal of Medicinal Chemistry*, 55(3), 1072-1081. <https://doi.org/10.1021/jm200912j>
- Chattaraj, P. K., & Poddar, A. (1999). Molecular reactivity in the ground and excited electronic states through density-dependent local and global reactivity parameters. *The Journal of Physical Chemistry A*, 103(43), 8691-8699. <https://doi.org/10.1021/JP991214+>
- Chen, L., Strohmeier, V., He, Z., Deshpande, M., Catalan-Dibene, J., Durum, S. K., Moran, T. M., Kraus, T., Xiong, H., & Faith, J. J. (2019). Interleukin 22 disrupts pancreatic function in newborn mice expressing IL-23. *Nature Communications*, 10, Article 4517. <https://doi.org/10.1038/s41467-019-12540-8>
- Czarnomysy, R., Surazyński, A., Muszynska, A., Gornowicz, A., Bielawska, A., & Bielawski, K. (2018). A novel series of pyrazole-platinum (II) complexes as potential anti-cancer agents that induce cell cycle arrest and apoptosis in breast cancer cells. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 33(1), 1006-1023. <https://doi.org/10.1080/14756366.2018.1471687>
- Dhote, Y. S., Moharil, M. P., & Jadhav, P. V. (2014). Isolation and characterization of amylase inhibitor from alkalophilic bacteria isolated from lonar crater and its insecticidal protein producing ability. *Biosciences, Biotechnology Research Asia*, 11(1), 329-333.
- Duong-Ly, K. C., & Gabelli, S. B. (2014). Salting out of proteins using ammonium sulfate precipitation. *Methods in Enzymology*, 541, 85-94. <https://doi.org/10.1016/B978-0-12-420119-4.00007-0>
- Egan, A. M., & Dinneen, S. F. (2019). What is diabetes? *Medicine*, 47(1), 1-4. <https://doi.org/10.1016/J.MPMED.2018.10.002>

- Elghalban, M. G., El Defarwy, A. M., Shah, R. K., & Morsi, M. A. (2014). α -Furil dioxime: DFT exploration and its experimental application to the determination of palladium by square wave voltammetry. *International Journal of Electrochemical Science*, 9, 2379-2396.
- Halim, M., & Halim, A. (2019). The effects of inflammation, aging and oxidative stress on the pathogenesis of diabetes mellitus (type 2 diabetes). *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 13(2), 1165-1172. <https://doi.org/10.1016/j.dsx.2019.01.040>
- Hames, B. D. (1998). *Gel electrophoresis of proteins: A practical approach* (3rd Ed.). OUP Oxford.
- Hollis, L. S., Stern, E. W., Amundsen, A. R., Miller, A. V., & Doran, S. L. (1987). Platinum complexes of vitamin C. NMR studies on the solution chemistry of cis-platinum (diamine)(ascorbate) complexes. *Journal of the American Chemical Society*, 109(12), 3596-3602. <https://doi.org/10.1021/ja00246a016>
- Hong, H. R., Oh, Y. I., Kim, Y. J., & Seo, K. W. (2019). Salivary alpha-amylase as a stress biomarker in diseased dogs. *Journal of Veterinary Science*, 20(5), Article e46. <https://doi.org/10.4142/jvs.2019.20.e46>
- Kalita, D., Holm, D. G., LaBarbera, D. V., Petrash, J. M., & Jayanty, S. S. (2018). Inhibition of α -glucosidase, α -amylase, and aldose reductase by potato polyphenolic compounds. *PLoS One*, 13(1), Article e0191025. <https://doi.org/10.1371/journal.pone.0191025>
- Kumar, S., & Chakravarty, S. (2018). Amylases. In *Enzymes in Human and Animal Nutrition* (pp. 163-180). Elsevier.
- López, J. L., Alvarez, F., Príncipe, A., Salas, M. E., Lozano, M. J., Draghi, W. O., Jofré, E., & Lagares, A. (2018). Isolation, taxonomic analysis, and phenotypic characterization of bacterial endophytes present in alfalfa (*Medicago sativa*) seeds. *Journal of Biotechnology*, 267, 55-62. <https://doi.org/10.1016/j.jbiotec.2017.12.020>
- Maher, F. T., Mukhlis, A. J., Abuod, A. I., & Najji, N. A. (2017). *In vivo* study of compounds 3-(acetyl Salicyloyl)-5, 6-O-isopropylidene-L-ascorbic acid, 2, 3-(acetyl Salicyloyl)-5, 6-O-isopropylidene-L-ascorbic acid and 2, 3, 5, 6-Tetra (acetyl Salicyloyl)-L-ascorbic acid. *Ibn Al-Haitham Journal for Pure and Applied Science*, 22(2), 1-7.
- Mahmood, N. (2016). A review of α -amylase inhibitors on weight loss and glycemic control in pathological state such as obesity and diabetes. *Comparative Clinical Pathology*, 25, 1253-1264. <https://doi.org/10.1007/s00580-014-1967-x>
- Mansour, A. M. (2013). Coordination behavior of sulfamethazine drug towards Ru (III) and Pt (II) ions: Synthesis, spectral, DFT, magnetic, electrochemical and biological activity studies. *Inorganica Chimica Acta*, 394, 436-445. <https://doi.org/10.1016/j.ica.2012.08.025>
- McGuire, J. P., Friedman, M. E., & McAuliffe, C. A. (1984). Studies of enzyme inhibition. The interaction of some platinum (II) complexes with fumarase and malate dehydrogenase. *Inorganica Chimica Acta*, 91(3), 161-165. [https://doi.org/10.1016/S0020-1693\(00\)81806-X](https://doi.org/10.1016/S0020-1693(00)81806-X)
- Mistry, B. D. (2009). *A handbook of spectroscopic data: Chemistry*. Oxford.
- Mohhmod, R. J. (2010). Kinetics of α -amylase enzyme in human serum. *Journal of Kerbala University*, 8(3), 237-244.

- Oboh, G., Agunloye, O. M., Adefegha, S. A., Akinyemi, A. J., & Ademiluyi, A. O. (2015). Caffeic and chlorogenic acids inhibit key enzymes linked to type 2 diabetes (in vitro): A comparative study. *Journal of Basic and Clinical Physiology and Pharmacology*, 26(2), 165-170. <https://doi.org/10.1515/jbcpp-2013-0141>
- Oboh, G., Isaac, A. T., Akinyemi, A. J., & Ajani, R. A. (2014). Inhibition of key enzymes linked to type 2 diabetes and sodium nitroprusside induced lipid peroxidation in rats' pancreas by phenolic extracts of avocado pear leaves and fruit. *International Journal of Biomedical Science*, 10(3), 208-216.
- Romero-Canelon, I., & Sadler, P. J. (2013). Next-generation metal anticancer complexes: multitargeting via redox modulation. *Inorganic Chemistry*, 52(21), 12276-12291. <https://doi.org/10.1021/ic400835n>
- Sabounchei, S. J., Shahriary, P., Salehzadeh, S., Gholiee, Y., Nematollahi, D., Chehregani, A., Amani, A., & Afsartala, Z. (2015). Pd (II) and Pd (IV) complexes with 5-methyl-5-(4-pyridyl) hydantoin: Synthesis, physicochemical, theoretical, and pharmacological investigation. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 135, 1019-1031. <https://doi.org/10.1016/j.saa.2014.08.002>
- Salomon, L. L. (1963). Preparation of 5, 6-O-isopropylidene-L-ascorbic acid. *Experientia*, 19(12), 619-620. <https://doi.org/10.1007/BF02151276>
- Silano, V., Barat Baviera, J. M., Bolognesi, C., Cocconcelli, P. S., Crebelli, R., Gott, D. M., Grob, K., Lambré, C., & Lampi, E. (2020). Safety evaluation of the food enzyme α -amylase from the genetically modified *Bacillus licheniformis* strain DP-Dzb45. *EFSA Journal*, 17(6), Article e06311. <https://doi.org/10.2903/j.efsa.2019.5738>
- Silverstein, R. M., Webster, F. X., & Kiemle, D. J. (2005). *Spectrometric Identification of Organic Compounds*, 7th Edition. John Wiley & Sons. Inc.
- Soliman, A. A., Alajrawy, O. I., Attaby, F. A., & Linert, W. (2016). New binary and ternary platinum (II) formamidine complexes: Synthesis, characterization, structural studies and in-vitro antitumor activity. *Journal of Molecular Structure*, 1115, 17-32. <https://doi.org/10.1016/J.MOLSTRUC.2016.02.073>
- Tilley, L. P., & Smith Jr, F. W. K. (2015). *Blackwell's five-minute veterinary consult: Canine and feline*. John Wiley & Sons.
- Varley, H., Gowenlock, A. H., McMurray, J. R., & McLauchlan, D. M. (1988). *Varley's practical clinical biochemistry*. Heinemann Medical Books.
- Whitcomb, D. C., & Lowe, M. E. (2007). Human pancreatic digestive enzymes. *Digestive Diseases and Sciences*, 52(1), 1-17. <https://doi.org/10.1007/s10620-006-9589-z>.
- Yalçın, G. (1995). Studies on cis-DDP,[Pt (Dach)(MePhSO) Cl]⁺ and [Pt (NH₃)₂ (N-Py) Cl]⁺ binding to fumarase. *Drug Metabolism and Drug Interactions*, 12(2), 105-116. <https://doi.org/10.1515/DMDI.1995.12.2.105>.
- Yang, H., Liu, L., Wang, M., Li, J., Wang, N. S., Du, G., & Chen, J. (2012). Structure-based engineering of methionine residues in the catalytic cores of alkaline amylase from *Alkalimonas amylolytica* for improved oxidative stability. *Applied and Environmental Microbiology*, 78(21), 7519-7526. <https://doi.org/10.1128/AEM.01307-12>

- Yu, H. (2019). *Kiwifruit effects on starch digestion by salivary amylase under simulated gastric conditions* (Master dissertation). University of Otago, New Zealand. <http://hdl.handle.net/10523/9448>
- Zhao, Y., Ouyang, X., Chen, J., Zhao, L., & Qiu, X. (2018). Separation of aromatic monomers from oxidatively depolymerized products of lignin by combining Sephadex and silica gel column chromatography. *Separation and Purification Technology*, 191, 250-256. <https://doi.org/10.1016/j.seppur.2017.09.039>
- Zümreoglu-Karan, B. (2006). The coordination chemistry of vitamin C: An overview. *Coordination Chemistry Reviews*, 250(17-18), 2295-2307. <https://doi.org/10.1016/j.ccr.2006.03.002>

Comparative Computational Study of Double Rotating Cylinder Embedded on Selig S1223 Aerofoil and Flat Plate for High Altitude Platform

Hidayatullah Mohammad Ali*, Azmin Shakrine Mohd Rafie, Mohd Faisal Abdul Hamid and Syaril Azrad Md. Ali

Department of Aerospace Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

The high-altitude platform was built as an alternative approach to address the weakness of the terrestrial and satellite communication networks. It can be an aircraft or balloon positioned 20 to 50 km above the earth's atmosphere. The use of the Magnus effect was not noticeable in the production of the high-altitude platform, while past research study has denoted its aerodynamic performance in generating greater lift and stall angle delay, which would be beneficial in creating such a flying device. This research delineates the proposed designs using the computational fluid dynamics approach utilizing ANSYS WORKBENCH 2019 software. The embedment of the rotating cylinder onto the design would best portray the use of the Magnus effect in generating higher lift coefficients with probable delay in stall angle. Hereby, the design of embedding rotating cylinder onto Selig S1223 aerofoil and the flat plate is proposed to test their aerodynamic performances for high altitude platform purposes. Here, Fluent fluid flow analysis was simulated for 500 RPM and 1000 RPM momentum injection with free stream velocities from 5 m/s to 30 m/s for different angles of attack of 0 to 20 degrees. The analysis has resulted in a greater impact on its lift coefficient and stall angle delay of about 39% and 53% enhancement for modified aerofoil while showing 128% and 204% betterment for modified flat plate than their respective unmodified model. Therefore, it is perceived that the CyFlaP has better stability yet is simplistic in a design suitable for HAP application.

Keywords: Computational fluid dynamic, flat plate, high altitude platform, rotating cylinder, Selig S1223 aerofoil

ARTICLE INFO

Article history:

Received: 20 November 2021

Accepted: 07 March 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.26>

E-mail addresses:

hidayatmaddali@gmail.com (Hidayatullah Mohammad Ali)

shakrine@upm.edu.my (Azmin Shakrine Mohd Rafie)

mohd_faisal@upm.edu.my (Mohd Faisal Abdul Hamid)

syaril@upm.edu.my (Syaril Azrad Md. Ali)

* Corresponding author

INTRODUCTION

The high-altitude platform (HAP) has attracted constant growth of interest internationally for the past few years. It is a new concept of infrastructure which can be easily deployed to support the current or existing systems. It was built as an alternative approach to address the weakness of the terrestrial and satellite communication networks. Simultaneously, the existing infrastructure's restrictions and disadvantages can be devoid of the application of HAP. HAP can be an aircraft, balloon, or airship positioned 20 km to 50 km above the earth and intended to provide surveillance, terrestrial, and telecommunication services (Gultom, 2016). It greatly gives an advantage of the ease of placement, low operating costs, low propagation delay, wide elevation angle, wide coverage. Moreover, it can be used for broadband services, broadcasts, and disastrous conditions. The HAP is available worldwide where several projects have been and are being conducted in the United States, Europe, and Asia involving national space agencies such as the National Aeronautics and Space Administration (NASA), the European Space Agency (ESA), the German Aerospace Center (DLR), the Korea Aerospace Research Institute (KARI), and the Japan Aerospace Exploration Agency (JAXA) (Fidler et al., 2010).

Furthermore, HAP is quasi-stationary vehicles that operate in the stratosphere region, which falls well above the clouds, civil air routes, and jet streams, but below the mesosphere and thermosphere region (Fidler et al., 2010; Tozer & Grace, 2001). Chunchuzov et al. (2015) denoted that the wind velocity in this region may go up to 10 m/s per 100 m and beyond. The relatively low wind speed and turbulence make it perfect for this region, so air navigation has no problems. However, it is important to note that the former design of the HAP is large to fly at such extreme operating conditions of height and wind speed. Based on history, the size of the HAP ranges from 23 m up to 305 m in length, and none incorporates the Magnus effect in their design despite the well variant of its high lift generation (D'Oliveira et al., 2016).

In 1997, Modi studied Magnus effect embedment onto different configurations on a symmetrical aerofoil and thus resulted in a tremendous effect on its aerodynamic coefficients (Modi, 1997). The researcher denoted that a rotating cylinder application onto the model proves that it improves the lift capability over the ranges of low to a medium angle of attack (α). Furthermore, Ahmed et al. (2014) have conducted a numerical assessment of such embedment for NACA 0024 aerofoil, which has revealed an improvement in lift coefficient (C_L) and stall angle delay of 36% and 122%, respectively, compared to its unmodified model. Following this, the numerical analysis by Huda et al. (2015) incorporated a leading-edge rotating cylinder onto the NACA 0010 aerofoil at a rotational speed of 30, 60, 90, and 120 RPM, thereby generating a surprising improvement in the maximum lift of 145% in comparison with the unmodified NACA 0010 aerofoil model. Similarly, Faisal et al. (2017) have undertaken experimental and numerical analyses both by applying the

leading-edge rotating cylinder onto NACA 0018. However, the scholars did not record the pre- and post-aerodynamic characteristics of the modified aerofoil. Additionally, Salam et al. (2019) studied the effect of rotating cylinder on NACA0021 and discovered an increase in C_L and reduction in drag coefficient (C_D) for all α . Furthermore, a delay in flow separation was also viewed at higher α in view of momentum injection by the leading-edge cylinder.

The rotating cylinder embedment onto the bluff body occurred in 1925 (Wolff, 1925). Following this, studies pertaining to cylinder embedment on the flat plate (CyFlaP) are usually influenced by the past research known as leading edge cylindrical aerofoil (LECA). In comparison, the research on the bluff body has only come into the picture within the last decade where the cylinder to bluff body configuration was experimentally studied by Modi et al. (1991) but based only on drag reduction for automotive purposes. Accordingly, the researcher denoted that configuration resulted in a reduction of 75% in C_D . Wang et al. (2013) investigated the use of flat plates in 2013 when they examined the flat plate configuration for its aerodynamic coefficient. At around 60% of the maximum lift coefficient (C_{Lmax}), they denoted that the flat plate could generate a high coefficient at higher α .

Therefore, this paper delineates two (2) proposed designs for the preliminary design of the HAP. First, the design of a rotating cylinder onto 1) Selig S1223 aerofoil and 2) flat plate are proposed. A two-dimensional flow testing has been simulated by using the Computational Fluid Dynamic (CFD) approach utilizing a Fluent analysis system from the ANSYS WORKBENCH 2019 Software for momentum injection of 500 and 1000 revolution per minute (RPM) with free stream velocities ranging from 5 m/s to 30 m/s for different α of 0 to 20 degrees ($^\circ$). Finally, the model is proposed to study its efficiency for aerodynamic generation purposes.

This study is a novel concept incorporating a new embedment of rotating cylinder onto Selig S1223 aerofoil and flat plate intended for HAP purposes. It can be used as a future reference and benchmark for designing HAP. The novelty of embedding such aerodynamic and bluff bodies together would further elevate the aerospace industry to explore such an objective sense. In addition, the results from this CFD simulation may be of interest to the CFD community for numerical studies that consider the Magnus effect in the same class of flow regions.

METHODOLOGY

The methodology is comprised of several subsections and thus offers an insight into the research flow for the study.

Geometry and Computational Modelling

The design parameters of the model were the most important consideration that had to be considered in this study. Therefore, a thorough study of the past research data has thus

implemented onto the CFD software. An embedment of rotating cylinder onto 1) Selig S1223 aerofoil and 2) flat plate are proposed. The design parameter was posed for the available resource data as follows.

Cylinder. In 2008, Badalamenti tested a rotating cylinder in the Handley Page Laboratory of Aerodynamics at City University by utilizing a wind tunnel (Badalamenti & Prince, 2008). For the effects of the rotating cylinder, he demonstrated the aerodynamic properties, which were then applied in comparison to other configurations from previous research. He also verified the data obtained by showing agreement with Betz’s analysis for the same aspect ratio (AR) of 4.7 (Barati et al., 2019). The formula used for the aspect ratio is Equation 1:

$$AR = \frac{b}{c} = \frac{b}{c} * \frac{b}{b} = \frac{b^2}{A} \quad [1]$$

Where b is the span of the cylinder, c is the chord, and A denotes the model area. Accordingly, the calculations and the formulas used were fully integrated for the study. Therefore, the data incorporated will be used for validation in a later subsection (refer to subsection 2.6). According to Badalamenti & Prince’s study (2008), the cylinder’s dimension was set at AR = 5.1, corresponding to a diameter (Ø) of 0.16 m, as shown in Figure 1. Duly noted that the cylinder is rotating clockwise, opposing the relative velocity to mimic the same condition (Badalamenti & Prince, 2008). Furthermore, Barati et al. (2019) have done the same concept over the clockwise rotation of the cylinder to aid the freestream flow on top of the cylinder while having the one at the bottom opposing it. Therefore, it ensures an upward force for the generated lift. For 2D planar simulations, the geometry is displayed as an XY plane, and the Z-axis is oriented away from the computer screen. Hence, a counterclockwise rotation is denoted as positive and clockwise rotation is negative based on the right-hand thumb rule. Therefore, the inserted value should be input as -500 or -1000 RPM (clockwise rotation) speed for the moving wall with the rotational motion for the cylinder’s wall.

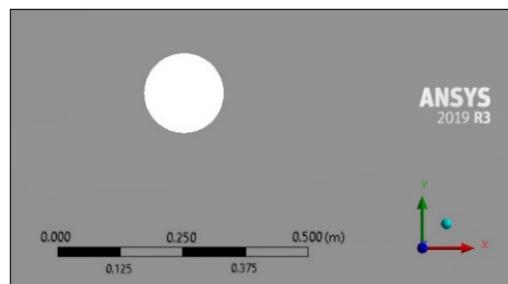


Figure 1. The geometry of the cylinder

Selig S1223 Aerofoil. The Selig S1223 aerofoil is well known for its high lift generation at a low Reynolds number (Re), originally patented by Michael S. Selig in 1995 (Selig et al., 1995). In addition, Oller et al. (2016) used the aerofoil in their study to test for hydrofoils’ effectiveness and obtained a tremendous effect for such fluid conditions. He further noted that the aerofoil yielded high separation flow aftward at high α . As a result, the LECA is investigated to determine the aerodynamic profile of such embedment.

Ergo, the aerofoil was an upscale version of the original one with a ratio of 1:1.45 from the unmodified model to meet the performance objective of the study (Figure 2). Validation of the unmodified aerofoil has been performed and is displayed in the next subsection in comparison to the experiment performed by Selig et al. (Selig et al., 1995; Selig et al., 1996).

LECA. The LECA design was presented in Figure 3, with a rotating cylinder embedded at the leading and trailing edges of the Selig S1223 aerofoil. This double embedment onto the aerofoil was adopted as per the study by Modi, where he incorporated several configurations of rotating cylinders onto a symmetrical aerofoil in 1997 (Modi, 1997). Therefore, the researcher's work inspired the creation of such a design, which was then used in this LECA investigation.

The embedment was set up along the camber line of the aerofoil while maintaining a gap of 0.005 m between the cylinder and the aerofoil. The gap was set to optimum configuration performance for the model's embedment (Abdulla & Hasan, 2018). The embedment of the cylinder at $3/20$ of the camber line was utilized to optimize its diameter to fit the aerofoil's leading edge and trailing edge, and thus satisfied with the experimental data by Michael S. Selig and James Guglielmo (Selig & Guglielmo, 1997).

Flat Plate. Another design that incorporates and utilizes the flat plate was developed to provide a simpler HAP design that can be compared to the LECA. The flat plate design by Torres (2002) and Wang et al. (2013) was duly referred on its capability for a flight at a higher value of α . As a result, the flat plate relevant to the researcher's study was established with a flat plate length of 1.00 m, as illustrated in Figure 4.

CyFlaP. The idea of the design for the CyFlaP would not spark without knowing its existence which was first introduced in 1997 by Modi (Modi, 1997). At that moment, Modi was simply testing the concept for its drag reduction for a vehicle instead of its

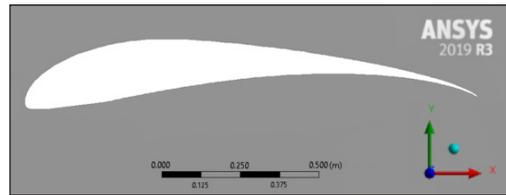


Figure 2. The geometry of the Selig S1223 aerofoil

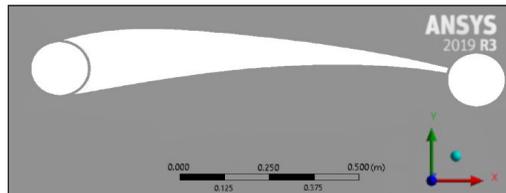


Figure 3. The geometry of the LECA

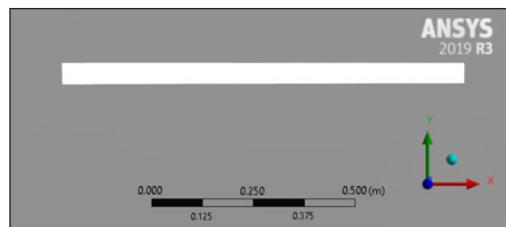


Figure 4. The geometry of the flat plate

lift generation. Therefore, the design will be used for the CyFlaP with dual cylinder embedment at both the flat plate leading edge and trailing edge, as shown in Figure 5. A gap of 0.005 m was always enforced between the cylinder and the flat plate.

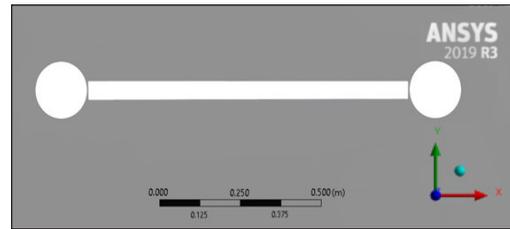


Figure 5. The geometry of the CyFlaP

Turbulence Model

Turbulence modeling, by definition, denotes a numerical technique designed to close a system of mean flow equations. It defines the design and usage of mathematical models to solve and simulate turbulence impacts, whereby their turbulence equations also provide simpler solutions using the averaging method. Turbulence modeling differs in the way they are calculated by extending the Navier-Stokes equations with additional terms consisting of some parameters such as local fluid velocity, wall distance, additional transport equations, eddy viscosity, turbulent kinetic energy, turbulence intensity, and others. Here, the Spalart-Allmaras, K-epsilon ($K-\epsilon$), and K-omega ($K-\omega$) models are some of the turbulence models that are often used in applications (Monk & Chadwick, 2017).

Spalart-Allmaras model is a one equation model which solves its modeled transport equation for the kinematic eddy viscosity (Kim et al., 1999). This model has produced promising results in simulation that involves wall-bounded flows. Moreover, it produces excellent results for boundary layers subjected to adverse pressure gradients, especially the one that occurs during vortex separation and turbulent flows over the wing surfaces (Kölzsch & Breitsamter, 2014). However, the model is less sensitive to numerical errors when non-layered meshes are used near the walls. The near-wall gradients of the transported variable used in this model are far smaller than in the $K-\epsilon$ and $K-\omega$.

The $K-\omega$ model was used for the current work with reference to the study by Mgaidi et al. (2018), which implemented two extra transport equations as a two-equation model to delineate the turbulent flow properties accordingly. Hence, the turbulent kinetic energy is referred to as K as the first variable transported, while the second transported variable is referred to as ω . However, it has difficulty converging and relies on an initial estimation at the beginning of the solution. Therefore, it is common in the industry to use the $K-\epsilon$ model first, where the kinetic energy dissipation rate is solved. However, the $K-\omega$ model is far more suitable for obtaining accurate results in many aerodynamic situations, especially for highly curved and separated flows (Ali et al., 2021a; Ali et al., 2021b; Mgaidi et al., 2018; Monk & Chadwick, 2017).

Antecedent to this, Menter developed a two-equation eddy-viscosity model, otherwise referred to as the SST $K-\omega$ turbulence model, which is used in the near-wall region to combine and blend with the SST $K-\omega$ model in the far-field free stream $K-\omega$ model (Menter,

1994). The model is an advanced extension of the original model proposed by Wilcox in 1988, which can demonstrate a strong turbulence formulation for both the adverse pressure gradients and the separating flow (Wilcox, 1988). Conversely, the SST formulation is built by the integrated modeling of the K- ω and K- ϵ formulations, whereby the amalgamation helps the SST in transitioning from the K- ω model to the K- ϵ model to prevent problems occurring in the inlet of the free-stream turbulence (Mgaidi et al., 2018). At the same time, the K- ω formulation is included in the inner portion of the boundary layer instead of its K- ϵ equivalent. Consequently, the general form of transport equations in the Fluent analysis system is as shown in Equations 2 and 3:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_j} \left[\Gamma_k \frac{\partial k}{\partial x_j} \right] + G_k - Y_k + S_k \quad [2]$$

$$\frac{\partial}{\partial t}(\rho \omega) + \frac{\partial}{\partial x_i}(\rho \omega u_i) = \frac{\partial}{\partial x_j} \left[\Gamma_\omega \frac{\partial \omega}{\partial x_j} \right] + G_\omega - Y_\omega + S_\omega \quad [3]$$

The generation of turbulence kinetic energy due to mean velocity gradient and output of ω is expressed in these equations as G_k and G_ω , respectively. Meanwhile, the effective diffusivity of k and ω due to turbulence is expressed as Γ_k and Γ_ω , respectively. The dissipation of k and ω is then presented as Y_k and Y_ω , respectively, whereas the D_ω presents the cross-diffusion concept. Finally, yet importantly, the S_k and S_ω are the input by the user-defined source terms.

Concerning the above argument, the SST K- ω model is identical to the standard K- ω model but contains the following refinements (Ali et al., 2021a; Khalil et al., 2018):

- (a) A blending function feature is used to multiply the standard K- ω model with the transformed K- ω model, which is then applied or added together.
- (b) In the blending function, the standard K- ω will be triggered for a near-wall region, while the transformed K- ω will be activated zero away from the surface.
- (c) The ω equation in the SST model is derived as a dampened cross-diffusion.
- (d) The discrepancy between the turbulent shear stress transport and the modeling constant is compensated for by adjusting the turbulent viscosity definition.

All these refinements will allow the SST K- ω model to be extremely effective and efficient for a wider variety of flows, including aerofoils, adverse pressure gradient flows, and transonic shock waves, relative to the standard K- ω model (Khan et al., 2020). Therefore, the research study will mainly implement the SST K- ω model, which allows the model to be directly used on the wall through the viscous sublayer. Many who have worked with this particular model have found that promising results have been obtained in situations with adverse pressure gradients and separating flows (Ali et al., 2021a; Ali et al., 2021b; Hamisu et al., 2019; Mgaidi et al., 2018; Monk & Chadwick, 2017).

Key Performance Parameter

This study's Key Performance Parameters (KPPs) directly and precisely describe the output target. Here, the C_L and C_D equations used for the model formulation are presented in Equations 4 and 5.

$$C_L = \frac{L}{\frac{1}{2} * \rho * V^2 * S} \quad [4]$$

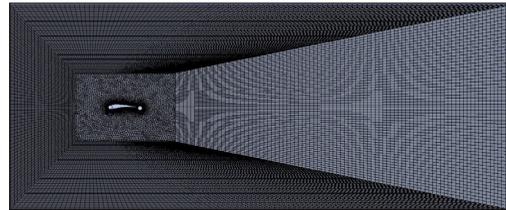
$$C_D = \frac{D}{\frac{1}{2} * \rho * V^2 * S} \quad [5]$$

where L is the lift force, D is the drag force, ρ is the density of the fluid (i.e., air), V is the free stream velocities, and S is the projected area of the model.

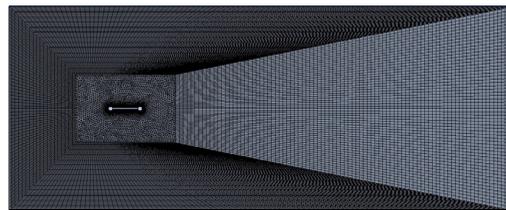
Grid Generation

The standard CFD process will need a mesh appropriate for the computational domain boundaries. It involves the generation of a computational mesh appropriate for the expected 2D Navier-Stokes equations, in which a set of grid points are defined for the domain and its boundaries. A grid structure like this is referred to as grid generation.

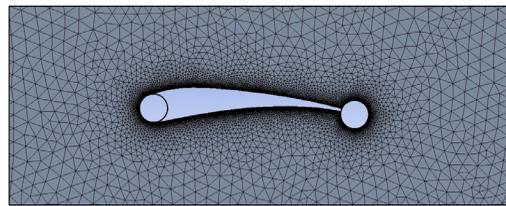
Mesh Topology. Two meshing zones were integrated into the Fluent fluid flow analysis system. The domain consists of an inner zone with a block of $2\emptyset \times 3\emptyset$ around the model and an outer zone with a block setup of $6\emptyset \times 15\emptyset$ across the inner zone [Figures 6(a) and 6(b)]. The block setup was applied based on the research work of Mgaidi et al. (2018) and Yao et al. (2016). The LECA and the CyFlaP models use the identical domain configuration throughout the investigation. The model surfaces were fixed with y-plus (y^+) less than 1, which was located from the



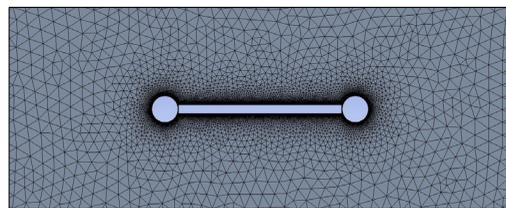
(a)



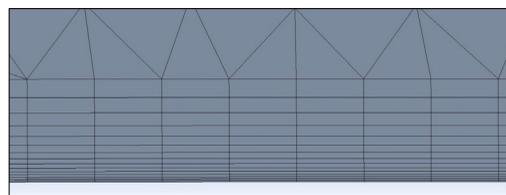
(b)



(c)



(d)



(e)

Figure 6. Domain and mesh generation for the LECA and CyFlaP model, from the domain setting for LECA (a) and CyFlaP (b), mesh generated for LECA (c) and CyFlaP (d), to the inflation set up on the model (e)

wall to the first mesh node in compliance with the inflation criterion of a maximum of 10 layers and with a growth rate of 1.2 to obtain good results [Figures 6(c), 6(d) and 6(e)]. This y^+ is a non-dimensional distance often used to determine how fine or coarse a mesh is for a particular flow pattern. It is a ratio between laminar and turbulent influence in a cell. Here, it was used to determine the proper size of the cells near its domain walls which were important in turbulence modeling. A faster flow near the wall will produce higher values of y^+ , so the grid size near the wall must be reduced. Therefore, an inflation layer between 10 and 15 layers on the domain wall will generally resolve the boundary layer and accurately predict any separation or reattachment points. In addition, the wall function strategy could be conveniently achieved for a particular turbulence model. The specifications of the boundary conditions and mesh environment for this analysis are provided, as per Tables 1 and 2.

Table 1
Specifications of boundary conditions

Boundary conditions	Type
Inlet	Velocity-inlet
Outlet	Pressure-outlet
Cylinder	Wall
Cylinder Aftward	Wall
Selig aerofoil	Wall
Flat Plate	Wall
Wall	Symmetry
Interior surface body	Interior
Surface body	Interior

Table 2
Mesh configurations

Mesh specifications	
Growth rate	1.2
Defeature size	5.e-004 m
Curvature minimum size	1.e-003 m
Curvature normal angle	18.0°
Smoothing	High
Inflation specifications	
Inflation option	First layer thickness
First layer height	1.e-005 m
Maximum layers	10
Growth rate	1.2

Mesh Independency Test. In order to monitor the solution grid independence before commencing with the model testing phase, a mesh independency test (MIT) was recommended (Abdulla & Hasan, 2018). It was obtained by establishing a new cell grid and testing it utilizing various alternatives. Ergo, as seen in Figures 7 and 8, MIT is conducted for the LECA and CyFlaP to obtain the best data quality. Consequently, at approximately $1.560e+005$ (LECA) and $1.640e+005$ (CyFlaP) cells, the grid refinement for C_L resulted in the highest accuracy and durability for CFD research from now on. At this point, the simulation solves time was denoted with an increase of elements where the system's response converges to a solution where further refinement of the mesh would not affect the solution. Therefore, refinement past this intersection point is an inefficient application of CFD; hence the selected number of cells would be appropriate in capturing the system behavior while reducing the solve time. The percentage error from the described grid refinement was less than 1% to achieve the optimal time efficiency.

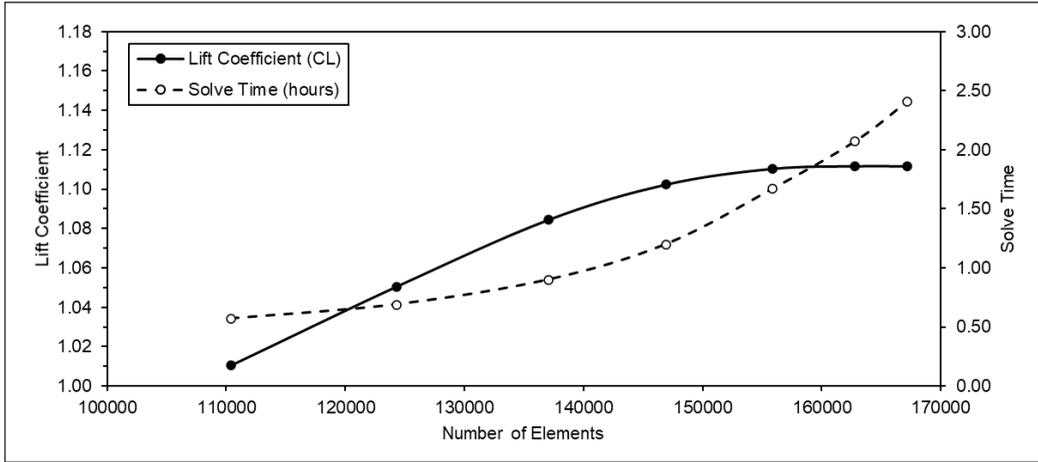


Figure 7. MIT for the LECA's convergence versus solve time

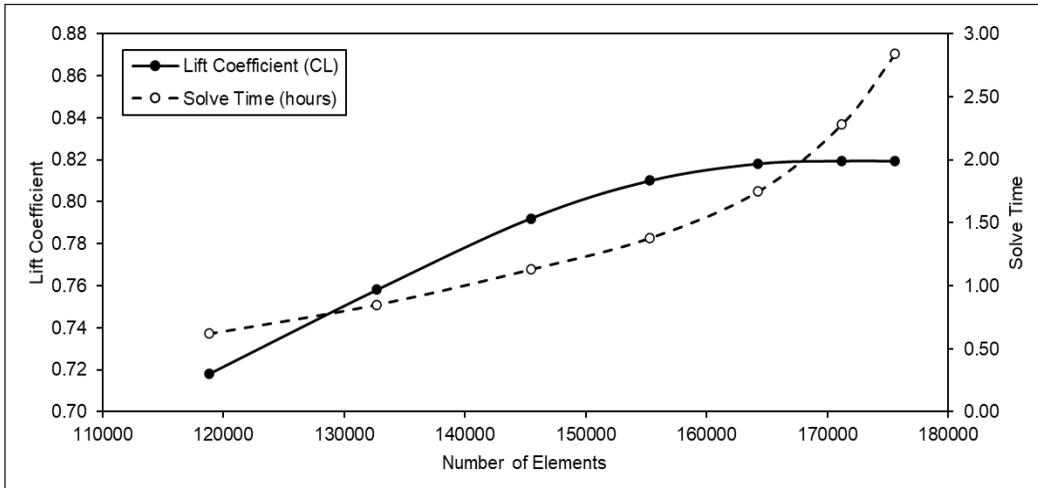


Figure 8. MIT for the CyFlaP's convergence versus solve time

Solver Setting

Before beginning the simulation solutions, it is critical to specify the settings for the targeted condition in the CFD solver interface. The best setup was introduced to evaluate the computational simulations, as seen in Table 3.

Validation

Validation for the cylinder, Selig S1123 aerofoil, and the flat plate was studied to obtain better results before starting deeper through the CFD process. The validation method has been referred to based on a few research as a guidance for the validation used (Boye et al., 2017; Gowree & Prince, 2012; Merryisha & Rajendran, 2019; Salam et al., 2019). Based

Table 3
ANSYS Fluent 19 solver setting

General	
Type	Pressure-based
Velocity formulation	Absolute
Time	Transient
2D space	Planar
Model	
Viscous	SST k-omega
Reference values	
Density (kg/m ³)	1.225
Pressure (pascal)	101325
Temperature (K)	288.16
Velocity (m/s)	5, 10, 15, 20, 25, 30
Viscosity (kg/m-s)	1.789e-005
Turbulence	
Specification method	Intensity and length scale
Turbulent intensity (%)	5
Turbulent length scale (m)	9.310e-002
Solution	
Method	Pressure-velocity coupling
Scheme	Coupled
Residual error	1.e-006
Spatial discretization	
Gradient	Green Gauss Node Based
Pressure	PRESTO!
Momentum	QUICK
Turbulent kinetic energy	QUICK
Specific dissipation rate	QUICK
Transient formulation	Bounded Second Order Implicit

on this validation analysis, the embedment of the LECA and CyFlaP was cautiously carried out. Badalamenti and Prince (2008), Torres (2002), and Selig et al. (1995) performed experimental fluid dynamics (EFD) on the rotating cylinder, Selig S1223 aerofoil, and flat plate, which were then validated by using the CFD approach, as illustrated in Figures 9 to 11. The validation analysis for each of the situations culminated in an error of less than 10%.

RESULTS AND DISCUSSION

The numerical results were obtained for flow over the proposed LECA and CyFlaP embedment for free stream velocities of 5 m/s up until 30 m/s at different α (i.e., 0°, 5°, 10°, 15°, and 20°) and for a rotational speed of 500 RPM and 1000 RPM. For LECA and CyFlaP conditions, a gap of 0.005 m was set between the cylinder, the aerofoil and/

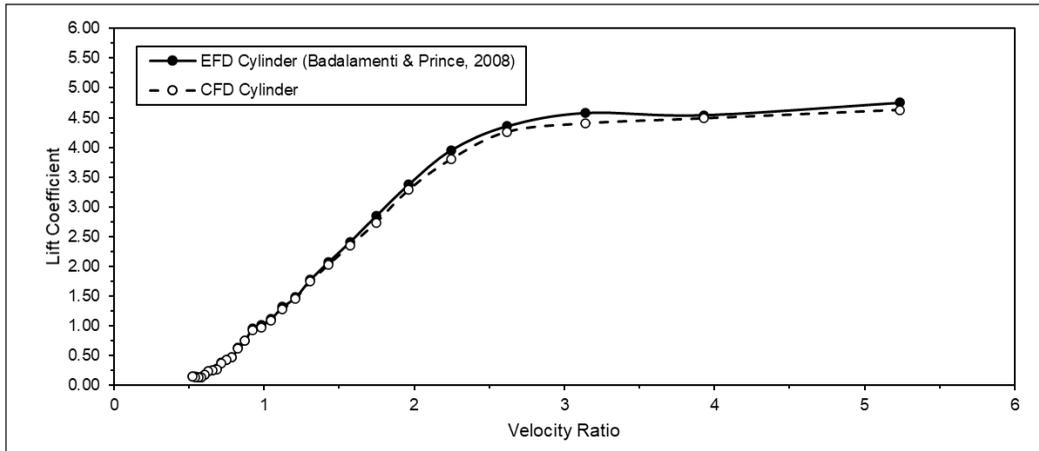


Figure 9. Validation of rotating cylinder

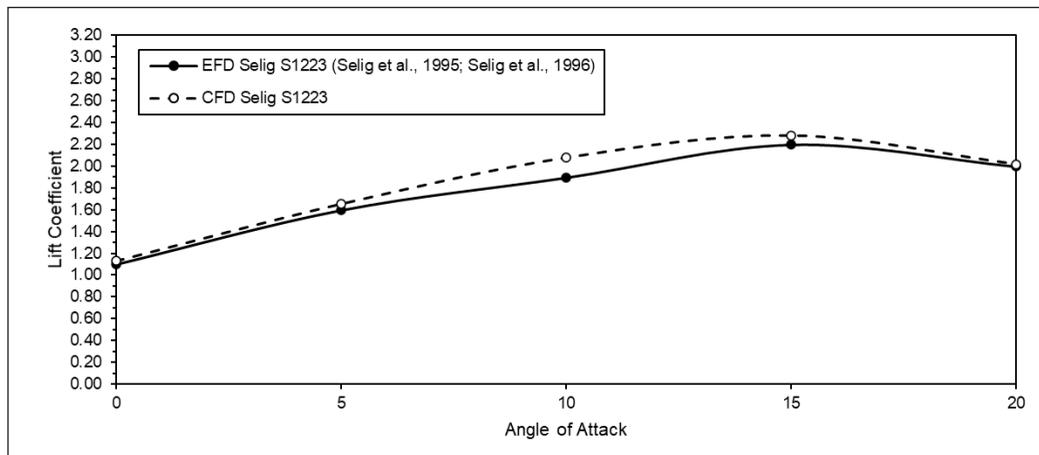


Figure 10. Validation of Selig S1223 aerofoil

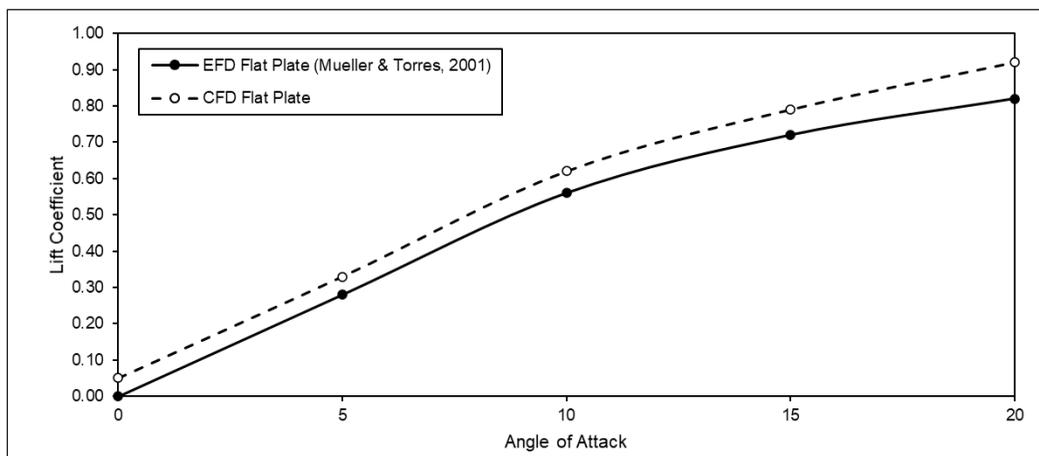


Figure 11. Validation of flat plate

or flat plate. Additionally, it is often necessary to estimate the turbulent intensity on the inlet when setting boundary conditions for a CFD simulation. For simulations at this class of stratosphere flow region in which the inlet involves a turbulent boundary layer for wall-bounded flows, the inlet turbulent intensity level is set to the recommended default “Medium,” which is 5%, together with its turbulent length scale of $9.310e-002$, respectively (ANSYS, 2013; Šidlof et al., 2017; Russo & Basse, 2016).

Numerical Simulation

In compliance with the following clauses, the numerical simulation was specified accordingly. As a result, the aerodynamic flow analysis is dominated by the dimensionless parameters C_L , C_D , RPM , Re , and α , resulting in a functional relationship with the numerical solution’s performance.

Effect of Magnus Effect. As shown in Figures 12 to 19, the computational analysis for the proposed LECA and CyFlaP resulted in the following results. The Magnus effect on the LECA and CyFlaP is demonstrated in Figures 12 to 19 by injecting a momentum injection of 500 RPM and 1000 RPM, respectively. In all cases for this analysis, the cylinder rotation was set on all the CFD simulations examined, rotating in the clockwise direction. The planned environment was to uphold the momentum injection on the aerofoil’s upper surface aligned with the free streamflow’s path.

Figures 12 and 13 present the aerodynamic performance of the Magnus effect on LECA at 500 RPM. Here, the performance of the LECA was increased at $\alpha = 0^\circ$ for 5 m/s to 10 m/s conditions yet still had some aerodynamic performance below the unmodified model from Selig at higher α for 15 m/s to 30 m/s conditions (Selig et al., 1995; Selig et al., 1996). Despite this variance, the application of cylinders at both sides of the aerofoil has resulted in slight C_L improvement. Next, the 5 m/s conditions have resulted in a 53% stall angle delay at $\alpha = 20^\circ$ beyond the unmodified model. The help of flow recovery at low wind speed may be the one that contributed the most to this model condition. However, the lowest C_L can be seen at $\alpha = 20^\circ$ with a 37% coefficient below the unmodified model for 15 m/s conditions, where it was caused predominantly by the shape of the aerofoil, which formed high separation and burbling at the aftward surfaces of the aerofoil as the flow broke away that clearly shows the reason of high C_D as seen in Figure 13. Therefore, it can be denoted that the LECA behaved the best at a higher α for lower wind velocity. A higher rotational speed would amplify the recovery of depleted C_L at the lower α later.

Figures 14 and 15 display the aerodynamic performance of the Magnus effect on LECA at 1000 RPM. The trends showed better C_L for lower freestream flow but not for larger value of freestream flow. For lower freestream flow, i.e., 5 m/s, LECA generates higher C_L and lower C_D than the rest of the free stream velocities, which has 39% better C_{Lmax}

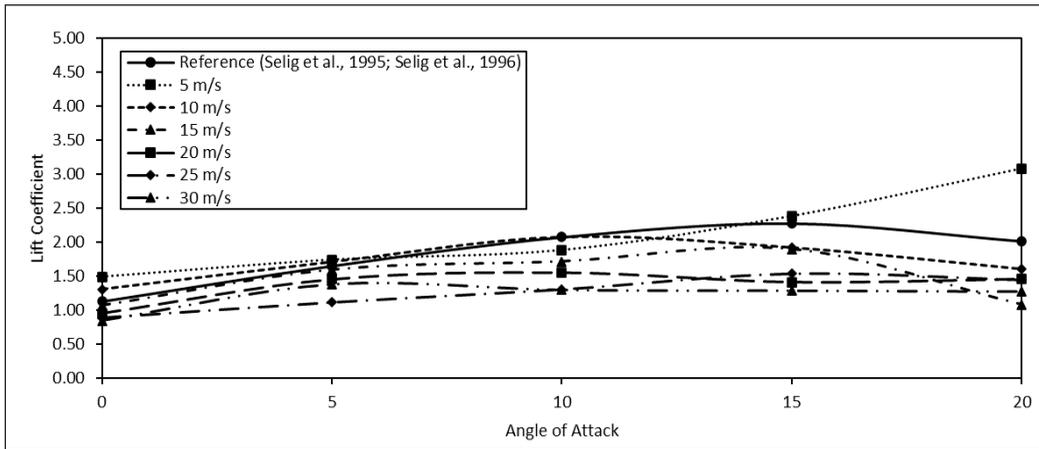


Figure 12. C_L of LECA versus α at 500 RPM of rotational speed

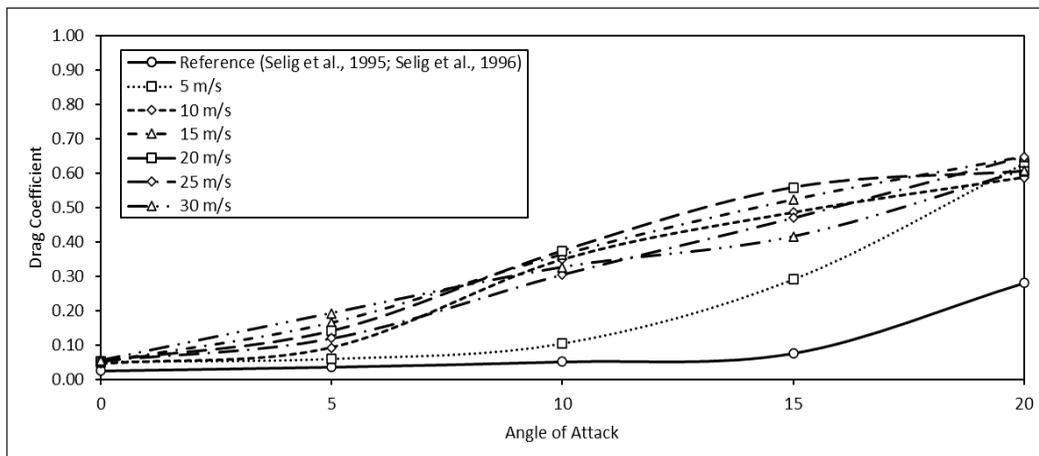


Figure 13. C_D of LECA versus α at 500 RPM of rotational speed

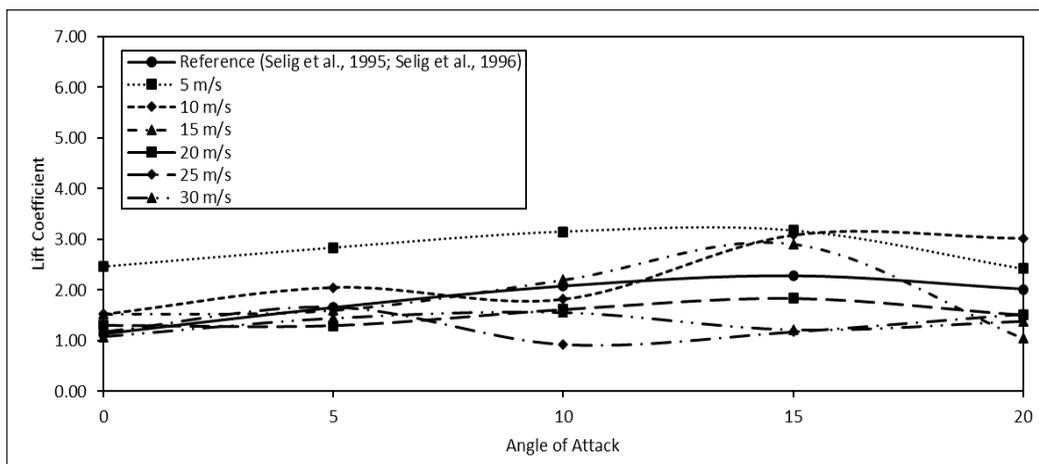


Figure 14. C_L of LECA versus α at 1000 RPM of rotational speed

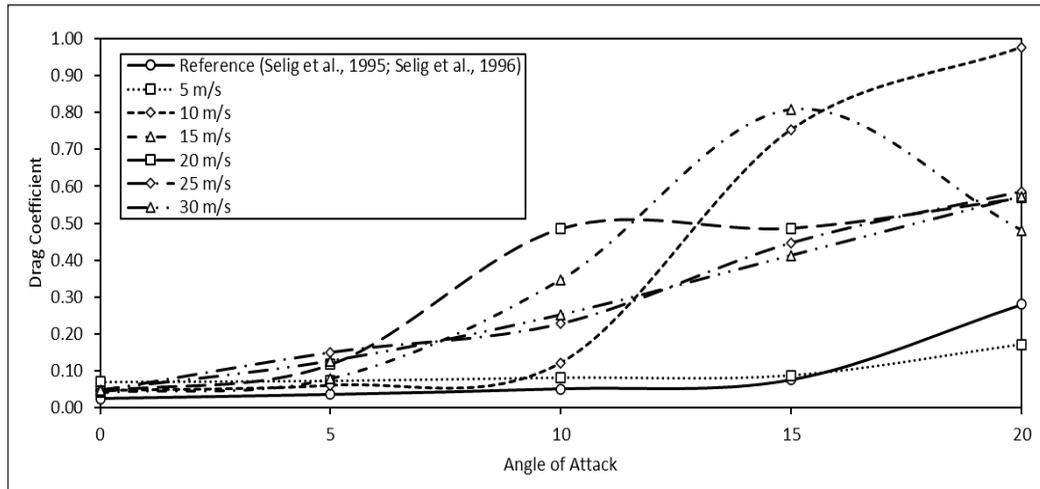


Figure 15. C_D of LECA versus α at 1000 RPM of rotational speed

than the unmodified model at $\alpha = 15^\circ$ before stalling. At a velocity of 10 m/s and 25 m/s, the trend rose then stalled from $\alpha = 5^\circ$ to 10° , yet rose back on higher α onwards. This diminution was caused by a flow separation aftward of the LECA, which was recovered with momentum injection from the trailing edge rotating cylinder. As for LECA at 15 m/s and 20 m/s free stream velocities, a slight plateau effect was seen at the initial flow from $\alpha = 0^\circ$ to 5° . This sudden plateau effect was caused by the gradually reducing aerodynamic coefficient but did not stall immediately. Despite some of the 1000 RPM momentum injections showing better performances against the 500 RPM yet, the high drag effect due to the swirling vortices aftward of the cylinder was not able to recover fully, which has affected the LECA's aerodynamic performance, especially in generating inconsistent C_L . An increase in the cylinder rotational speed may help elevate its aerodynamic performance for certain free stream velocities conditions. However, it may not be suitable for the intended HAP, designed to operate at high altitudes with such inevitable turbulence.

Figures 16 and 17 show the aerodynamic performance of the Magnus effect on CyFlaP at 500 RPM. At this condition, the highest C_L was denoted at 5 m/s conditions with a 128% improvement to the unmodified model. However, the lowest was 26% for 30 m/s free stream velocities. Moreover, a slight plateau effect can be seen at $\alpha = 5^\circ$ to 10° , but the condition may recover with a further increment of momentum injection on the model. Overall, the C_L performance showed satisfactory results beyond the unmodified model (Torres, 2002). Figure 17 showed the C_D of CyFlaP, which showed high values of C_D due to the swirling vortices carried away from the trailing edge section of the model. However, it is significantly better than the LECA's C_D and is on par with the unmodified model C_D . The matter can be resolved with higher momentum injection to elevate its aerodynamic performance, reducing the drag even further.

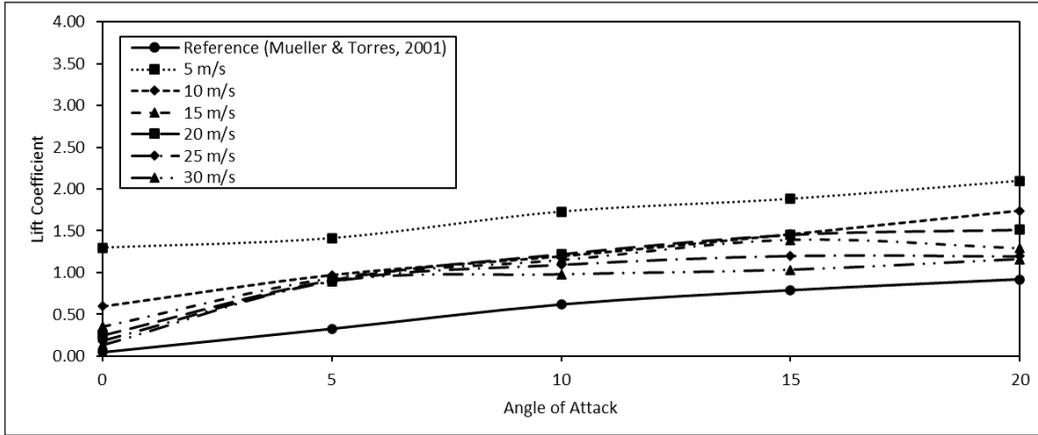


Figure 16. C_L of CyFlaP versus α at 500 RPM of rotational speed

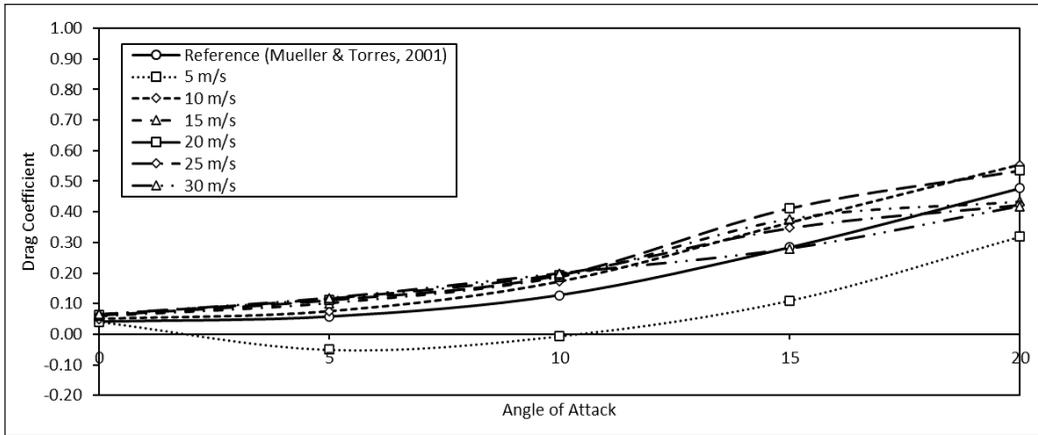


Figure 17. C_D of CyFlaP versus α at 500 RPM of rotational speed

Figures 18 and 19 show the aerodynamic performance of the Magnus effect on CyFlaP at 1000 RPM. The graph trend for CyFlaP is better than LECA embedment, which showed a tremendous effect on all the free stream velocities and α . The freestream velocities of 5 m/s to 15 m/s resulted in a modest increase in performance, which might be attributed to the separation bubble. It recovered for a higher α , which further elevated its aerodynamic coefficient and stall angle delay by up to 204% better than the unmodified model. Based on this trend, the CyFlaP may go for a higher α beyond the $\alpha = 20^\circ$. The negative C_D trend in Figures 17 and 19 is equivalent to a positive thrust which resulted in a high C_L value recorded. Nevertheless, for higher free stream velocities of 20 m/s to 30m/s, the trend showed an increase in C_L yet plateau from $\alpha = 5^\circ$ to 15° . The same plateau effect from LECA and CyFlaP is due to the partial separation bubble during the flight. However, it may be recovered with an increment of the momentum injection, which helps reattach the air stream towards the LECA and CyFlaP surface body.

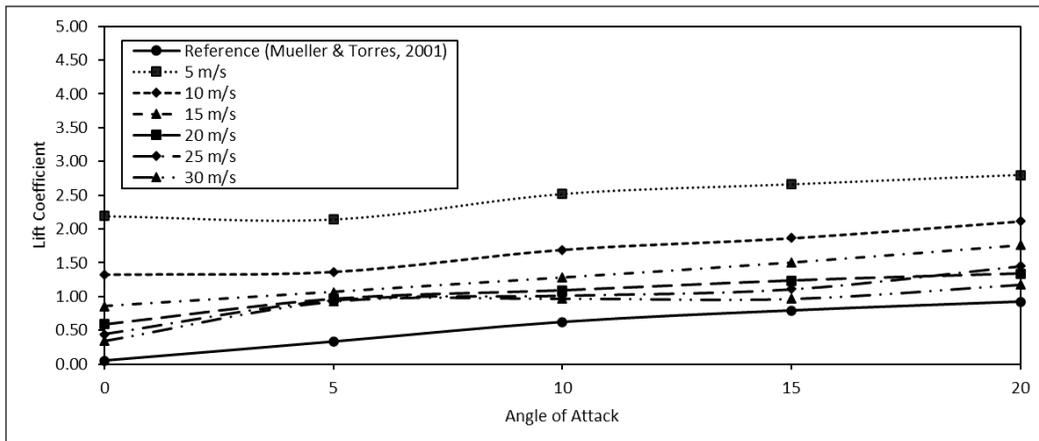


Figure 18. C_L of CyFlaP versus α at 1000 RPM of rotational speed

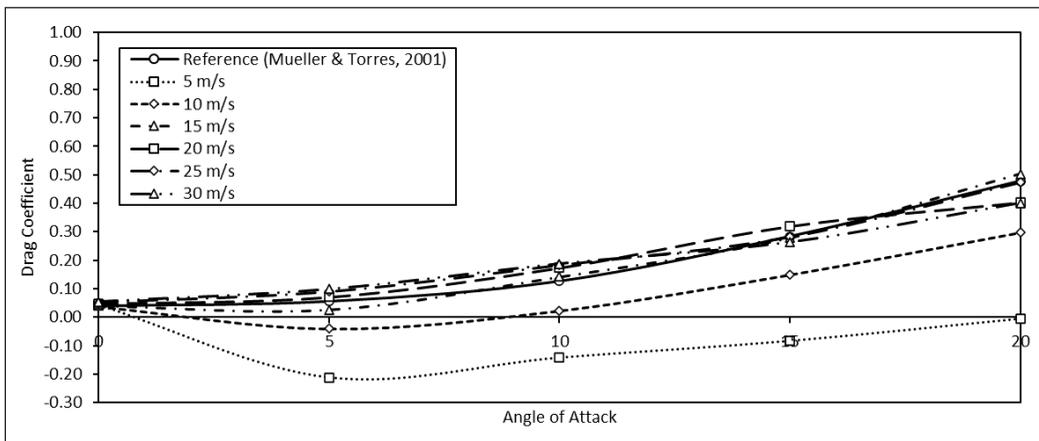


Figure 19. C_D of CyFlaP versus α at 1000 RPM of rotational speed

Comparatively, the unmodified Selig S1223 aerofoil has better C_L than the unmodified flat plate in terms of aerodynamic performance. In this case, the momentum injection for the flat plate, known as the CyFlaP, works well in attaining its aerodynamic coefficient and is thus more stable and suitable for the application of HAP than the modified LECA. Furthermore, the wobble in the LECA trend may not be suitable for HAP as this may be dangerous and difficult to control at such a high altitude. Previously, the highest recorded data on the leading-edge cylinder embedment has a maximum lift of 145% for NACA 0010 aerofoil by Huda et al. (2015), followed by 36% in C_L by Ahmed et al. (2014). On the contrary, this research study has proved that the LECA generated 39% and 53% improvement while the CyFlaP has denoted 128% and 204% enhancement on its C_L and stall angle delay, respectively. These recorded data clearly show that the use of double rotating cylinders on both designs would be beneficial in stabilizing its aerodynamic performance, especially

on the CyFlaP, where the coefficient may extend beyond the recorded α . Finally, it should be noted that the LECA and CyFlaP designs perform better at lower free stream velocities, which are appropriate for the stratosphere region.

Effect on Velocity Magnitude Contour.

The flow evolution was derived from the CFD post-processing for this investigation. The velocity inlet of 5 m/s to 30 m/s was used to express the physical vector quantity. For free stream velocities of 5 m/s to 30 m/s at 1000 RPM, the velocity magnitude contour for twelve (12) different conditions at $\alpha = 20^\circ$ is shown in Figure 20. Here, partial flow separation can be seen behind the LECA and smooth flow contour for the CyFlaP at 5 m/s free stream velocity. A higher value of free stream velocities resulted with flow separation starting from 10 m/s onwards, where a swirling vortex started to occur aftward of the model.

Meanwhile, the trailing edge rotating cylinder momentum injection aids in reattaching the flow to the surface before catastrophic phenomena such as stall occur. A higher momentum injection would be suggested to study the flow reattachment as moving surface boundary control. Flow reattachment would be sufficient to bring the stable flight back into its path for HAP application.

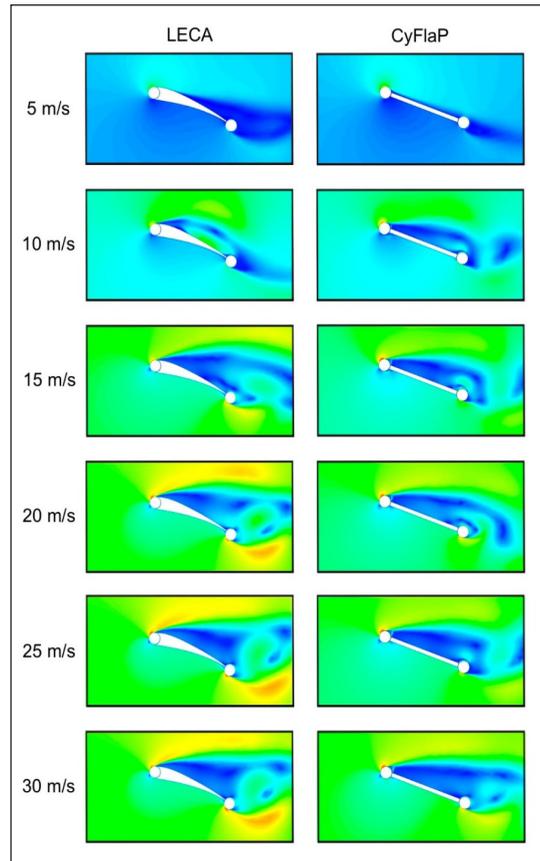


Figure 20. Velocity magnitude contour on LECA (Left) and CyFlaP (Right) at $\alpha = 20^\circ$

Effect on Pressure Coefficient Contour. Another set of flow evolution for pressure coefficient (C_p) was derived from the CFD post-processing to study its pressure distribution along the surface body of the model, as shown in Figure 21. This analysis represented 12 conditions for free stream velocities ranging from 5 to 30 m/s at 1000 RPM momentum injection at $\alpha = 20^\circ$. Here, the partial flow separation from the previous clause indicates the existence of Kármán vortex street. Furthermore, it was a condition in which swirling vortices emerge aftward the model, affecting the present flow over time. In Figures 20 and 21, a larger formation of swirling vortices can be seen aftward of the LECA, which

may be the factor with the wobble for its aerodynamic coefficient, resulting in high C_D , which may cause vibration and inconsistent flight.

Based on Bernoulli's principle, low pressure on the upper surface and higher pressure on the lower surface body resulted in higher lift and lower drag generation. The rotation of the rotating cylinder is further enhanced, thus reattaching the flow onto its surface body whenever flow separation occurs. Furthermore, a formation of higher pressure was depicted towards the lower rotating cylinder surfaces with an increment of free stream velocities, concluding upon the success of Magnus effect generation for the model's embedment in HAP application. Therefore, a momentum injection played a vital role in improving the flow boundary layer back onto its surface body. Therefore, a higher momentum injection is suggested to reattach the flow further, thus reducing its drag generation.

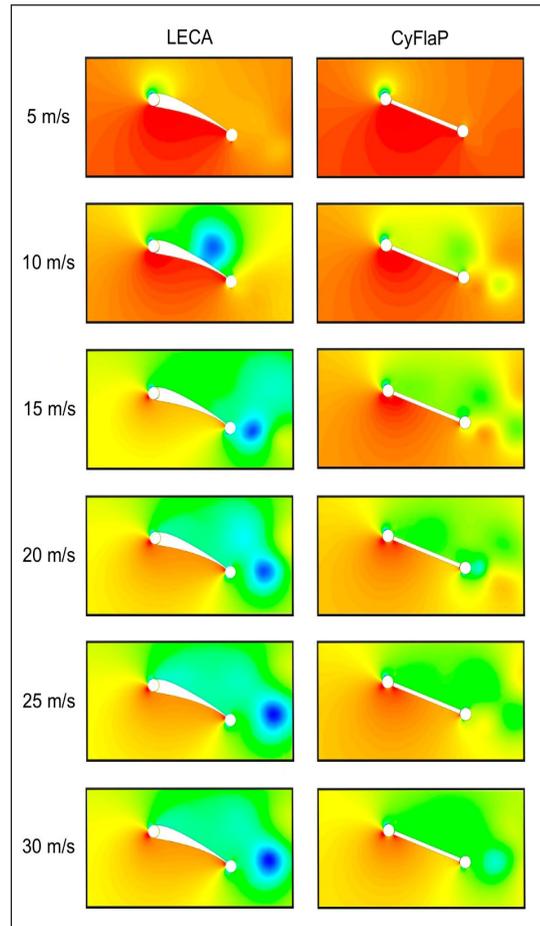


Figure 21. C_p contour on LECA (Left) and CyFlaP (Right) at $\alpha = 20^\circ$

CONCLUSION

The numerical simulation analysis by embedding rotating cylinders onto LECA and CyFlaP has tremendously affected their aerodynamic coefficients. The analysis yielded the effects on C_L and stall angle delay of about 39% and 53%, respectively, for LECA while showing 128% and 204% better for CyFlaP than their respective unmodified model. It is believed that the CyFlaP provides more stability while being simple in design, making it appropriate for HAP applications. Despite having a better aerodynamic coefficient than the CyFlaP, the LECA is unsuitable for the task due to uneven flow and excessive vortices. The momentum injection on the model showed a rise in C_L , reduction in C_D , and yet extended its stall angle delay, thus concluded upon the successful Magnus effect application on the model. Overall, further research, such as momentum injection increment, will be recommended to improve its aerodynamic performance for the intended HAP application.

ACKNOWLEDGEMENT

The Ministry of Education (MoE) Malaysia supported this research under the research grant scheme of Fundamental (FRGS) Phase 1/2018 with MoE ref. code FRGS/1/2018/TK09/UPM/02/2 and project code of 03-01-18-1950FR. The support is gratefully acknowledged.

REFERENCES

- Abdulla, N. N., & Hasan, M. F. (2018). Effect of gap between airfoil and embedded rotating cylinder on the airfoil aerodynamic performance. *Research & Development in Material Science*, 3(4), 1-10. <https://doi.org/10.31031/rdms.2018.03.000567>
- Ahmed, S., Nazari, A., & Wahba, E. (2014). Numerical analysis of separation control over an airfoil section. *International Review of Aerospace Engineering*, 7(2), 61-68. <https://doi.org/10.15866/irease.v7i2.2057>
- Ali, H. M., Rafie, A. S. M., Ali, S. A. M., & Gires, E. (2021a). Computational analysis of the rotating cylinder embedment onto flat plate. *CFD Letters*, 13(12), 133-149. <https://doi.org/10.37934/cfdl.13.12.133149>
- Ali, H. M., Rafie, A. S. M., & Ali, S. A. M. (2021b). Numerical analysis of leading-edge cylinder aerofoil on Selig S1223 for moving surface boundary control. *Journal of Aeronautics, Astronautics and Aviation*, 53(2), 143-153. [https://doi.org/10.6125/JoAAA.202106_53\(2\).06](https://doi.org/10.6125/JoAAA.202106_53(2).06)
- ANSYS. (2013) *ANSYS fluent theory guide*. ANSYS, Inc. https://www.academia.edu/38091499/ANSYS_Fluent_Theory_Guide
- Badalamenti, C., & Prince, S. (2008). *Effects of endplates on a rotating cylinder in crossflow*. In *AIAA Applied Aerodynamics Conference*. American Institute of Aeronautics Ins. <https://doi.org/10.2514/6.2008-7063>
- Barati, E., Zarkak, M. R., & Esfahani, J. A. (2019, April 30 - May 2). *Effect of rotational direction of circular cylinder for mixed convection at subcritical Reynolds Number*. In *27th Annual International Conference of Iranian Society of Mechanical Engineers (ISME 2019)* (pp. 1-6). Tehran, Iran.
- Boye, T. E., Nwaoha, T. C., Olusegun, S. D., & Ashiedu, F. I. (2017). A validation method of computational fluid dynamics (CFD) simulation against experimental data of transient flow in pipes system. *American Journal of Engineering Research*, 6(6), 67-79.
- Chunchuzov, I., Kulichkov, S., Perepelkin, V., Popov, O., Firstov, P., Assink, J. D., & Marchetti, E. (2015). Study of the wind velocity-layered structure in the stratosphere, mesosphere, and lower thermosphere by using infrasound probing of the atmosphere. *Journal of Geophysical Research: Atmosphere*, 120(17), 8828-8840. <https://doi.org/10.1002/2015JD023276>
- D'Oliveira, F. A., Melo, F. C. L. D., & Devezas, T. C. (2016). High-altitude platforms - Present situation and technology trends. *Journal of Aerospace Technology and Management*, 8(3), 249-262. <https://doi.org/10.5028/jatm.v8i3.699>
- Faisal, K. M., Salam, M. A., Ali, M. T., Sarkar, M. S., Safa, W., & Sharah, N. (2017). Flow control using moving surface at the leading edge of aerofoil. *Journal of Mechanical Engineering*, 47(1), 45-50. <https://doi.org/10.3329/jme.v47i1.35420>

- Fidler, F., Knappek, M., Horwath, J., & Leeb, W.R. (2010). Optical communications for high-altitude platforms. *Journal of Selected Topics in Quantum Electronics*, 16(5), (1058-1070). <https://doi.org/10.1109/JSTQE.2010.2047382>
- Gowree, E. R., & Prince, S. A. (2012). A computational study of the aerodynamics of a spinning cylinder in a crossflow of high Reynolds number. In *Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences (ICAS'12)* (pp. 1138-1147). Academia.
- Gultom, A., & Yuniarti, D. (2016). Kajian teknologi high altitude platform (HAP) [Study of high altitude platform (HAP) technology]. *Buletin Pos Dan Telekomunikasi*, 14(1), 1-11. <https://doi.org/10.17933/bpostel.2016.140102>
- Hamisu, M. T., Jamil, M. M., Umar, U. S., & Sa'ad, A. (2019). Numerical study of flow in asymmetric 2D plane diffusers with different inlet channel lengths. *CFD Letters*, 11(5), 1-21.
- Huda, M. N., Ahmed, T., Ahmed, T. S. M., Salam, M. A., Afsar, M. R., Faisal, K. M., & Ali, M. T. (2015). *Study of NACA 0010 symmetric airfoil with leading edge rotating cylinder in a subsonic wind tunnel*. ResearchGate.
- Khalil, H., Saqr, K., Eldrainy, Y., & Abdelghaffar, W. (2018). Aerodynamics of a trapped vortex combustor: A comparative assessment of RANS based CFD models. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 43(1), 1-19.
- Khan, S. A., Bashir, M., Baig, M. A. A., & Ali, F. A. G. M. (2020). Comparing the effect of different turbulence models on the CFD predictions of NACA0018 airfoil aerodynamics. *CFD Letters*, 12(3), 1-10. <https://doi.org/10.37934/cfdl.12.3.110>
- Kim, S. E., Choudhury, D., & Patel, B. (1999). Computations of complex turbulent flows using the commercial code FLUENT. In M. D. Salas, J. N. Hefner & L. Sakell (Eds.), *Modeling complex turbulent flows* (pp. 259-276). Springer. https://doi.org/10.1007/978-94-011-4724-8_15
- Kölzsch, A., & Breitsamter, C. (2014). Vortex-flow manipulation on a generic delta-wing configuration. *Journal of Aircraft*, 51(5), 1380-1390. <https://doi.org/10.2514/1.C032231>
- Menter, F. R. (1994). Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA Journal*, 32(8), 1598-1605. <https://doi.org/10.2514/3.12149>
- Merryisha, S., & Rajendran, P. (2019). *CFD validation of NACA 2412 airfoil*. ResearchGate. <https://doi.org/10.13140/RG.2.2.16245.42723>
- Mgaidi, A. M., Rafie, A. S., Ahmad, K. A., Zahari, R., Hamid, M. F. A., & Marzuki, O. F. (2018). Numerical and experimental analyses of the flow around a rotating circular cylinder at subcritical regime of Reynolds number using K-E and K- Ω -SST turbulent models. *ARPJ Journal of Engineering and Applied Sciences*, 13(3), 954-960.
- Modi, V. J. (1997). Moving surface boundary-layer control: A review. *Journal of Fluids Structures*, 11(6), 627-663. <https://doi.org/10.1006/jffs.1997.0098>
- Modi, V. J., Fernando, M. S. U. K., & Yokomizo, T. (1991). Moving surface boundary-layer control-Studies with bluff bodies and application. *AIAA Journal*, 29(9), 1400-1406. <https://doi.org/10.2514/3.10753>

- Monk, D., & Chadwick, E. A. (2017, July 3-6). *Comparison of turbulence models effectiveness for a delta wing at low Reynolds numbers*. In *7th European Conference for Aeronautics and Space Sciences (EUCASS)* (pp. 1-12). Milan, Italy. <https://doi.org/10.13009/EUCASS2017-653>
- Oller, S. A., Nallim, L., & Oller, S. (2016). Usability of the Selig S1223 profile airfoil as a high lift hydrofoil for hydrokinetic application. *Journal of Applied Fluid Mechanics (JAFM)*, *9*(2), 537-542. <https://doi.org/10.18869/acadpub.jafm.68.225.24302>
- Russo, F., & Basse, N. T. (2016). Scaling of turbulence intensity for low-speed flow in smooth pipes. *Flow Measurement and Instrumentation*, *52*, 101-114. <https://doi.org/10.1016/j.flowmeasinst.2016.09.012>
- Salam, M. A., Deshpande, V., Khan, N. A., & Ali, M. T. (2019). Numerical analysis of effect of leading-edge rotating cylinder on NACA0021 symmetric airfoil. *European Journal of Engineering and Technology Research*, *4*(7), 11-17. <https://doi.org/10.24018/ejeng.2019.4.7.1385>
- Selig, M. S., & Guglielmo, J. J. (1997). High-lift low Reynolds number aerofoil design. *Journal of Aircraft*, *34*(1), 72-79. <https://doi.org/10.2514/2.2137>
- Selig, M. S., Guglielmo, J. J., Broeren, A. P., & Giguère, P. (1995). *Summary of low-speed airfoil data, Volume 1*. SoarTech Publications.
- Selig, M. S., Lyon, C. A., Giguère, P., Ninham, C. P., & Guglielmo, J. J. (1996). *Summary of low-speed airfoil data, Volume 2*. SoarTech Publications.
- Šidlof, P., Antoš, P., Šimurda, D., & Štěpán, M. (2017). Turbulence intensity measurement in the wind tunnel used for airfoil flutter investigation. *EPJ Web of Conferences*, *143*, Article 02107. <https://doi.org/10.1051/epjconf/201714302107>
- Torres, G. E. (2002). *Aerodynamics of low aspect ratio wings at low Reynolds numbers with applications to micro air vehicle design and optimization* (Publication No. 3040583). (Doctoral dissertation). University of Notre Dame, USA. <https://www.proquest.com/docview/305522704?pq-origsite=gscholar&fromopenview=true>
- Tozer, T., & Grace, D. (2001). High-altitude platforms for wireless communications. *IEE Electronics & Communication Engineering Journal*, *13*(3), 127-137. <https://doi.org/10.1049/ecej:20010303>
- Wang, S., Zhang, X., He, G., & Liu, T. (2013). A lift formula applied to low-Reynold-number unsteady flows. *Physics of Fluids*, *25*, Article 093605. <https://doi.org/10.1063/1.4821520>
- Wilcox, D. C. (1988). Reassessment of the scale-determining equation for advanced turbulence models. *American Institute of Aeronautics and Astronautics Journal*, *26*(11), 1299-1310. <https://doi.org/10.2514/3.10041>
- Wolff, E. B. (1925). *Preliminary investigation of the effect of a rotating cylinder in a wing* (No. NACA-TM-307). National Advisory Committee for Aeronautics. <https://ntrs.nasa.gov/api/citations/19930086915/downloads/19930086915.pdf>
- Yao, Q., Zhou, C. Y., & Wang, C. (2016). Numerical study of the flow past a rotating cylinder at supercritical Reynolds number. In *Proceedings of the 2016 4th International Conference on Mechanical Materials and Manufacturing Engineering* (p. 813-816). Atlantis Press. <https://doi.org/10.2991/mmme-16.2016.159>

Risk Assessment on Robotic Surgery Using Bayesian Network

Teh Raihana Nazirah Roslan^{1*} and Chee Keong Ch'ng²

¹*Othman Yeop Abdullah Graduate School of Business, Universiti Utara Malaysia, 50300 Kuala Lumpur, Malaysia*

²*Department of Decision Sciences, School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia*

ABSTRACT

In moving towards Industrial Revolution 4.0, healthcare and medicine are one of the biggest areas of concern which is beneficial to maintaining healthy living. This study seeks to identify the potential problems and risks related to high-technology medical approaches, namely the da Vinci robotic surgical systems, specifically used for thyroidectomy surgery. In particular, the risks embedded in robotic surgeries in terms of health and economy are investigated. Furthermore, a probabilistic risk analysis was conducted to assess the risk among surgeons of the da Vinci robotic surgery using event tree analysis and Bayesian network. This research revealed that the probability of success for surgeons without prior robotic surgery experience was 0.10. It highlights the importance of proper training for medical practitioners in handling advanced medical equipment by considering the related risk involved in patients.

Keywords: Bayes' theorem, event tree analysis, healthcare, high technology medical, probabilistic risk analysis, robotic surgery, thyroid surgery

ARTICLE INFO

Article history:

Received: 21 November 2021

Accepted: 11 March 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.27>

E-mail addresses:

raihana@uum.edu.my (Teh Raihana Nazirah Roslan)

chee@uum.edu.my (Chee Keong Ch'ng)

*Corresponding author

INTRODUCTION

Healthcare and medicine are among the priority areas affected by the current wave of Industrial Revolution 4.0. Some applications in these areas involve surgical instruments, clinical devices, implants, and healthcare equipment such as telemedicine and robotic surgery. Consequently, conventional surgery methods often produce traumatic effects in patients posed a high risk in a large wound,

and longer recovery time will be improved. The robotics application implemented in the medical field assisted the human in terms of helping doctors to perform the complicated task such as conducting surgery on the patient in the operation theatre, which required a long duration of time, deep focus, accuracy, and other activities that the doctors' ability cannot do. The robotic surgery system in healthcare is an advanced development in the surgical area. This great system has been approved to be applied and implemented in treating diseases requiring surgeries.

The most pronounced development is in robotic surgery (Talib, 2017). The only robotic system that is most widely used in procedures of surgical, especially for application in laparoscopic procedures, urological procedures, and mitral valve repair surgery, that has been approved by the Food and Drug Administration (FDA), is the robotic system of da Vinci Surgical System (Ng & Tam, 2014). This type of surgery can reduce pain and discomfort, and among the popular ones is the thyroidectomy surgery conducted through the da Vinci robotic surgical system (Park et al., 2015). The da Vinci System has been widely used worldwide for robotic colorectal surgeries since then (Zakaria et al., 2018).

The da Vinci surgical system refers to the surgery system of the thyroid gland, which involves teleoperating, as shown in Figure 1. The doctors would control a surgical robot comfortably, where the robot's arms would follow da Vinci motions with tools and an endoscope for surgery (Olanrewaju et al., 2013). This system involves three components: a surgeon's console, a patient-side robotic cart including four robotic arms handled by the surgeon, and a 3D high-dimensional vision system (Ng & Tam, 2014). This procedure was safer and feasible compared to traditional open surgery. The duration of such surgery would also be shortened (Park et al., 2015). Moreover, compared robotic versus conventional laparoscopic colorectal surgery, it was found that the robotic approach was as safe and feasible as a conventional one, although it involved higher costs (Ng & Tam, 2014). DeSouza et al. (2010) also claimed that robotic procedure took longer surgery times and greater costs but was found safer and feasible. Ng and Tam (2014) highlighted that robotic surgery improved visualization and enhancement of the images involved with the shorter learning curve and improved musculoskeletal strain to surgeons. Therefore, the application of robotic surgery in thyroidectomy can overcome the traditional thyroid surgery approach's flaws in shortening the learning curve. Here, the learning curve refers to analyzing an individual's stage and total operation time (Park et al., 2015).

Before undergoing robotic thyroidectomy surgery, surgeons must completely understand the thyroid gland's anatomy and the lymph node compartments. Next, surgeons must know how to describe the overall process of thyroidectomy in detail. Then, surgeons will be trained in robotic thyroidectomy for 6 to 12 months. Finally, they can conduct robotic thyroidectomy independently monitored by supervising consultants (Park et al., 2015).



Figure 1. The da Vinci robotic surgery system

In this article, we will assess the risks involved among surgeons when using robotic thyroidectomy surgery using probabilistic risk analysis through the Bayesian network. Zoullouti et al. (2019) successfully utilized a Bayesian approach to healthcare management to investigate the probability of success for surgeons based on previous surgery experiences.

METHODOLOGY

Risk Management in Healthcare

Risk management for healthcare can be defined as identifying, assessing, and mitigating the possible risks to health institutions' visitors, staff, and assets that require organization. Risk management in its best form may be proactive in identifying and managing the risks. However, if an incident happens after the event handling, it should still be tackled in line with the risk management principles outlined in Figure 2 (Alam, 2016).

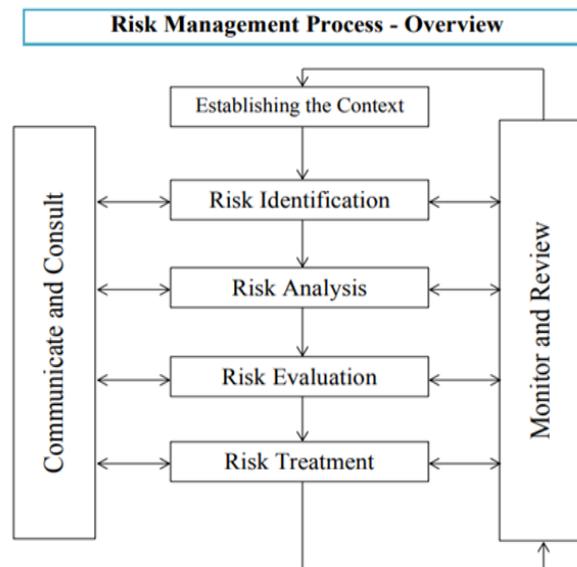


Figure 2. Steps of risk management in healthcare

Five steps are involved in the risk management process, starting by establishing the context. In healthcare risk management, context is crucial. There are many high-priority areas in the hospital, such as ICU (Intensive care unit), O.R. (Operation room), and E.R. (Emergency room), together with other miscellaneous functions related to patients' care. The second step is the identification of risk. This process involves awareness among healthcare professionals and staff on health care services. The third step is analyzing the risks involved. The need to analyze the risk was to develop and understand the risk when it is identified. The fourth step involves evaluating the risk, where underlying causes of the risk, emergency arrangements, and healthcare training are brainstormed and evaluated. The evaluation of risk is divided into two; to accept the risk or to treat the risk. The final step involves treating risk, normally consisting of three methods: controlling, transferring, and avoiding risk. The risk treatment plan should be able to propose actions, mobilizes resources, and establish a timeframe for undertaken actions (Agency for Healthcare Research and Quality, 2013).

The risk management of healthcare was very important to determine and eliminate all the risks. Because predominantly, the underlying causes of medical errors are recognized as communication problems, inadequate information flow, human-related problems, organizational transfer of knowledge, staffing patterns and workflow, inadequate policies and procedures, and technical failures (Agency for Healthcare Research and Quality, 2013).

Sources of Data

Park et al. (2015) obtained the data from a secondary source. This data consists of 125 patients who went for robotic thyroid surgery by two trained surgeons. Before the robotic thyroidectomy, Surgeon 1 had experienced 47 open conventional thyroidectomy surgeries but no experience in endoscopic thyroidectomy or robotic surgery. Meanwhile, Surgeon 2 had more than 200 experiences in open conventional thyroidectomy surgery and five experiences in endoscopic thyroidectomy surgery or robotic surgery. Table 1 illustrates the data that will be used in this project. For this study, data for Surgeon 2 is only for display purposes since our target is Surgeon 1 without former robotic surgery experiences.

From Table 1, 5 variables (age, sex, body mass index (BMI), the extent of surgery, and operation time) were considered. Total operation time was defined as from the first incision to completion of skin closure, which includes docking and undocking robots. Besides that, the learning curve was analyzed by individual-level time and overall operation time. The learning curve was described as the increased performance and experience with time, increasing productivity. The extent of surgery was divided into three groups of surgery. The first surgery group was less than total thyroidectomy, meaning that the surgery group did not involve thyroid patients. Total thyroidectomy was the second surgery group involving only thyroid patients. The third surgery group, total thyroidectomy + MRND, was the

Table 1

Data of patient demographics, extent of surgery, and operation time

	Population (N = 125)	Surgeon 1 (n = 76)	Total	Surgeon 2 (n = 49)	Total
Mean age	39.1	39.5	39.5	38.5	38.5
Sex					
Female	92	58	58	34	34
Male	33	19	19	14	14
BMI, kg/m ²	22.6	22.4	22.4	22.8	22.8
Extent of surgery					
Less than total thyroidectomy	113(90.4%)	72(94.7%)	72	41(83.7%)	41
Total thyroidectomy	9(7.2%)	2(2.63%)	2	7(14.3%)	7
Total thyroidectomy + MRND	3(2.4%)	2(2.63%)	2	1(2.04%)	1
Operation time, min					
Less than total thyroidectomy	100.8	96.0	6912	112.4	4608.4
Total thyroidectomy	134.2	96.5	193	145	1015
Total thyroidectomy + MRND	284.7	234.1	468.2	50.6	50.6

Source. The robotic thyroidectomy learning curve for beginning surgeons with little or no experience of endoscopic surgery (Park et al., 2015)

surgery of patients with thyroid disease and other diseases or symptoms found during the surgery. Here, MRND means Modified Radical Neck Dissection. The operation time was defined in minutes. We calculate the total time of each operation using Equation 1:

$$\sum x = \mu \times n, \tag{1}$$

where μ represents the mean value, x is the total operation time, and n is the sample size.

The mean value for *less than total thyroidectomy* was 100.8, and the sample n was 113. These values gave us Equation 2:

$$\frac{\sum x}{113} = 100.8 \tag{2}$$

$$\sum x = 11390.4.$$

The total operation time for 113 patients was 11390.4 minutes. Besides that, for *total thyroidectomy* with the mean value of 134.2 and $n = 9$, we obtain Equation 3:

$$\frac{\sum x}{9} = 134.2 \quad (3)$$

$$\sum x = 1207.8.$$

Thus, the total operation time for nine patients was 1207.8 minutes. Besides, the calculation for *total thyroidectomy* + *MRND* was repeated as above, using the mean of 284.7 and $n = 3$ in Equation 4:

$$\frac{\sum x}{3} = 284.7 \quad (4)$$

$$\sum x = 854.1.$$

Here, the total operation time for three patients was 854.1 minutes.

Appendix A shows the framework of robotic healthcare, where successful robotic surgery requires details of operation time and extent of surgery. The surgery was considered successful when the robotic surgery's operation time was less than the history data or the duration of the surgery without using robots. It also means that robotic surgery is more advantageous than conventional surgery. For example, robotic thyroidectomy eliminated the need for anterior neck incision for patients, resulting in reduced pain and swallowing discomfort compared to conventional thyroidectomy (Lee et al., 2010). Perez and Schwautzberg (2019) mentioned that the increased cost of robotic surgery was partly due to the high cost of fixed equipment. Robotic surgery could be cost-effective if these fixed costs were spread to a higher volume. Weaver and Steele (2016) stated that the learning curve, particularly for trained laparoscopic surgeons, is expected to be shorter than conventional laparoscopic surgery as robots are meant to be more intuitive than laparoscopy, although the curve is still lengthy and profound. Besides that, Mattos (2016) said that robots would enable more accurate and safer operations by increasing surgeons' agility, control, and accuracy.

Event Tree Analysis

Spouge (1999) proposed event tree analysis (ETA) as an inductive procedure that shows all possible outcomes of an accident, normally based on two major assumptions. First, the likelihood of events or basic events is assumed to be exact and precisely known. Second, interdependencies of events or basic events are assumed, independent. These assumptions are understandably not always true. There were situations when some inherent uncertainties were available in data collection and defining the relationships of events or basic events (Sadiq et al., 2008). Sometimes, the events surrounding trees may depend on one another (Ferson, 2004). In this article, a predictive approach was applied where the number of

accidental events X in some specified operations was focused. A link g between X and observables on a more detailed system level was established and denoted in $Z = (Z_1, Z_2, \dots, Z_m)$. The number of hazardous situations of a certain type occurring during some operations was indicated as Z_i . 1 was denoted as specific safety barrier fails, while 0 otherwise—for example, the model given in Figure 3 consists of Z_1, Z_2 , and Z_3 . The number of hazardous situations to occur refers to Z_1 , while Z_2 indicates 1 if the first safety barrier fails and 0 otherwise. Next, Z_3 is equal to 1 if the second safety barrier fails and 0 otherwise (Aven & Eidesen, 2007).

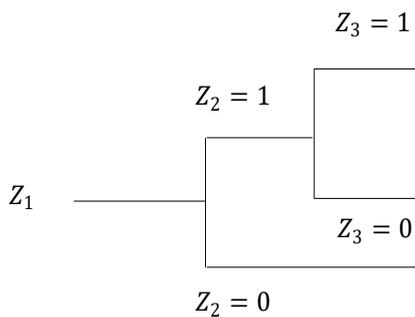


Figure 3. An event tree

$X = g(Z) = Z_1 \cdot Z_2 \cdot Z_3$, as X is given by the number of hazardous situations where both safety barriers fail.

Bayesian Network

The applications of Bayesian are already found in a wide range of activities, including in sciences, engineering, medicine, sport, and others. Lately, a Bayesian network strategy has started to be utilized in designing applications. A Bayesian network is a graphical induction method to communicate

the causal connections among factors. Bayesian networks are utilized either to anticipate the likelihood of obscure factors or to refresh the likelihood of realized factors given the specific condition of different factors through the course of likelihood spread or thinking. The modeling depends on Bayes’ theorem. Because of this capacity, Bayesian networks have given a promising structure to a framework for security investigation and hazard the executives. Bayesian networks are progressively utilized for the development of framework unwavering quality models, hazard the executives, and wellbeing examination considering probabilistic and dubious information. Like fault trees, Bayesian networks comprise both subjective and quantitative parts.

Many researchers have examined various methods in mishap situation examination, not many of whom have involved Bayesian networks. Sklet (2004) subjectively looked at a few ordinarily utilized techniques, for example, fault tree examination, occasion tree investigation, and hindrance examination for mishap investigation. The examination was made in view of standards, for example, graphical portrayal and the capacity to help security boundaries. Nivolianitou et al. (2004) utilized fault tree, occasion tree, and Petri nets for a subjective mishap situation examination in an alkali stockpiling plant, reasoning that Petri nets can fuse the proof through situation investigation and consequently are more fitting for dynamic mishap investigation. Zheng and Liu (2009) made a correlation

among a few broadly involved strategies for mishap anticipating. Although fault tree as a situation examination strategy and Bayesian networks were momentarily talked about, the principal center in their exploration was given to techniques, for example, relapse models, time-series techniques, and neural organizations.

Additionally, Simon et al. (2007) gave a comprehensive factual survey of Bayesian network application in various regions like gambling and upkeep investigation, in which Bayesian network was subjectively contrasted and different techniques, for example, fault trees, Markov chains, and Petri nets. The current work is pointed toward showing the equals among fault trees and Bayesian networks in the particular area of mishap displaying and process security examination, which has not been concentrated on so far. The paper additionally examines the demonstrating potential presented by Bayesian networks, making them a better technique looked at than fault trees for dynamic security investigation.

According to Oppermann (2018), in finance, the Bayes' theorem can be used to rate the risk of lending money to potential borrowers. In the healthcare field, the Bayesian network can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test. In this work, we will find the probability of *total thyroidectomy + MRND* occurring, given that the event of *total thyroidectomy + MRND* for Surgeon 1 had occurred. The general formula for Bayes' theorem is as Equation 5:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (5)$$

where

$P(A|B)$ = the probability of event A occurring, given event B has occurred

$P(B|A)$ = the probability of event B occurring, given event A has occurred

$P(A)$ = the probability of event A

$P(B)$ = the probability of event B

RESULTS AND DISCUSSION

In assessing the risks, all aspects of the event need to be identified to improve the field of robotic thyroidectomy surgery. Therefore, we begin with the posed risks, followed by the assessment through event tree analysis and the Bayesian network for probabilistic risk analysis. In this era of globalization, the government has introduced many new technologies in the field of healthcare to provide greater efficiency in services.

Health Risk

The health risk is the health consequence of a specific disease or condition. The robotic surgery impacted the precise dissection in a confined space and reduced in blood loss of the

patients. The other complications after surgery, like lower transfusion rates and death rates in the hospital, can be reduced through the robotic surgery approach. Moreover, a shorter learning curve can be achieved (Ng & Tam, 2014). This approach's noticeable drawback is the speed of transferring information from the operator to the robot and the lag time from operator to execution involving surgeon and patient in a different location (Olanrewaju, 2013). The difficult robotic surgery procedures can be done quickly as in open surgery by competent surgeons and nurses. The operating time was sped up with the help of a regular assistant—the surgical training for robotic surgery using VR simulators in the operating room limit the operative time. However, the acceleration of the learning curve in real robotic surgery does not directly imply the improvement of da Vinci's performance. Using the VR simulation of the robotic surgery tasks performed by the trainee, less psychomotor stress and actively manipulated clutch and camera pedals of the robotic system were improved (Cho et al., 2013). However, this surgery could be smoothly conducted when surgery team members became familiar and comfortable with the procedures (Sahabudin, 2006).

Economic Risk

The economic risk involved is the high cost of the da Vinci robotic thyroidectomy apparatus, but this risk could become cost-effective when the employability rate of such an approach is high (Park et al., 2015). Even though the current costs are high, the wider dissemination of this technology and the increase in competition from manufacturers may drive the costs down (Ng & Tam, 2014). This issue is also related to the training needed for surgeons in the operating room (Cho et al., 2013). Training for the trainees to conduct robotic surgery is essential, specifically in suturing techniques. A short course is insufficient for those who do not have laparoscopic experience (Sahabudin, 2006).

Moreover, financial limitations are found in applying the da Vinci robotic surgery system. The cost involved in using robotic surgery is a higher burden in terms of limited availability, cost, and the learning curve of robotic surgery. Since the da Vinci robotic surgery system applied proprietary software, which physicians cannot modify, there is no freedom for users to modify the standard operating system. The jobs that relate to intellectual capitalism, creativity, imagination, leadership, analysis, humor, common sense, screen or script-writing, and scientific endeavors will be volatile against the technological revolution. Another risk of this technology is that it might be able to replace labor tasks and roles, reducing human function that may leading to unemployment (Zakaria, 2018).

Event Tree Analysis

In this section, we establish the links between the parameters used based on the operation time in surgery using robotic thyroidectomy. Event tree analysis is a model used to show

logical numbers for failure and success from a population or individual (Freeman, 2022). The general event tree in this study is as follows:

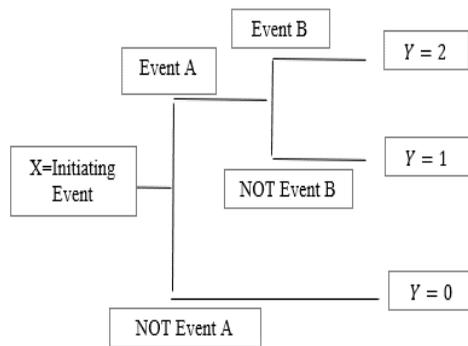


Figure 4. The general event tree

Figure 4 portrays the possible pathways of this study. We assess the *total thyroidectomy + MRND* as $Y = 2$ when both events succeed, *total thyroidectomy* as $Y = 1$ when one event gains success and the other does not occur, and *less than total thyroidectomy* as $Y = 0$, which is a failure or does not achieve their total operation time. To perform the event tree analysis, we need to perform the following steps:

Step 1: The system needs to be defined by stating the variables involved.

Total thyroidectomy + MRND,
total thyroidectomy, and
less than total thyroidectomy

Step 2: The accident scenarios were identified by performing a system assessment for each event.

$Y = 2$ (*total thyroidectomy* occurs, and *MRND* occurs)

$Y = 1$ (*total thyroidectomy* occurs, and *MRND* does not occur)

$Y = 0$ (*less than total thyroidectomy*)

Step 3: Define the initiating event by using the main sources of events.

Operation time of robotic thyroidectomy surgeries for population

Step 4: Identify Events A and B, either failure or success.

Event A = *total thyroidectomy*

Event B = *MRND*

Step 5: Calculate the overall probability for each path.

Based on the data in Table 1, the total operation time of the overall event was 13452.3 minutes. It was obtained by summing up the operation time for *total thyroidectomy + MRND*, which was 854.1 minutes, *total thyroidectomy*, which was 1207.80 minutes, and *less than total thyroidectomy* was 11390.40 minutes, respectively. Therefore, to calculate the probability of each event occurring, we need to use every event's minute, then divide by overall minute by a population of all events, as shown in Equation 6.

$$P(\text{Total Thyroidectomy} + \text{MRND}) = \frac{854.1 \text{ minutes}}{13452.3 \text{ minutes}} = 0.0635$$

$$P(\text{Total Thyroidectomy}) = \frac{1207.80 \text{ minutes}}{13452.3 \text{ minutes}} = 0.0897$$

$$P(\text{Less than Total Thyroidectomy}) = \frac{11390.40 \text{ minutes}}{13452.3 \text{ minutes}} = 0.8467$$

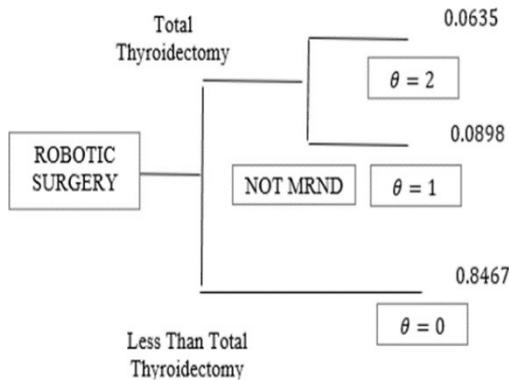


Figure 5. Event tree analysis in robotic thyroidectomy

When Events A and B succeed or take place, we choose $\theta = 2$. Meanwhile, when one event occurs, and the other event does not occur, we state $\theta = 1$. Finally, if Events A and B did not occur, we set $\theta = 0$.

From Figure 5, it can be observed that the condition for *total thyroidectomy* occurring in Event A and *MRND* occurring in Event B implies that Event A and Event B also took place. Here, the value is 0.0635 for $\theta = 2$. For the case when Event A did not occur, it would be considered $\theta = 0$ and

recognized as *less than total thyroidectomy* with a value of 0.8467. The last event was when Event A occurred, but Event B did not occur; here, θ was denoted as 1. This event was for the surgery population in robotic thyroidectomy if *total thyroidectomy* occurred, in which the surgery was just for thyroid but not for *MRND*. The value calculated was 0.0897.

Bayesian Network

The Bayesian network allows combining prior information about population parameters with evidence from the information contained in the sample to guide the statistical inference process (Gleason & Harris, 2019). This method allows us to integrate information from historical data and the parameters of the population undergoing thyroid surgery. The information from the past population is defined as the prior distribution for thyroidectomy surgery using robotics. We call the information from the past population the prior distribution for Thyroidectomy Surgery using Robotic for IR 4.0. In this study, all surgeons showed success in handling robotic surgery. Therefore, we assumed that the experience from Surgeon 1 would give a positive response mean a positive result to a person who undergoes thyroidectomy surgery. We later look at the angle from which operation time for a surgeon that successfully performed this operation using the same steps as the previous history.

Based on Table 1, when Surgeon 1 performs *total thyroidectomy* for Event A but at the same time Event B, which is *MRND* occurred, the value for this situation was 468.2

times. On the other hand, if Surgeon 1 performs only Event A (*total thyroidectomy*), and the patient was not required to undergo surgery for *MRND* (Event B), the value obtained was 193. Lastly, when Surgeon 1 was performing the *less than total thyroidectomy* surgery, the value was 6912 operation time. All these values can be written using conditional probabilities as Equations 7 and 8:

$$\begin{aligned} (MRND \text{ (by surgeon 1)} \cap \text{Total Thyroidectomy}) &= 468.2 \\ (\text{Surgeon 1} \cap \text{Total Thyroidectomy}) &= 193 \\ (\text{Surgeon 1} \cap \text{Less than Total Thyroidectomy}) &= 6912 \end{aligned} \tag{7}$$

We used the results for each event as follows:

$$P(B | A) = \frac{P(B \cap A)}{P(A)}$$

$$\begin{aligned} (MRND \text{ (by surgeon 1)} \cap \text{Total Thyroidectomy}) &= 468.2 \\ (MRND \text{ (by surgeon 1)} \cap \text{Total Thyroidectomy}) &= 468.2 \\ P(MRND \text{ Operation time By surgeon 1} | \text{Total Thyroidectomy} + MRND) &= \frac{468.2}{854.1} = 0.54818 \\ P(\text{Operation time By surgeon 1} | \text{Total Thyroidectomy}) &= \frac{193}{1207.80} = 0.1598 \\ P(\text{Operation time By surgeon 1} | \text{Less than Total Thyroidectomy}) &= \frac{6912}{11390.40} = 0.6068 \end{aligned} \tag{8}$$

By using the

$$\begin{aligned} P(MRND \text{ Operation time By surgeon 1} | \text{Total Thyroidectomy} + MRND) &= \\ P(X = 1 | \theta = 2) &= 0.54818 \\ P(\text{Operation time By surgeon 1} | \text{Total Thyroidectomy}) &= P(X = 1 | \theta = 1) = 0.1598 \\ P(\text{Operation time By surgeon 1} | \text{Less than Total Thyroidectomy}) &= \\ P(X = 1 | \theta = 1) &= 0.6068 \end{aligned} \tag{9}$$

By combining data on the history and experience of Surgeon 1 (Equation 9), we can compute the posterior probability for the probability of patients having *Total Thyroidectomy* + *MRND* undergoing surgery by Surgeon 1, denoted as $P(\theta = 2 | X = 1)$.

To achieve this, first, we calculate $P(X=1)$ in Equation 10:

$$\begin{aligned}
P(X = 1) &= P(X = 1 | \theta = 2)P(\theta = 2) + P(X = 1 | \theta = 1)P(\theta = 1) \\
&+ P(X = 1 | \theta = 0)P(\theta = 0) \\
P(X = 1) &= 0.54818(0.0635) + (0.1598)(0.0898) + (0.6068)(0.8467) \\
&= (0.03481) + (0.01435) + (0.51378) = 0.56294
\end{aligned} \tag{10}$$

Here, $P(X = 1)$ means the operation time performed by Surgeon 1. By using historical data $P(X = 1 | \theta = 2) = 0.54818$.

$$P(\theta = 2 | X = 1) = \frac{P(X = 1 | \theta = 2)P(\theta = 2)}{P(X = 1)} = \frac{(0.54818)(0.0635)}{0.56294} = 0.10 \tag{11}$$

The result for posterior distribution when using the Bayes' Theorem Analysis was 0.10 (Equation 11). This result shows that the probability of success by Surgeon 1 when performing *Total Thyroidectomy + MRND surgery* was 0.10. Therefore, based on the history data in Park et al. (2015), Surgeon 1 did not have experience handling robotic surgery, and his success probability in this operation was 0.10.

CONCLUSION

In this study, we used event tree analysis and the Bayesian network to show probability of the learning curve (operation time) for Surgeon 1 handling the *Total Thyroidectomy + MRND surgery*. This approach can be applied to other robotic or technological equipment that requires large funding for a better understanding of the said innovation. Although expensive, the government should consider investing in this technology as it reflects more advantages, such as reducing the potential of unnecessary risk of injuries in normal surgery. Furthermore, constructive strategies should be developed to provide necessary training tools for surgeons with minimal training costs. This type of technology will be potential in the future, where its adoption will produce safe and efficient procedures with minimal damage. Patients will also reduce hospitalization; thus, the bed occupancy in major hospitals can be reduced to cater to more emergency matters.

ACKNOWLEDGEMENT

This research was self-supported by the authors.

REFERENCES

AHRQ. (2013). *National healthcare disparities report*. Agency for Healthcare Research and Quality. <http://www.ahrq.gov/research/findings/nhqrdr/nhdr13/chap4.html>

- Alam, A. Y. (2016). Steps in the process of risk management in healthcare. *Journal of Epidemiology and Preventive Medicine*, 2(2), 1-5.
- Aven, T., & Eidesen, K. (2007). A predictive Bayesian approach to risk analysis in health care. *BMC Medical Research Methodology*, 7(1), 1-8.
- Cho, J. S., Hahn, K. Y., Kwak, J. M., Kim, J., Baek, S. J., Shin, J. W., & Kim, S. H. (2013). Virtual reality training improves da Vinci performance: A prospective trial. *Journal of Laparoendoscopic and Advanced Surgical Techniques*, 23(12), 992-998. <https://doi.org/10.1089/lap.2012.0396>
- DeSouza, A. L., Prasad, L. M., Park, J. J., Marecik, S., Blumetti, J., & Abcarian, H. (2010). Robotic assistance in right hemicolectomy: Is there a role? *Journal of Disease of the Colon & Rectum*, 53(7), 1000-1006. <https://doi.org/10.1007/DCR.0b013e3181d32096>
- Ferson, S., Hajagos, J., Berleant, D., Zhang, J., Tucker, W. T., Ginzburg, L., Nelsen, R., Oberkampf, W. L., & Hall, C. (2004). *Dependence in Dempster-Shafer theory and probability bounds analysis*. Sandia National Laboratories.
- Freeman, J. (2022). *Event tree analysis - The risk assessment application tool*. Edraw. <https://www.edrawsoft.com/event-tree-introduction.html>
- Gleason, P. M., & Harris, J. E. (2019). The Bayesian approach to decision making and analysis in nutrition research and practice. *Journal of the Academy of Nutrition and Dietetics*, 119(12), 1993-2003. <https://doi.org/10.1016/j.jand.2019.07.009>
- Lee, J., Nah, K. Y., Kim, R. M., Ahn, Y. H., Soh, E. Y., & Chung, W. Y. (2010). Differences in postoperative outcomes, function, and cosmesis: Open versus robotic thyroidectomy. *Surgical Endoscopy*, 24, 3186-3194. <https://doi.org/10.1007/s00464-010-1113-z>
- Mattos, L. S., Caldwell, D. G., Peretti, G., Mora, F., Guastini, L., & Cingolani, R. (2016). Microsurgery robots: Addressing the needs of high-precision surgical interventions. *Swiss Medical Weekly*, 146, 1-14. <https://doi.org/10.4414/smw.2016.14375>
- Ng, A. T. L., & Tam, P. C. (2014). Current status of robot-assisted surgery. *Hong Kong Medical Journal*, 20(3), 241-250. <https://doi.org/10.12809/hkmj134167>
- Nivolianitou, Z. S., Leopoulos, V. N., & Konstantinidou, M. (2004). Comparison of techniques for accident scenario analysis in hazardous systems. *Journal of Loss Prevention in the Process Industries*, 17(6), 467-475. <https://doi.org/10.1016/j.jlp.2004.08.001>
- Olanrewaju, O. A., Faieza, A. A., & Syakirah, K. (2013). Current trend of robotics application in medical. *IOP Conference Series: Materials Science and Engineering*, 46, Article 012041. <https://doi.org/10.1088/1757-899X/46/1/012041>
- Oppermann, A. (2018). *Bayes' theorem: The holy grail of data science*. Towards Data Science. <https://towardsdatascience.com/bayes-theorem-the-holy-grail-of-data-science-55d93315defb>
- Park, J. H., Lee, J., Hakim, N. A., Kim, H. Y., Kang, S. W., Jeong, J. J., & Chung, W. Y. (2015). Robotic thyroidectomy learning curve for beginning surgeons with little or no experience of endoscopic surgery. *Head and Neck*, 37(12), 1705-1711. <https://doi.org/10.1002/hed.23824>

- Perez, R. E., & Schwaitzberg, S. D. (2019). Robotic surgery: Finding value in 2019 and beyond. *Annals of Laparoscopic and Endoscopic Surgery*, 4(3), 1-7. <http://dx.doi.org/10.21037/ales.2019.05.02>
- Sadiq, R., Saint-Martin, E., & Kleiner, Y. (2008). Predicting risk of water quality failures in distribution networks under uncertainties using fault-tree analysis. *Urban Water Journal*, 5(4), 287-304. <https://doi.org/10.1080/15730620802213504><https://doi.org/10.12809/hkmj134167>
- Sahabudin, R. M., Arni, T., Ashani, N., Arumuga, K., Rajenthiran, S., Murali, S., & Menon, M. (2006). Development of robotic program: An Asian experience. *World Journal of Urology*, 24, 161-164. <https://doi.org/10.1007/s00345-006-0069-z>
- Simon, C., Weber, P., & Levrat, E. (2007). Bayesian networks and evidence theory to model complex systems reliability. *Journal of Computers*, 2(1), 33-43.
- Sklet, S. (2004). Comparison of some selected methods for accident investigation. *Journal of Hazardous Materials*, 111(1-3), 29-37. <https://doi.org/10.1016/j.jhazmat.2004.02.005>
- Spouge, J. (1999). *A guide to quantitative risk assessment for offshore installations*. CMPT Publication.
- Talib, Y. Y. A. (2017, August 28). Pembedahan robotic [Robotic surgery]. *Harian Metro*. <https://www.hmetro.com.my/hati/2017/08/258470/pembedahan-robotik>
- Weaver, A., & Steele, S. (2016) Robotics in colorectal surgery. *F1000 Research*, 5, Article 2373. <https://doi.org/10.12688/f1000research.9389.1>
- Zakaria, A. D., Toh, J. W. T., & Kim, S. H. (2018). Future perspectives in robotic colorectal surgery. In N. Kim, K. Sugihara & J. T. Liang (Eds.), *Surgical treatment of colorectal cancer* (pp. 315-325). Springer. https://doi.org/10.1007/978-981-10-5143-2_29
- Zheng, X., & Liu, M. (2009). An overview of accident forecasting methodologies. *Journal of Loss Prevention in the Process Industries*, 22(4), 484-491. <https://doi.org/10.1016/j.jlp.2009.03.005>
- Zoullouti, B., Amghar, M., & Nawal, S. (2019). Using Bayesian networks for risk assessment in healthcare system. In D. McNair (Ed.), *Bayesian networks: advances and novel applications* (pp. 39-53). IntechOpen. <https://doi.org/10.5772/intechopen.80464>



Review Article

A Review on the Effect of Zein in Scaffold for Bone Tissue Engineering

Istikamah Subuki^{1,2*}, Khairun Nor Ashikin Nasir¹ and Nur Azrini Ramlee^{1,2}

¹*School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia*

²*Circular Industrial Research Laboratory, School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia*

ABSTRACT

Natural pharmaceutical ingredients have been widely used in recent decades due to their safety and biocompatibility. Zein, a plant-derived natural protein, has several advantages over other synthetic polymers in bone tissue engineering (BTE). This study of zein protein focuses more on its application in BTE as potential biopolymer material used in scaffold development. The use of zein in BTE has shown its benefits in the production of scaffolds. Therefore, attention has been given to studies of the effect of zein usage in bone scaffold development, as it offers a great ability based on its porosity, mechanical strength, in vitro degradation study, cell proliferation, and osteogenic differentiation, which is important for healing bone tissue damage. Therefore, this review aims to critically analyze the current research on the method of scaffold fabrication and the effect of zein usage in scaffolds for BTE. In addition, the common methods used in creating the scaffold are addressed.

Keywords: Bone tissue engineering, protein, polymer, scaffold, zein

ARTICLE INFO

Article history:

Received: 22 November 2021

Accepted: 14 February 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.28>

E-mail addresses:

istikamah@uitm.edu.my (Istikamah Subuki)

khairunashikin7@gmail.com (Khairun Nor Ashikin Nasir)

azrini@uitm.edu.my (Nur Azrini Ramlee)

*Corresponding author

INTRODUCTION

Zein consists of 45–50% of the protein in maize and is the primary storage protein of maize. It was first identified in 1897 due to its solubility in aqueous alcohol solutions (Shukla & Cheryan, 2001). Corn, often known as maize, is the only cereal crop native to the Americas and one of the essential food and industrial crops in the United States (Shukla & Cheryan, 2001).

Because of its negative nutrient supply and low water solubility, zein isolate is not instantly consumed by humans (Labib, 2018). However, commercial value has been placed on the ability of zein and its resins to make strong, glossy, hydrophobic grease-proof coatings resistant to microbes. Zein offers promising applications in fiber, adhesives, coatings, ceramics, paint, cosmetics, textiles, chewing gums, and biodegradable plastics (Shukla & Cheryan, 2001; Subuki et al., 2018). Table 1 shows the usage of zein in the earliest years of its discovery. Zein can be categorized into four main groups based on its solubility and sequence homology which is α -zein, accounting for 70–85% of the total zein mass; β -zein, and γ -zein, the second most common component; and δ -zein (Esen, 1987). Several hydrophobic and neutral amino acids, such as leucine, proline, and alanine, are present in all zein classes, as well as some polar amino acid traces such as glutamine. Zein differs from other proteins as it contains a very small amount of lysine and tryptophan, including a few arginine and histidine residues. In addition, it has unique solubility characteristics due to its amino acid composition, which is limited to acetone, acetic acid, aqueous alcohols, and aqueous alkaline solutions (Lawton, 2002). The features of zein-based formulations have shown remarkable properties of this natural material, including excellent heat, water, abrasion, and humidity resistance. In addition, zein protein from maize can increase the potential of biomolecules having a longer shelf-life (Tortorella et al., 2021).

In this manuscript, the effect of zein in scaffolding for bone tissue engineering (BTE) is assessed in view of the characteristics and advantages of zein as the material used, scaffolding fabrication, and current research on the zein addition effect in BTE.

Table 1

Zein usage in the industry (Shukla & Cheryan, 2001)

Applications	Year
Adhesive, binders	1944
Fibers	1950
Printing inks	1951
Medical tablets coating	1970
Drug controlled release application	1984
Paper surfaces, glossy magazine covers	1994
Biodegradable plastics, films	1997
Chewing gum	1999
Coating	1999

Current Research on Zein Protein in Bone Tissue Engineering (BTE)

The present study of zein protein focuses more on its application in bone tissue engineering (BTE) as a potential biopolymer material used in scaffold development. The use of Zein

in BTE has shown benefits in the production of scaffolds, where it is required in BTE application to build adequate scaffolding materials that can help heal tissue damage by rejuvenating new tissues (Bharadwaz & Jayasuriya, 2020; Gong et al., 2006). Biomaterials, cells, and signaling factors are the combination for BTE to heal or restore bone defects through scaffolding. Figure 1 shows the steps involved in the repair or regeneration process: (1) Culture cells and create a scaffold; (2) Introduce the cells into the scaffold, creating a “cell-scaffold” structure; (3) Implantation; (4) Formation of new bone and degradation of the scaffold; and (5) Bone repair is achieved (Tong et al., 2017).

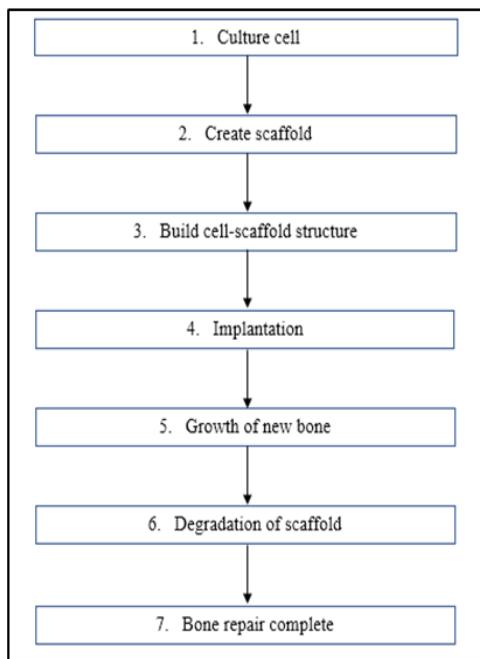


Figure 1. Steps involved in the repair or regeneration process (Tong et al., 2017)

scaffold can allow the cells to bind, normal function, proliferate, differentiate, and create new cell structures (Ghassemi et al., 2018). Scaffold, transplant, template, or artificial ECM are all terms for support. Biocompatibility is essential in a scaffold (Qu et al., 2019). Biocompatibility of a scaffold is also related to the condition of the pores where the required pore size is sufficient to allow for cell migration and is adequate for the cell binding of the scaffold (Chocholata et al., 2019). A highly porous microstructure with interconnected pores and a wide surface area is widely recognized to be favorable for tissue ingrowth in the scaffold. Pore diameters between 100 and 350 μm and porosities greater than 90% are preferable for bone regeneration (Yoshimoto et al., 2003).

Biodegradable polymers based on proteins are widely used in biomedical applications because they show excellent biocompatibility, particularly those similar to the components of the extracellular matrix (ECM) (Demir et al., 2017). Therefore, zein protein has been widely studied and used as a scaffold material; outcomes of the research have shown acceptable mechanical properties and degradation portrayed by the zein-based scaffold (Hum & Boccaccini, 2018; Jing et al., 2018; Qu et al., 2008). Most scaffolds are analyzed on their biocompatibility, biodegradability, mechanical properties, and degradation rate. In the design of bone scaffolds, the following considerations should be taken:

Biocompatibility

In BTE application, biocompatibility is one of the key characteristics of a scaffold, as the

Biodegradability

The scaffold must degrade in a timely manner after implantation to facilitate adequate tissue regeneration (Bitar & Zakhem, 2014). In addition, biodegradable scaffolds have gained much attention because of their irremovable features after the first implantation surgery, as they may be absorbed naturally by the body (Patel & Fisher, 2008). In summary, materials from scaffold degradation must also be non-toxic, so they will not cause any harm to the human body during the degradation process (Qu et al., 2019; Tariverdian et al., 2019).

Mechanical Strength

Next, in the creation of a scaffold for BTE application, mechanical strength is an important characteristic that is mostly regulated by the volume or size of the pore (Bose et al., 2013; Roseti et al., 2017). Therefore, the scaffold must provide sufficient mechanical strength to enable new tissue replacement (Prasadh & Wong, 2018). Cortical bone has a strength of 100 to 230 MPa on average. Scaffold biomaterials should fall within this range. However, the scaffold is supported by plates, screws, wires, or pins to avoid failure (Prasadh & Wong, 2018). Table 2 shows some characteristics of the human bone.

Table 2

Mechanical properties of human bone tissue (Shamaz & Halima, 2015)

Property	Cortical bone	Cancellous bone
Compressive strength (MPa)	100–230	2–12
Tensile strength (MPa)	50–150	10–20
Strain to failure (%)	1–3	5–7
Young's modulus (GPa)	7–30	0.5–0.05

Advantages of Zein Proteins for Bone Tissue Engineering Compared to Other Proteins.

Many commonly used proteins act as polymers in BTE application, including collagen, silk, zein, soy protein, and wheat gluten. Zein is categorized as one of the plant-based protein biopolymers, which makes it preferable to other proteins. In comparison to animal proteins and synthetic polymers, plant proteins have various advantages, including relatively inexpensive, safe, biocompatibility, and the ability to be manufactured from renewable sources (Berardi et al., 2018; Chen et al., 2019; Reddy & Yang, 2013). Moreover, zein can be processed using a variety of solvents compared to other plant-based proteins, which makes them easy to handle (Anderson & Lamsa, 2011; Reddy & Yang, 2011). In addition, in comparison to wet mechanical properties, zein protein shows a better performance than other proteins, as it is an important property because the scaffold is accessible to the physiological fluid inside the body (Maji & Dasgupta, 2017).

FABRICATION OF SCAFFOLD

The use of zein in BTE has shown to be advantageous in the production of scaffolds. Zein is classified as a natural polymer that can be used as one of the biomaterials for scaffold fabrication. Bio-functional elements in the polymers ensure bioactivity, a biomimetic surface, and natural remodeling. However, immunogenicity, microbial contamination (i.e., endotoxin), limited tunability, unstable degradation rate, and low mechanical strength are among their key drawbacks, which limit its application in bone tissue engineering (Francesca et al., 2020). Therefore, researchers often combine two or more materials with diverse qualities to enhance the properties of materials. Examples of these combinations are copolymers, polymer-polymer blends, and polymer–ceramic.

Bone scaffolds can be created in a few ways and categorized into two types which are conventional and additive manufacturing (AM). Conventional techniques have limitations in building structures with complex porosity (Bajaj et al., 2014). AM techniques in producing three-dimensional scaffolds have progressed to the point that they can now construct complicated scaffolds. Furthermore, because harmful solvents are not employed in making three-dimensional (3D) scaffolds, biocompatibility issues observed in some conventionally manufactured scaffolds are no longer an issue (Torabi et al., 2015). The three-dimensional scaffold design is sent to the machine using computer-aided design (CAD) and computer-aided manufacturing (CAM) software in AM technology used to create the scaffold (Sah & Pramanik, 2011).

Salt Leaching

The salt leaching method entails mixing the polymer with a water-soluble porogen, such as sodium chloride or sodium citrate, after dissolving it in an organic solvent and then transferring the mixture into a mold (Chocholata et al., 2019; Mikos & Temenoff, 2000; Rezwani et al., 2006).

In 2006, a study conducted by Gong et al. (2006) used the salt leaching method using several different measurements of the porogen's mass fraction and particle size. First, the mixture was shaped into three-dimensional scaffolds, with different particle sizes ranging from 38.5 to 220 nm and mass fraction ratio of 1:1.4 and 1:2.5 w/w (zein:NaCl). Then, the samples were leached using a water bath and lyophilized. Scaffolds were 10 mm diameter cylindrical rods with 3–25 mm height. Additionally, this method has been employed by Wang et al. (2008) to demonstrate the application of zein in bone tissue engineering, where porous zein scaffolds were produced using the salt leaching technique based on its adhesive characteristics and solubility, which differ significantly from other native protein biomaterials and demonstrate remarkable compressive performance.

Meanwhile, another study on using PCL and zein in the fabrication of scaffolds for BTE application was conducted by Wu et al. (2012). The sample was made by employing

a solvent casting–particulate leaching process with sodium chloride particles to combine poly(ϵ -caprolactone) (PCL) with zein. The method used is quite similar to work done by Gong et al. (2006). In addition, Hum (2016) utilized the leaching approach to create porous zein scaffolds, with salt as the porogen. As a result, zein and sodium chloride particles were homogeneously mixed. Subsequently, the mixture was compressed into cylindrical specimens. Finally, the samples were leached in a water bath, rinsed in ultrapure water, then lyophilized for later usage to obtain a porous structure.

Gas Foaming

Gas foaming is a way to avoid utilizing harmful organic solvents. Gas-foaming agents, such as carbon dioxide and nitrogen, are employed in this approach, known as non-flammable, low-toxic gas (Ji et al., 2011; Mikos & Temenoff, 2000). This method allows high-pressure water to enter biodegradable polymers, resulting in a porous structure with an average pore size of 30–700 μm and a porosity of up to 85% (Thavorniyutikarn et al., 2014b). In addition, it is an easy and cost-effective scaffolding method (Nam et al., 2000). This method was used by Salerno et al. (2010) to prepare the multi-phase PCL/TZ and PCL/TZ–hydroxyapatite particles (HA) scaffold. Firstly, in a batch foaming device, chosen protein or plasticizer systems were foamed with CO_2 and N_2 mixtures as blowing agents (Salerno et al., 2006). Salerno et al. (2010b) studied using the CO_2 foaming method. The scaffold samples were created using PCL, thermoplastic zein (TZ), and HA. Another study by Salerno et al. (2011) used a combination of gas foaming–leaching methods to create porous multi-phase PCL–TZ and PCL–TZ–HA composite scaffolds. PCL, TZ, and HA were first mixed in an internal mixer. The most recent study on the usage of zein in BTE applications using the supercritical CO_2 foaming method was conducted by Subuki et al. (2020).

3D Printing

Several commonly used approaches for creating these 3D scaffolds use a conventional method. However, they all have the same drawbacks, including poor scaffold design, pore network and size, and unsatisfactory 3D scaffolds (An et al., 2015). In addition, the processes of these methods are not adaptable enough (Gungor-Ozkerim et al., 2018). On the other hand, rapid prototyping, solid free-form fabrication, biofabrication, bioprinting, and additive manufacturing are all 3D printing processes that could tackle all the issues regarding the usage of conventional methods (Hospodiuk et al., 2016).

Three-dimensional (3D) printing is additive manufacturing (AM) method, and fast prototyping is a method of layer-by-layer combining of materials to create products from 3D model data, as opposed to subtractive manufacturing methods (Wang et al., 2020). Ru et al. (2018) study create ternary scaffolds using 3D printing. The scaffolds were created using three components: zein (ZN), PCL, and nano magnesium silicate (n-MS). The ingredient was chosen due to their advantages given to the BTE application.

Freeze Drying

Freeze drying, also known as emulsification, is a promising scaffold preparation procedure. Freeze drying works on sublimation, in which frozen water in polymer nanocomposites is immediately transformed from solid to a gas state without liquefaction (Wahid et al., 2018).

The resulting pores using this method are constantly produced, but their porosity is greater than 90%, and their size ranges from 20 to 200 μm (Xiaohao & Peter, 2004). The emulsion freeze-drying process can be paired with particle leaching, sucrose, or sodium chloride, which can be added to the emulsion to create porosity. Particles can be cleaned after freeze drying (Alizadeh et al., 2013). This method was used by Shahbazarab et al. (2018) to produce the ZN/chitosan (CS)/nHAp biocomposite scaffold. The CS solution was mixed with the ZN solution and swirled. The nHAp was introduced to the CS-ZN solution and mixed for 24 hours. The sample was then subjected to 24-well culture plates and pre-frozen; it was freeze-dried at -80°C for 48 hours before being stored.

Polymer Coating

Polymer coatings are regularly used for integrating drug delivery transport and bone tissue engineering scaffolds (Philippart et al., 2015; Li et al., 2014; Yao et al., 2013). Polymer coatings are thin polymer layers applied on flat or uneven surfaces (Francis & Roberts, 2016). For example, in BTE, scaffolds are usually coated with polymer to enhance scaffold performance.

A study by Fereshteh et al. (2015) has produced highly porous bioactive glass (BG)-based scaffolds with a porosity of 90%, where it applied a dual PCL/zein coating on the scaffold surface to improve the mechanical properties of the scaffold. The scaffolds were made via the foam replication technique (Chen et al., 2006). In addition, the potential of zein as a covering material for such scaffolds was studied by Hum (2016). Dip coating was used to make zein-coated BG scaffolds, which were then characterized in microstructure, bioactivity, and mechanical strength. Another study conducted by Arango-ospina et al. (2021) used zein and manuka honey (MH) coating to BG scaffold. Table 3 shows the method used by researchers in the production of the scaffold.

Table 3

Summary of the current progress of zein (ZN) usage in BTE applications

Method	Material	Reference
Salt leaching	ZN	Gong et al. (2006)
Salt leaching	ZN, plasticizer	Wang et al. (2008)
Salt leaching	ZN, PCL	Wu et al. (2012)

Table 3 (Continue)

Method	Material	Reference
Salt leaching	ZN	Hum (2016)
Gas foaming	PCL, TZ, HA	Salerno et al. (2010a)
Gas foaming	PCL, TZ, HA	Salerno et al. (2010b)
Gas foaming-leaching	PCL, TZ, HA	Salerno et al. (2011)
Gas foaming	PCL, TZ, HA	Subuki et al. (2020)
3D printing	ZN, PCL, nMS	Ru et al. (2018)
Freeze drying	ZN, CS, nHAp	Shahbazarab et al. (2018)
Polymer coating	PCL, ZN	Fereshteh et al. (2015)
Polymer coating	ZN	Hum (2016)
Polymer coating	ZN, MH	Arango-ospina et al. (2021)

EFFECT OF ZEIN IN SCAFFOLD

Most scaffolds are analyzed on their biocompatibility, biodegradability, mechanical properties, and degradation rate. In addition, considerations such as scaffold porosity, mechanical properties, in vitro degradation study, cell proliferation, and osteogenic differentiation should be taken when designing a bone scaffold. This section reviews the effect of zein usage in bone scaffolds based on those considerations mentioned.

Scaffold Porosity

Macroporosity (pore size more than 100 μm) is frequently necessary for enhancing osteogenesis and angiogenesis (Murphy et al., 2010). In addition, interconnected macropores are required to increase the body's fluid circulation and the cell growth to the core of the implant (Reinwald et al., 2014). More notably, researchers discovered that microporosity (pore size less than 10 μm) plays a vital role in scaffold osteoinduction (Bohner et al., 2017; Polak et al., 2011).

Figure 2 depicts the scaffold's shape and microstructure, as Gong et al. (2006) reported. The pores on the surface of scaffolds produced were bigger and more uniform when manufactured with large porogen particle sizes. The porosities of zein scaffolds with various sodium chloride mass fractions and particle sizes range from 75.3 to 79.0%. Despite the pores' interior morphology and structure, all produced scaffolds had high porosity above 75%, which was deemed favorable to cell growth and survival (Thavornytikarn et al., 2014a).

Wu et al. (2012) discovered that the incorporation of zein in PCL scaffold could produce macropores larger than 300 μm , with many micropores ranging from 5 to 20 μm . Moreover, Figure 3A shows that the 3D model displays a highly porous structure and pore linkages. Meanwhile, the micro-CT scan in Figure 3B reveals a well-connected pore structure. For

cell migration and proliferation, consistent and suitable size of well-connected pores is required for 3D porous scaffolds (Chocholata et al., 2019; Murphy & O'Brien, 2010).

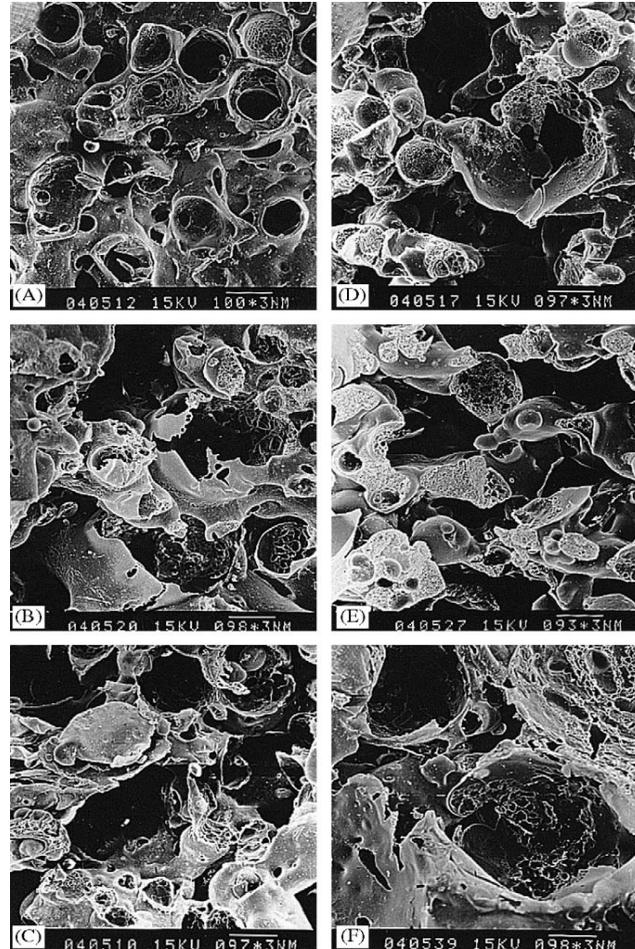


Figure 2. SEM illustrates the outer and inner morphology of scaffolds constructed with a porogen mass fraction of 1.4 and different particle sizes (Gong et al., 2006)

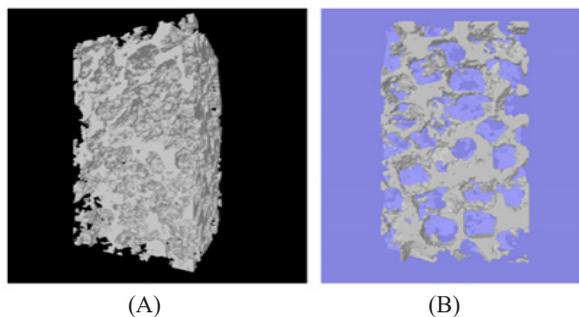


Figure 3. Micro-CT images of the zein/PCL-40 scaffold: (A) 3D model; and (B) semi-transparent overlay of a cross-section of the 3D model (Wu et al., 2012)

Besides that, Salerno et al. (2010) compared the structures of the PCL/TZ–HA with the PCL/TZ. As a result, the PCL/TZ–HA has smaller pores and a narrower pore size distribution, as demonstrated in Figure 4. Moreover, Salerno et al. (2010b) observed an increase in pore size as the foaming temperature (T_f) went from 80 to 100°C.

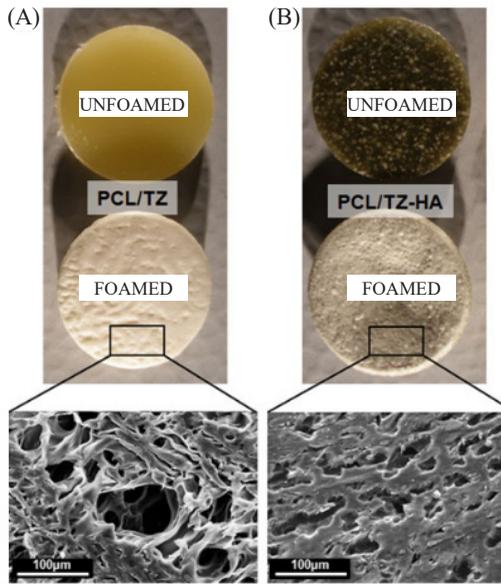


Figure 4. Optical and SEM microscope images of (a) PCL/TZ; and (b) PCL/TZ–HA before and after undergoing the gas foaming step (Di Maio et al., 2010)

In addition, Subuki et al. (2020) discovered that biocomposite at 20 wt% incorporations with TZ exhibits strong visible interconnection of the porous structure, as seen in Figure 5(a) when contrasted to Figure 5(b), in which the effect of the foaming process at 10% displays low interconnection.

Furthermore, Shahbazarab et al. (2018) performed SEM analysis, revealing that the ZN comprises particles with a diameter of 100–500 nm and a high amount of interconnected pores in pure ZN. Besides that, according to SEM images, the pore size of the PCL/zein coating experiment conducted by Fereshteh et al. (2015) was in the range of 200–450 nm, which is ideal for use in bone tissue engineering.

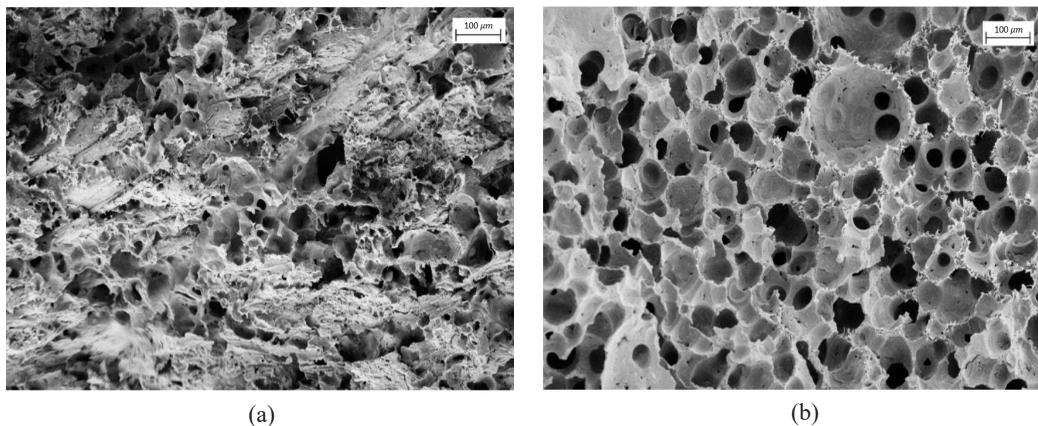


Figure 5. SEM Micrograph Images of the Cross-Section of: (a) 20 wt% TZ; and (b) 10 wt% TZ (Subuki et al., 2020)

Mechanical Properties

The scaffold must provide sufficient mechanical strength to enable new tissue replacement (Prasadh & Wong, 2018). The biostability of the scaffold is improved by the material interface's strength, flexibility, and absorption. The mechanical properties of the scaffold should be closely related to the mechanical properties of the surrounding bone. Wang et al. (2008) reported that Young's modulus and strength increased with decreasing mass percentage of the porogen. As a result, the compressive modulus and compressive strength of the porous zein scaffold were in the trabecular bone range (compressive modulus: 10–2000 MPa; compressive strength: 2–180 MPa).

On the other hand, with increasing zein content, the compressive stress value of the composite scaffolds decreased, as Wu et al. (2012) and Shahbazarab et al. (2018) reported in Figures 6 and 7, respectively. In addition, the tensile properties analysis conducted by Salerno et al. (2010) shows that the yielding stress, the stress at break, and the elongation at the break of the scaffold decrease with the incorporation of ZN. It was observed that the compressive strength of the tested scaffold decreased with increasing wt% of ZN in the scaffold (Ru et al., 2018), which could be due to the presence of zein that caused a reduced degree of crystallinity.

Although the creation of porous scaffolds made of natural materials for bone regeneration has yielded promising results, they all have one weakness: lack of mechanical strength due to their porosity. It means that they are insufficiently strong to meet the demands of load-bearing applications (Wang et al., 2008). Furthermore, larger porosity was associated with decreased Young's modulus and compressive strength, consistent with previous research by Zhao et al. (2018) and Sabudin et al. (2019) that revealed an inverse relationship between porosity and mechanical characteristics.

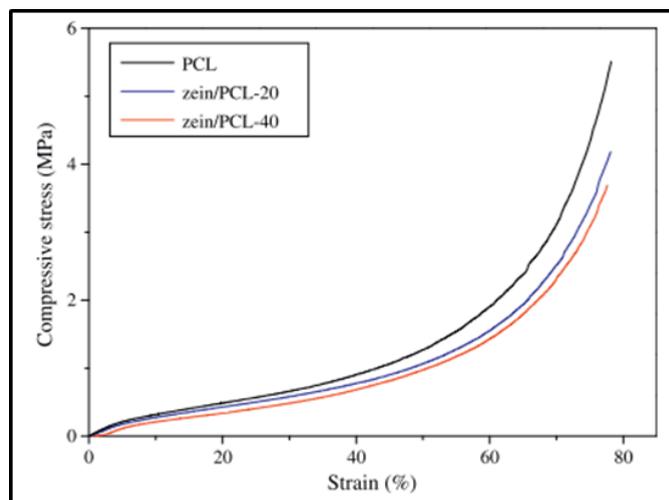


Figure 6. Compressive stress-strain curves of the different scaffolds (Wu et al., 2012)

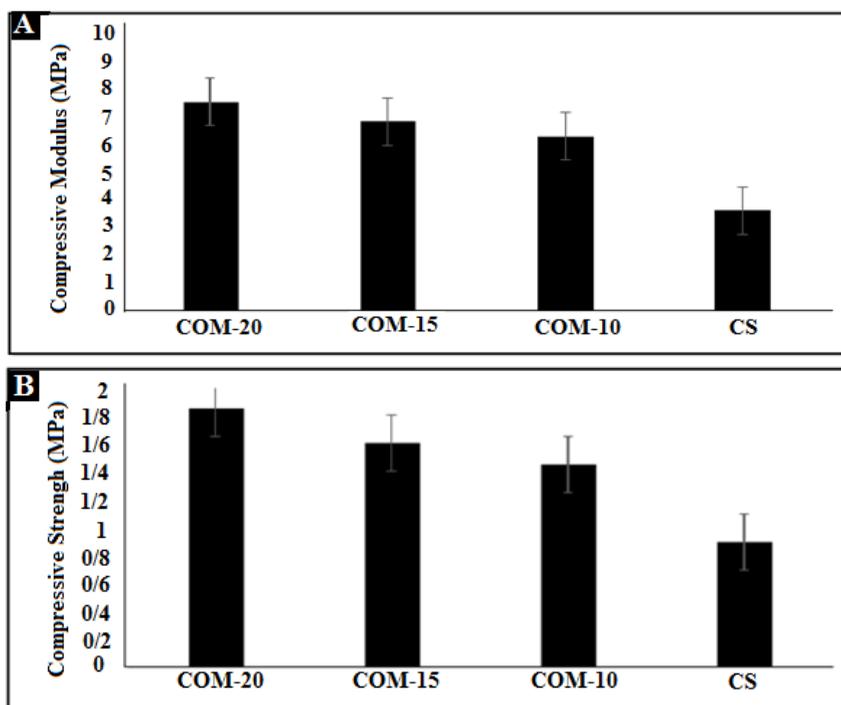


Figure 7. Mechanical properties of tested composite scaffolds (Shahbazarab et al., 2018)

***In Vitro* Degradation Study**

The scaffold must degrade in a timely manner after implantation to facilitate adequate tissue regeneration (Bitar & Zakhem, 2014). Gong et al. (2006) reported that during the 14-day *in vitro* incubation period, up to 36% and 89% rates of degradation were recorded when incubated with collagenase and pepsin, respectively. This result indicated that a greater porogen mass proportion results in increased porosity, leading to a faster scaffold degradation rate. Wu et al. (2012) found that the weight loss of the zein/PCL composite scaffolds was significant, and it increased as the amount of zein in the composite increased, as shown in Figure 8. According to Salerno et al. (2010), during the rate of degradation analysis, PCL/TZ and PCL/TZ–HA showed considerable variations, as shown in Figure 9. A study by Fereshteh et al. (2015) also shows that during the 28-day immersion in PBS, the PCL/zein coating degraded noticeably. The quicker degradation rates found for the PCL/TZ and PCL/TZ–HA may be due to the capacity of TZ to swell and degrade in water.

Moreover, Ru et al. (2018) and Shahbazarab et al. (2018) reported that scaffolds with higher wt% of ZN lose significantly more weight than other scaffolds. The analysis of the degradation rate obtained shows that the scaffold degradation rate increases with the increasing mass fraction of zein. Furthermore, in water-based systems, zein-coated scaffolds could exhibit a high level of stability. The steady degradation posed by zein is advantageous to the goal of tissue engineering strategy for natural tissue replacement of

the temporary matrix (Gomes et al., 2008). From the analysis obtained, it can be said that the rate of degradation of composite scaffolds can be controlled to match the rate of tissue regeneration by manipulating the amount of zein in the composite.

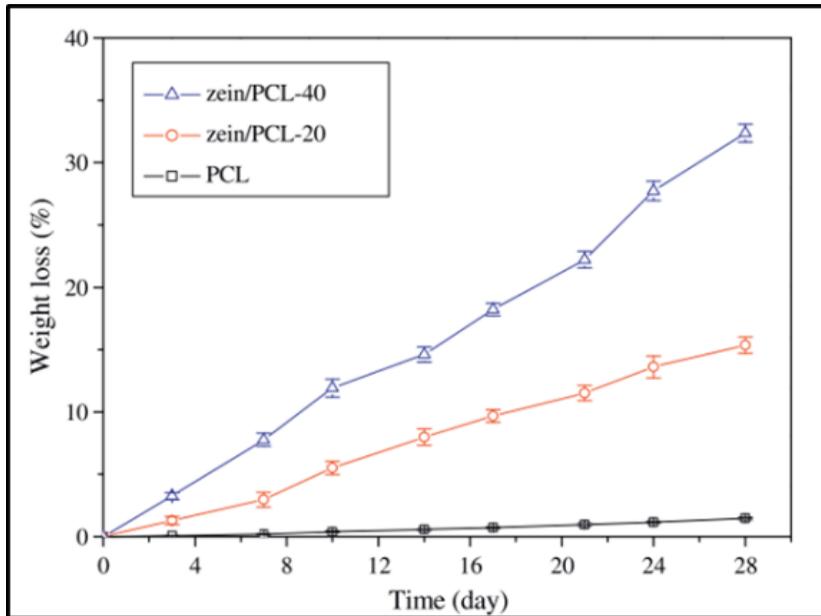


Figure 8. Weight loss of scaffold analysis graph (Wu et al., 2012)

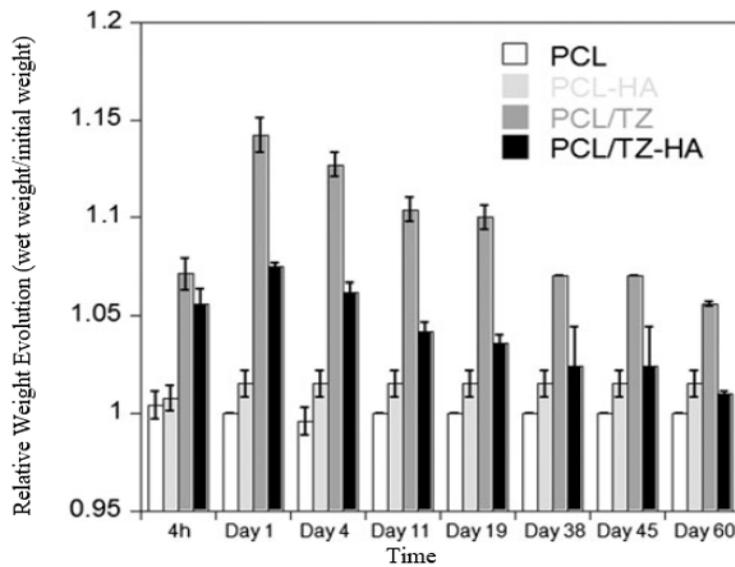


Figure 9. Effect of biomaterial composition on relative weight loss during degradation analysis (Salerno et al., 2010)

Cell Proliferation and Osteogenic Differentiation

For materials used in tissue engineering, good biocompatibility is required in addition to adequate mechanical capabilities, outstanding pore characteristics, and degradation behavior. Gong et al. (2006) show that the dexamethasone group had much more activity than the control group, showing that MSCs successfully differentiated themselves towards osteoblasts and proliferated on the porous zein scaffolds. The findings of SEM and CLSM investigations in Figure 10 verified the ability of multi-phase scaffolds to allow for both cell types' attachment and colonization, as Salerno et al. (2010b). The cells adhered to the pores of the scaffolds and colonized them, expanding and forming bridges between opposite pore walls, as seen in Figure 10a. Meanwhile, the CLSM results shown in Figure 10b demonstrate that the hMSCs were stable on Day 1 after seeding, adhering to and populating the scaffold surface, and following the topology of the pore walls.

According to Ru et al. (2018), the OD values (cell proliferation) for 20 wt% and 10 wt% of ZN incorporated in scaffolds were higher than no ZN at all, indicating that ZN scaffolds increased cell proliferation. The number of newly generated bone tissues (NBs) in the tested scaffolds grew. Furthermore, as the ZN content in the scaffolds grew, the amount of NBs development in the scaffolds increased, as seen in Figure 11 according to Ru et al. (2018). Shahbazarab et al. (2018) also showed that cells were adhered and spread out through SEM images of cells cultured on scaffolds. It was also discovered that adding CS and ZN to porous scaffolds improved MG-63 cell adhesion, growth, and proliferation. Therefore, it is deemed that the topography of the scaffold surface was influenced by ZN, which could have had a significant impact on cell adhesion, proliferation, and osteogenic differentiation. One possible reason for this impact is the higher hydrophilicity and degradation rate of TZ, which may increase serum protein absorption and HA particle exposure, ultimately increasing rMSC osteogenic development. All these findings show that ZN had improved osteogenic properties. Table 4 shows the summary of the outcome of the findings from the previous paper.

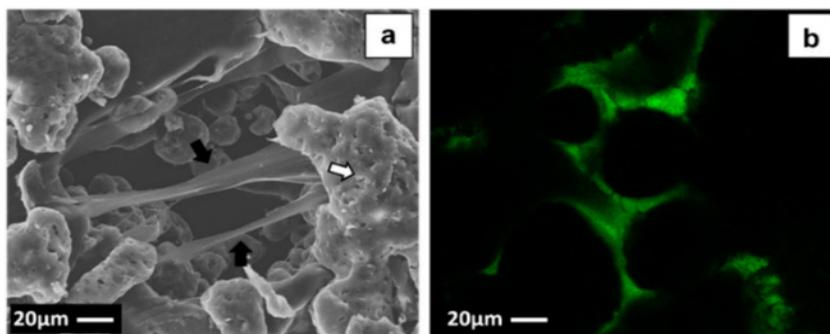


Figure 10. a) SEM image of MG63 seeding; b) CLMS image of hMSCs seeding Salerno et al. (2010b)

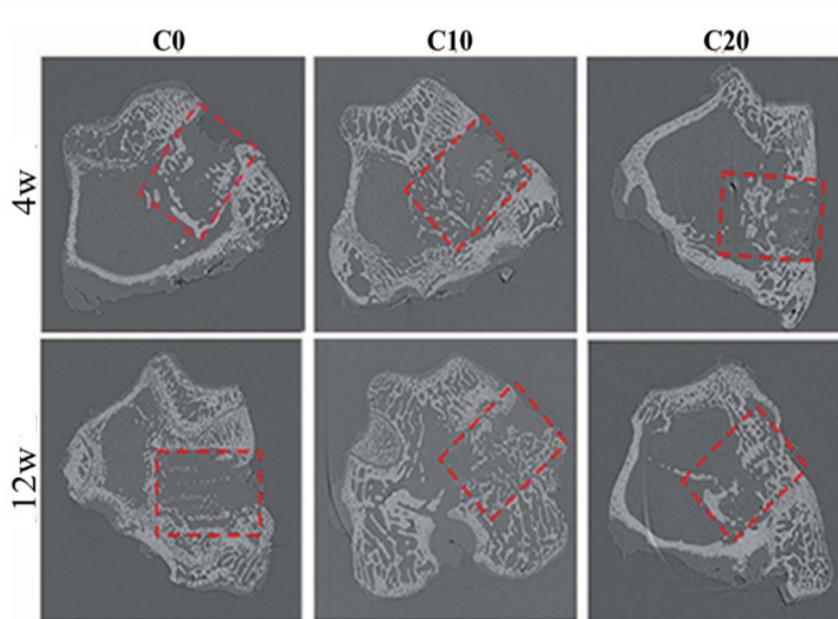


Figure 11. The 2D graphics of bone-scaffolds from micro-CT after tested scaffolds implanted to femoral defects of rabbits at 4 and 12 weeks (Ru et al., 2018)

Table 4

Summary of findings outcome from the previous paper on the usage of ZN and various materials used to combine with ZN, along with technique employed and its advantages/disadvantages

Method	Material	Advantages	Disadvantages	Reference
Salt leaching	ZN	Reliable scaffolds with an interconnective, open pore structure and pore size up to 300 mm were produced using the salt leaching technique. Mesenchymal stem cells (MSCs) can successfully adhere, develop, and proliferate on the zein scaffold.	The mechanical properties of the scaffold decrease with the increased mass fraction of zein.	Gong et al. (2006)

Table 4 (Continue)

Method	Material	Advantages	Disadvantages	Reference
Salt leaching	ZN, plasticizer	Plasticizer concentrations of 20% for (oleic acid) OA and 15% for stearic acid (SA) were sufficient to improve the flexibility and tensile properties of porous zein scaffolds.	Even when the mechanical properties of the OA plasticized zein scaffold meet the requirements for a hard tissue engineering replacement, it is not strong enough.	Wang et al. (2008)
Salt leaching	ZN, HA	<p>The pore interconnectivity of the HA-coated scaffolds was comparable to that of the zein scaffolds. The HA-coated zein scaffolds exhibited the typical structure of HA crystals, indicating the influence of immersion and apatite production.</p> <p>The compressive modulus of HA-coated zein scaffolds was acceptable for bone tissue engineering. hBMSC development was better on both the HA-coated and uncoated zein scaffolds.</p>	N/A	Qu et al. (2008)

Table 4 (Continue)

Method	Material	Advantages	Disadvantages	Reference
Salt leaching	ZN, PCL	The incorporation of zein showed a high hydrophilic nature that could help cell adhesion and proliferation. The scaffold's degradation rate was found to be higher in the zein/PCL scaffold.	Increasing zein content will cause the scaffold to have low mechanical strength.	Wu et al. (2012)
Salt leaching	ZN	The compressive strength of pure zein is found to be within the ranges required by natural cancellous bone.	Because of its insolubility, pure zein did not lose weight after 14 days in PBS, implying that it has high stability in water-based solutions. There is no hydroxyapatite on the surface of zein; hence no bioactive behavior can be discovered.	Hum (2016)
Gas foaming	PCL, TZ, HA	The increased relaxation temperature of the PCL/TZ and PCL/TZ-HA, which is close to human body temperature, may allow a cyclic mechanical solicitation to be dissipated. TZ improved the wettability of PCL. The topography of the scaffold surface was also changed by TZ and HA, which could have had a significant impact on cell adhesion, proliferation, and osteogenic differentiation.	N/A	Salerno et al. (2010)

Table 4 (Continue)

Method	Material	Advantages	Disadvantages	Reference
Gas foaming	PCL, TZ, HA	<p>Satisfactory correlation with those found in the literature on the capability of PCL and TZ to help adhesion and proliferation of the cell.</p> <p>HA content increases, and the mechanical properties of the scaffold also increase.</p> <p>PCL-TZ has the fastest degradation rate.</p> <p>Scaffold managed, allowing the cell for adhesion and colonization in the tested culture cell.</p> <p>The findings are consistent with previous research that has combined natural polymers such as chitosan and starch with PCL to create scaffolds with improved hydrophilicity and a faster breakdown rate.</p>	N/A	Rosa et al. (2005); She et al. (2007); Salerno et al. (2010b)
Gas foaming-leaching	PCL, TZ, HA	The use of PCL and TZ together allowed for precise control of scaffold degradation.	N/A	Salerno et al. (2011)
Gas foaming	PCL, TZ, HA	<p>At temperatures up to 250°C, all the composite samples showed a stable thermal property.</p> <p>The composite that contains 20 wt% of TZ exhibited better morphology of pore structure in the sample composite than other compositions of TZ.</p>	N/A	Subuki et al. (2020)

Table 4 (Continue)

Method	Material	Advantages	Disadvantages	Reference
3D printing	ZN, PCL, nMS	<p>The combination of those advantages of both biodegradable polymers and bioactive materials would result in enhanced bio-performance of the composite scaffolds.</p> <p>The apatite-mineralization capability of the scaffolds was enhanced with an increase in ZN content.</p> <p>The rate of degradability of the scaffolds has been improved with the addition of ZN.</p> <p>Tested cell culture responses, such as replication and division to the scaffolds, have been promoted with increasing ZN.</p>	N/A	Ru et al. (2018)
Freeze drying	ZN, CS, nHAp	Including CS and ZN (with a reduction in the amount of nHAp) improved the adhesion, growth, and proliferation of MG-63 cells on porous scaffolds.	N/A	Shahbazarab et al. (2018)

CONCLUSION

With excellent flexibility and compressibility, zein can create strong, glossy, hydrophobic, greaseproof coatings prone to microbial harm. Due to its advantages in the specialty food, pharmaceutical, and biodegradable plastic industries, zein has been of interest for commercialization. As a result, zein protein is widely used in many sectors due to its usefulness as an industrial and specialty polymer. In addition, the current research on zein has been mostly found in BTE applications, as zein displays great ability for use in the scaffold.

From the current research on the zein addition effect in BTE, it can be concluded that zein shows suitable interconnectivity as a material for scaffolding, which aids in achieving optimum vascular ingrowth. The scaffold degradation rate also increases with the increasing mass fraction of zein. In addition, the tested cell culture can successfully adhere, develop, and proliferate on the zein scaffold. Furthermore, the incorporation of zein shows high hydrophilic nature that could assist in the adhesion and proliferation of the cell. This research also demonstrates that the 3D printing method is preferred for scaffold fabrication, as it could overcome all issues regarding the use of the conventional method. Moreover, scaffold produced by the 3D printing method shows better performance for bone tissue engineering applications.

However, this research shows that the increased zein content will cause the scaffold to have low mechanical strength. Therefore, zein has to be combined with other components such as PCL, HA, and n-MS to enhance the mechanical strength of the composite scaffolds.

ACKNOWLEDGMENT

The authors would like to thank for the financial support by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2019/TK02/UITM/02/19), Universiti Teknologi MARA (UiTM), Malaysia.

REFERENCES

- Alizadeh, M., Abbasi, F., Khoshfetrat, A. B., & Ghaleh, H. (2013). Microstructure and characteristic properties of gelatin/chitosan scaffold prepared by a combined freeze-drying/leaching method. *Materials Science & Engineering: C*, 33(7), 3958-3967. <https://doi.org/10.1016/j.msec.2013.05.039>
- An, J., Teoh, J. E. M., Suntornnond, R., & Chua, C. K. (2015). Design and 3D printing of scaffolds and tissues. *Engineering*, 1(2), 261-268. <https://doi.org/10.15302/J-ENG-2015061>
- Anderson, T. J., & Lamsa, B. P. (2011). Zein extraction from corn, corn products, and coproducts and modifications for various applications: A review. *Cereal Chemistry*, 88(2), 159-173. <https://doi.org/10.1094/CCHEM-06-10-0091>
- Arango-ospina, M., Lasch, K., Weidinger, J., & Boccaccini, A. R. (2021). Manuka honey and zein coatings impart bioactive glass bone tissue scaffolds antibacterial properties and superior mechanical properties. *Front Mater*, 7, Article 610889. <https://doi.org/10.3389/fmats.2020.610889>
- Bajaj, P., Schweller, R. M., Khademhosseini, A., West, J. L., & Bashir, R. (2014). 3D biofabrication strategies for tissue engineering and regenerative medicine. *Annual Review of Biomedical Engineering*, 16, 247-276. <https://doi.org/10.1146/annurev-bioeng-071813-105155>
- Berardi, A., Bisharat, L., Alkhatib, H. S., & Cespi, M. (2018). Zein as a pharmaceutical excipient in oral solid dosage forms: State of the art and future perspectives. *AAPS PharmSciTech*, 19, 2009-2022. <https://doi.org/10.1208/s12249-018-1035-y>

- Bharadwaz, A., & Jayasuriya, A. C. (2020). Recent trends in the application of widely used natural and synthetic polymer nanocomposites in bone tissue regeneration. *Materials Science and Engineering: C, 110*, Article 110698. <https://doi.org/10.1016/j.msec.2020.110698>
- Bitar, K. N., & Zakhem, E. (2014). Design strategies of biodegradable scaffolds for tissue regeneration. *Biomedical Engineering and Computational Biology, 6*, 13-20. <https://doi.org/10.4137/BECB.S10961>
- Bohner, M., Baroud, G., Bernstein, A., Do, N., Galea, L., Hesse, B., Heuberger, R., Meille, S., Michel, P., von Rechenberg, B., Sague, J., & Seeherman, H. (2017). Characterization and distribution of mechanically competent mineralized tissue in micropores of β -tricalcium phosphate bone substitutes. *Materials Today, 20*(3), 106-115. <https://doi.org/10.1016/j.mattod.2017.02.002>
- Bose, S., Vahabzadeh, S., & Bandyopadhyay, A. (2013). Bone tissue engineering using 3D printing. *Materials Today, 16*(12), 496-504. <https://doi.org/10.1016/j.mattod.2013.11.017>
- Chen, H., Wang, J., Cheng, Y., Wang, C., Liu, H., Bian, H., Pan, Y., Sun, J., & Han, W. (2019). Application of protein-based films and coatings for food packaging: A review. *Polymers, 11*(12), Article 2039. <https://doi.org/10.3390/polym11122039>
- Chen, Q. Z., Thompson, I. D., & Boccaccini, A. R. (2006). 45S5 bioglass-derived glass-ceramic scaffolds for bone tissue engineering. *Biomaterials, 27*(11), 2414-2425. <https://doi.org/10.1016/j.biomaterials.2005.11.025>
- Chocholata, P., Kulda, V., & Babuska, V. (2019). Fabrication of scaffolds for bone-tissue regeneration. *Materials, 12*(4), Article 568. <https://doi.org/10.3390/ma12040568>
- Demir, M., Ramos-Rivera, L., Silva, R., Nazhat, S. N., & Boccaccini, A. R. (2017). Zein-based composites in biomedical applications. *Journal of Biomedical Materials Research Part A, 105*(6), 1656-1665 <https://doi.org/10.1002/jbm.a.36040>
- Esen, A. (1987). A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (*Zea mays* L.). *Journal of Cereal Science, 5*(2), 117-128. [https://doi.org/10.1016/S0733-5210\(87\)80015-2](https://doi.org/10.1016/S0733-5210(87)80015-2)
- Fereshteh, Z., Nooeaid, P., Fathi, M., Bagri, A., & Boccaccini, A. R. (2015). The effect of coating type on mechanical properties and controlled drug release of PCL/zein coated 45S5 bioactive glass scaffolds for bone tissue engineering. *Materials Science and Engineering: C, 54*, 50-60. <https://doi.org/10.1016/j.msec.2015.05.011>
- Francesca, D., Emanuela, J., Monica, S., & Manuela, T. R. (2020). Natural and synthetic polymers. In S. Chandra & Y. Ohama (Eds.), *Polymers in concrete*, (pp. 5-25). CRC Press. <https://doi.org/10.1201/9781003068211-2>
- Francis, L. F., & Roberts, C. C. (2016). Dispersion and solution processes. In *Material Processing* (pp. 415-512). Academia Press. <https://doi.org/10.1016/B978-0-12-385132-1.00006-9>
- Ghassemi, T., Shahroodi, A., Ebrahimzadeh, M. H., Mousavian, A., Movaffagh, J., & Moradi, A. (2018). Current concepts in scaffolding for bone tissue engineering. *Archives of Bone and Joint Surgery, 6*(2), 90-99. <https://doi.org/10.22038/abjs.2018.26340.1713>
- Gomes, M. E., Azevedo, H. S., Moreira, A. R., & Ell, V. (2008). Starch-poly (lactic acid) fibre-mesh scaffolds for bone tissue engineering applications: Structure, mechanical properties and degradation behaviour. *Journal of Tissue Engineering and Regenerative Medicine, 2*(5), 243-252.

- Gong, S., Wang, H., Sun, Q., Xue, S. T., & Wang, J. Y. (2006). Mechanical properties and *in vitro* biocompatibility of porous zein scaffolds. *Biomaterials*, 27(20), 3793-3799. <https://doi.org/10.1016/j.biomaterials.2006.02.019>
- Gungor-Ozkerim, P. S., Inci, I., Zhang, Y. S., Khademhosseini, A., & Dokmeci, M. R. (2018). Bioinks for 3D bioprinting: An overview. *Biomaterials Science*, 6(5), 915-946. <https://doi.org/10.1039/c7bm00765e>
- Hospodiuk, M., Dey, M., Sosnoski, D., & Ibrahim, T. (2016). The bioink: A comprehensive review on bioprintable materials. *Biotechnology Advances*, 35(2), 217-239. <https://doi.org/10.1016/j.biotechadv.2016.12.006>
- Hum, J., Naseri, S., & Boccaccini, A. R. (2018). Bioactive glass combined with zein as composite material for the application in bone tissue engineering. *Biomedical Glasses*, 4(1), 72-81. <https://doi.org/10.1515/bglass-2018-0007>
- Hum, J. K., (2016). *Bioactive glass combined with natural derived proteins as composite materials for the application in bone tissue engineering* (Doctoral dissertation). University of Erlangen-Nuremberg, Germany. <https://d-nb.info/1103801953/34>
- Ji, C., Annabi, N., Khademhosseini, A., & Dehghani, F. (2011). Fabrication of porous chitosan scaffolds for soft tissue engineering using dense gas CO₂. *Acta Biomaterialia*, 7(4), 1653-1664. <https://doi.org/10.1016/j.actbio.2010.11.043>
- Jing, L., Wang, X., Liu, H., Lu, Y., Bian, J., Sun, J., & Huang, D. (2018). Zein increases the cytoaffinity and biodegradability of scaffolds 3D-printed with zein and poly(ϵ -caprolactone) composite ink. *ACS Applied Materials and Interfaces*, 10(22), 18551-18559. <https://doi.org/10.1021/acsami.8b04344>
- Labib, G. (2018). Overview on zein protein: A promising pharmaceutical excipient in drug delivery systems and tissue engineering. *Expert Opinion on Drug Delivery*, 15(1), 65-75. <https://doi.org/10.1080/17425247.2017.1349752>
- Lawton, J. W. (2002). Zein: A history of processing and use. *Cereal Chemistry*, 79(1), 1-18. <https://doi.org/10.1094/CCHEM.2002.79.1.1>
- Li, W., Nooeaid, P., Roether, J. A., Schubert, D. W., & Boccaccini, A. R. (2014). Preparation and characterization of vancomycin releasing PHBV coated 45S5 Bioglass[®]-based glass-ceramic scaffolds for bone tissue engineering. *Journal of the European Ceramic Society*, 34(2), 505-514. <https://doi.org/10.1016/j.jeurceramsoc.2013.08.032>
- Maji, K., & Dasgupta, S. (2017). Effect of β -tricalcium phosphate nanoparticles additions on the properties of gelatin-chitosan scaffolds. *Bioceramics Development and Applications*, 7(2), Article 10000103. <https://doi.org/10.4172/2090-5025.1000103>
- Mikos, A. G., & Temenoff, J. S. (2000). Formation of highly porous biodegradable scaffolds for tissue engineering. *Electronic Journal of Biotechnology*, 3(2), 1995-2000.
- Murphy, C. M., & O'Brien, J. F. (2010). Understanding the effect of mean pore size on cell activity in collagen-glycosaminoglycan scaffolds. *Cell Adhesion & Migration*, 4(3), 377-381. <https://doi.org/10.4161/cam.4.3.11747>
- Murphy, C. M., Haugh, M. G., & Brien, F. J. O. (2010). The effect of mean pore size on cell attachment, proliferation and migration in collagen-glycosaminoglycan scaffolds for bone tissue engineering. *Biomaterials*, 31(3), 461-466. <https://doi.org/10.1016/j.biomaterials.2009.09.063>

- Nam, Y. S., Yoon, J. J., & Park, T. G. (2000). A novel fabrication method of macroporous biodegradable polymer scaffolds using gas foaming salt as a porogen additive. *Journal of Biomedical Materials Research*, 53(1), 1-7. [https://doi.org/10.1002/\(SICI\)1097-4636\(2000\)53:1<1::AID-JBM1>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-4636(2000)53:1<1::AID-JBM1>3.0.CO;2-R)
- Patel, M., & Fisher, J. P. (2008). *Biodegradable materials for tissue engineering*. McGraw-Hill Global Education Holdings, LLC. <https://doi.org/10.1036/1097-8542.YB080510>
- Philippart, A., Boccaccini, A. R., Fleck, C., Schubert, D. W., & Roether, J. A. (2015). Toughening and functionalization of bioactive ceramic and glass bone scaffolds by biopolymer coatings and infiltration: A review of the last 5 years. *Expert Review of Medical Devices*, 12(1), 93-111. <https://doi.org/10.1586/17434440.2015.958075>
- Polak, S. J., Lan, S. K., Wheeler, M. B., Maki, A. J., Clark, S. G., & Wagoner, A. J. (2011). Analysis of the roles of microporosity and BMP-2 on multiple measures of bone regeneration and healing in calcium phosphate scaffolds. *Acta Biomaterialia*, 7(4), 1760-1771. <https://doi.org/10.1016/j.actbio.2010.12.030>
- Prasadh, S., & Wong, R. C. W. (2018). Unraveling the mechanical strength of biomaterials used as a bone scaffold in oral and maxillofacial defects. *Oral Science International*, 15(2), 48-55. [https://doi.org/10.1016/S1348-8643\(18\)30005-3](https://doi.org/10.1016/S1348-8643(18)30005-3)
- Qu, H., Fu, H., Han, Z., & Sun, Y. (2019). Biomaterials for bone tissue engineering scaffolds: A review. *RSC Advances*, 9, 26252-26262. <https://doi.org/10.1039/c9ra05214c>
- Qu, Z. H., Wang, H. J., Tang, T. T., Zhang, X. L., Wang, J. Y., & Dai, K. R. (2008). Evaluation of the zein/inorganics composite on biocompatibility and osteoblastic differentiation. *Acta Biomaterialia*, 4(5), 1360-1368. <https://doi.org/10.1016/j.actbio.2008.03.006>
- Reddy, N., & Yang, Y. (2011). Potential of plant proteins for medical applications. *Trends in Biotechnology*, 29(10), 490-498. <https://doi.org/10.1016/j.tibtech.2011.05.003>
- Reddy, N., & Yang, Y. (2013). Thermoplastic films from plant proteins. *Journal of Applied Polymer Science*, 130(2), 729-738. <https://doi.org/10.1002/app.39481>
- Reinwald, Y., Johal, R. K., Ghaemmaghami, A. M., Rose, F. R. A. J., Howdle, S. M., & Shakesheff, K. M. (2014). Interconnectivity and permeability of supercritical fluid-foamed scaffolds and the effect of their structural properties on cell distribution. *Polymer*, 55(1), 435-444. <https://doi.org/10.1016/j.polymer.2013.09.041>
- Rezwan, K., Chen, Q. Z., Blaker, J. J., & Roberto, A. (2006). Biodegradable and bioactive porous polymer/inorganic composite scaffolds for bone tissue engineering. *Biomaterials*, 27(18), 3413-3431. <https://doi.org/10.1016/j.biomaterials.2006.01.039>
- Rosa, D. S., Lopes, D. R., & Calil, M. R. (2005). Thermal properties and enzymatic degradation of blends of poly (ϵ -caprolactone) with starches. *Polymer Testing*, 24(6), 756-761. <https://doi.org/10.1016/j.polymertesting.2005.03.014>
- Roseti, L., Parisi, V., Petretta, M., Cavallo, C., Desando, G., Bartolotti, I., & Grigolo, B. (2017). Scaffolds for bone tissue engineering: State of the art and new perspectives. *Materials Science and Engineering: C*, 78, 1246-1262. <https://doi.org/10.1016/j.msec.2017.05.017>
- Ru, J., Wei, Q., Yang, L., Qin, J., Tang, L., Wei, J., Guo, L., & Niu, Y. (2018). Zein regulating apatite mineralization, degradability, *in vitro* cells responses and *in vivo* osteogenesis of 3D-printed scaffold of n-MS/ZN/PCL ternary composite. *RSC Advances*, 8, 18745-18756. <https://doi.org/10.1039/c8ra02595a>

- Sabudin, S., Marzuke, M. A., & Hussin, Z. (2019). Effect of mechanical properties on porous calcium phosphate scaffold. *Materials Today: Proceedings*, 16(Part 4), 1680-1685. <https://doi.org/10.1016/j.matpr.2019.06.036>
- Sah, M. K., & Pramanik, K. (2011). Computational approaches in tissue engineering. *International Journal of Computer Applications*, 27(4), 0975-8887.
- Salerno, A., Di Maio, E., Iannace, S., & Netti, P. A. (2011). Tuning the microstructure and biodegradation of three-phase scaffolds for bone regeneration made of PCL, Zein, and HA. *Journal of Cellular Plastics*, 47(3), 245-260. <https://doi.org/10.1177/0021955X11404832>
- Salerno, A., Oliviero, M., Di Maio, E., Netti, P. A., Rofani, C., Colosimo, A., Guida, V., Dallapiccola, B., Palma, P., Procaccini, E., Berardi, A. C., Velardi, F., Teti, A., & Iannace, S. (2010a). Design of novel three-phase PCL/TZ-HA biomaterials for use in bone regeneration applications. *Journal of Materials Science: Materials in Medicine*, 21, 2569-2581. <https://doi.org/10.1007/s10856-010-4119-0>
- Salerno, A., Zeppetelli, S., Di Maio, E., Iannace, S., & Netti, P. A. (2010b). Novel 3D porous multi-phase composite scaffolds based on PCL, thermoplastic zein and ha prepared via supercritical CO₂ foaming for bone regeneration. *Composites Science and Technology*, 70(13), 1838-1846. <https://doi.org/10.1016/j.compscitech.2010.06.014>
- Salerno, A., Oliviero, M., Maio, E. Di, & Iannace, S. (2006). Thermoplastic foams from zein and gelatin. *International Polymer Processing*, 22(5), 480-488. <https://doi.org/10.3139/217.2065>
- Shahbazarab, Z., Teimouri, A., Chermahini, A. N., & Azadi, M. (2018). Fabrication and characterization of nanobiocomposite scaffold of zein/chitosan/nanohydroxyapatite prepared by freeze-drying method for bone tissue engineering. *International Journal of Biological Macromolecules*, 108, 1017-1027. <https://doi.org/10.1016/j.ijbiomac.2017.11.017>
- Shamaz, M., & Halima, S. B. (2015). Bone Tissue Engineering and Bony Scaffolds. *International Journal of Dentistry & Oral Health*, 1(1), 15-20. <https://doi.org/10.25141/2471-657X-2015-1.0001>
- She, H., Xiao, E. X., & Liu, E. R. (2007). Preparation and characterization of polycaprolactone-chitosan composites for tissue engineering applications. *Journal of Material Science*, 42, 8113-8119. <https://doi.org/10.1007/s10853-007-1706-7>
- Shukla, R., & Cheryan, M. (2001). Zein: The industrial protein from corn. *Industrial Crops and Products*, 13(3), 171-192. [https://doi.org/10.1016/S0926-6690\(00\)00064-9](https://doi.org/10.1016/S0926-6690(00)00064-9)
- Subuki, I., Adnan, N., & Sharudin, R. W. (2018). Biodegradable scaffold of natural polymer and hydroxyapatite for bone tissue engineering: A short review. *AIP Conference Proceedings*, 2031(1), 1-5. <https://doi.org/10.1063/1.5066975>
- Subuki, I., Akhbar, S., & Nor Wahid, F. K. (2020). Influence of thermoplastic PEG, GLY and Zein in PCL/TZ and HAp bio composite via solid state supercritical CO₂ foaming. *Scientific Research Journal*, 17(2), 177-190. <https://doi.org/10.24191/srj.v17i2.9534>
- Tariverdian, T., Sefat, F., Gelinsky, M., & Mozafari, M. (2019). Scaffold for bone tissue engineering. In M. Mozafari, F. Sefat & A. Atala (Eds.), *Handbook of Tissue Engineering Scaffolds: Volume One* (pp. 189-209). Elsevier Ltd. <https://doi.org/10.1016/B978-0-08-102563-5.00010-1>

- Thavornnyutikarn, B., Chantarapanich, N., & Chen, Q. (2014a). Bone tissue engineering scaffolding: Computer-aided scaffolding techniques. *Progress in Biomaterials*, 3, 61-102. <https://doi.org/10.1007/s40204-014-0026-7>
- Thavornnyutikarn, B., Chantarapanich, N., Sitthiseripratip, K., Thouas, G. A., & Chen, Q. (2014b). Bone tissue engineering scaffolding: computer-aided scaffolding techniques. In *Progress in Biomaterials* (pp. 61-102). Springer. <https://doi.org/10.1007/s40204-014-0026-7>
- Tong, W., Suihuai, Y., Dengkai, C., & Yanen, W. (2017). Bionic design, materials and performance of bone tissue scaffolds. *Materials*, 10(10), Article 1187. <https://doi.org/10.3390/ma10101187>
- Torabi, K., Farjood, E., & Hamedani, S. (2015). Rapid prototyping technologies and their applications in prosthodontics, a Review of Literature. *Journal of Dentistry*, 16(1), 1-9.
- Tortorella, S., Maturi, M., Vetri Buratti, V., Vozzolo, G., Locatelli, E., Sambri, L., & Franchini, M. C. (2021). Zein as a versatile biopolymer: Different shapes for different biomedical applications. *RSC Advances*, 11, 39004-39026. <https://doi.org/10.1039/d1ra07424e>
- Wahid, F., Khan, T., Hussain, Z., & Ullah, H. (2018). Nanocomposite scaffolds for tissue engineering; properties, preparation and applications. In A. M. A. Inamudin & A. Mohammad (Eds.), *Applications of Nanocomposite Materials in Drug Delivery* (pp. 701-735). Elsevier. <https://doi.org/10.1016/B978-0-12-813741-3.00031-5>
- Wang, C., Huang, W., Zhou, Y., He, L., He, Z., Chen, Z., He, X., Tian, S., Liao, J., Lu, B., Wei, Y., & Wang, M. (2020). 3D printing of bone tissue engineering scaffolds. *Bioactive Materials*, 5(1), 82-91. <https://doi.org/10.1016/j.bioactmat.2020.01.004>
- Wang, H. J., Gong, S. J., & Wang, J. Y. (2008). Mechanical improvement of zein protein as scaffold for bone tissue engineering. *Materials Science and Technology*, 24(9), 1045-1052. <https://doi.org/10.1179/174328408X341735>
- Wu, F., Wei, J., Liu, C., O'Neill, B., & Ngothai, Y. (2012). Fabrication and properties of porous scaffold of zein/PCL biocomposite for bone tissue engineering. *Composites Part B: Engineering*, 43(5), 2192-2197. <https://doi.org/10.1016/j.compositesb.2012.02.040>
- Xiaohao, L., & Peter, X. M. (2004). Polymeric scaffolds for bone tissue engineering. *Annals of Biomedical Engineering*, 32(3), 477-486.
- Yao, Q., Noeaid, P., Roether, J. A., Dong, Y., Zhang, Q., & Boccaccini, A. R. (2013). Bioglass®-based scaffolds incorporating polycaprolactone and chitosan coatings for controlled vancomycin delivery. *Ceramics International*, 39(7), 7517-7522. <https://doi.org/10.1016/j.ceramint.2013.03.002>
- Yoshimoto, H., Shin, Y. M., Terai, H., & Vacanti, J. P. (2003). A biodegradable nanofiber scaffold by electrospinning and its potential for bone tissue engineering. *Biomaterials*, 24(12), 2077-2082. [https://doi.org/10.1016/S0142-9612\(02\)00635-X](https://doi.org/10.1016/S0142-9612(02)00635-X)
- Zhao, H., Li, L., Ding, S., Liu, C., & Ai, J. (2018). Effect of porous structure and pore size on mechanical strength of 3D-printed comby scaffolds. *Materials Letters*, 223, 21-24. <https://doi.org/10.1016/j.matlet.2018.03.205>



Robust Hybrid Classification Methods and Applications

Friday Zinzendoff Okwonu^{1,2}, Nor Aishah Ahad^{1*}, Innocent Ejiro Okoloko³,
Joshua Sarduana Apanapudor⁴, Saadi Ahmad Kamaruddin^{1,2,5} and Festus
Irimisose Arunaye⁴

¹*School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*

²*Institute of Strategic Industrial Decision Modeling, School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*

³*Faculty of Computing, Dennis Osadebay University, Anwai Rd., Asaba, Nigeria*

⁴*Department of Mathematics, Faculty of Science, Delta State University, P.M.B.1, Abraka, Nigeria*

⁵*Centre for Testing, Measurement, and Appraisal (CeTMA), Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*

ABSTRACT

The sample mean classifier, such as the nearest mean classifier (NMC) and the Bayes classifier, is not robust due to the influence of outliers. Enhancing the robust performance of these methods may result in vital information loss due to weighting or data deletion. The focus of this study is to develop robust hybrid univariate classifiers that do not rely on data weighting or deletion. The following data transformation methods, such as the least square approach (LSA) and linear prediction approach (LPA), are applied to estimate the parameters of interest to achieve the objectives of this study. The LSA and LPA estimates are applied to develop two groups of univariate classifiers. We further applied the predicted estimates from the LSA and LPA methods to develop four hybrid classifiers. These

classifiers are applied to investigate whether cattle horn and base width length could be used to determine cattle gender. We also used these classification methods to determine whether shapes could classify banana variety. The NMC, LSA, LPA, and hybrid classifiers showed that cattle gender could be determined using horn length and base width measurement. The analysis further revealed that shapes could determine banana variety. The comparative results using the two data

ARTICLE INFO

Article history:

Received: 22 November 2021

Accepted: 31 January 2022

Published: 23 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.29>

E-mail addresses:

o.friday.zinzendoff@uum.edu.my (Friday Zinzendoff Okwonu)

aishah@uum.edu.my (Nor Aishah Ahad)

okolokoimno@yahoo.com (Innocent Ejiro Okoloko)

japanapudor@yahoo.com (Joshua Sarduana Apanapudor)

s.ahmad.kamaruddin@uum.edu.my (Saadi Ahmad Kamaruddin)

ifaranaye@yahoo.com (Festus Arunaye Irimisose)

*Corresponding author

sets demonstrated that all the methods have over 90% performance prediction accuracy. The findings affirmed that the performance of the NMC, LSA, LPA, and the hybrid classifiers satisfy the data-dependent theory and are suitable for classifying agricultural products. Therefore, the proposed methods could be applied to perform classification tasks efficiently in many fields of study.

Keywords: Classification, least squares, linear prediction, prediction errors, robust

INTRODUCTION

This article focuses on univariate classification methods. Classification methods often assign an object to the actual groups based on certain rules (Tang et al., 2014). Univariate classification methods have been discussed extensively in different fora (Gupta & Govindarajulu, 1973; Huberty & Holmes, 1983). The most frequently applied classifiers are based on the group mean estimates, univariate time series (Karimi-Bidhendi et al., 2018; Song et al., 2020), and the Bayes probability rule (Harianto et al., 2020; Ye, 2020). Unfortunately, these classifiers are influenced by outliers, thereby resulting in a high misclassification rate. The outliers are weighted or deleted, thereby resulting in significant information loss to minimize the misclassification rate. Other robust estimates used as a plug-in to robustify the mean classifiers are the minimum covariance determinant (MCD) (Hubert et al., 2018; Hubert & Debruyne, 2010; Leys et al., 2019), the S and M estimators (Almetwally & Almongy, 2018; Campbell et al., 1999; Croux et al., 1994; Kordestani et al., 2020; Verardi & McCathie, 2012). However, these estimators applied to develop robust classifiers often result in vital information loss. Thus, the Bayes rule is a unique univariate classifier that does not depend on the mean and covariance methods but may perform poorly if the data set in one group is significantly larger than the data set in the other. To avert the above problems, we proposed robustifying the above methods by estimating the parameters of interest using the least square approach (LSA) and linear prediction approach (LPA). The LSA and the LPA estimate drastically minimize the loss of information and hence are better estimates to be applied as a plug-in to learn or train the classical classifiers.

The concept and applications of the linear prediction approach (LPA) have been discussed in detail (Atal, 2006; Manolakis & Proakis, 1996). Linear prediction is based on the theory of estimation (Marple & Carey, 1989). It is a robust and dependable predictive estimator (Srivastava, 2017). Prediction based on linear or multivariate methods applies information on linear or multivariate variables. For example, let δ be the dependent random variable and $x_i, i = 1, 2, 3, \dots, k$ be the independent random variable, otherwise called the “predictor random variable.” Useful information can be obtained if the Borel functions are defined (Bickel & Doksum, 2015; Dobler, 2002; Lindley, 1999; Penenberg, 2015). The variables used in defining the Borel functions are random variables that generate a subspace

w of the Hilbert space H_{sp} . This concept produces the most tractable prediction variable δ , that relies on x_i . The assumption is that the tractable predictor and the independent variables, x_i have a normal joint distribution (Jaeger, 2006).

The least-square approach (LSA), like other estimation procedures and its variant, has received extensive coverage (He et al., 2021; Drygas, 2021; Yao et al., 2020; Kern, 2016; Miller, 2006). This unique technique can be traced to Galton (1886) though coined by Legendre in 1800s. Pearson and Fisher expanded the work of Galton in diverse ways. The main objective of applying this procedure is to estimate and fit the given data set into the function to obtain numerical value (Miller, 2006). It can be done by considering pairs of observations (Y_n, X_n) , $n = 1, 2, \dots, K$, which consist of the dependent random variables Y_n and the independent random variables X_n . To perform prediction involves the linear combination of these variables (Y_n, X_n) .

The LSA tends to minimize the parameters of interest by estimation based on the sum of squares deviation (Kern, 2016). It is also applied to determine the line of best fit of the given data set. The LSA has been applied to provide solutions to a power line (Girshin et al., 2016), pressure detection (Sun et al., 2015), motor induction (Koubaa, 2006), data fitting (Chen & Liu, 2012), and identification of groundwater pollution (He et al., 2021). On the other hand, the LPA has been applied to solve different problems, including the travel time and modeling the prediction of Covid-19 outbreak (Ogundokun et al., 2020; Olarenwaju & Harrison, 2020), climate change (Hasselmann & Barnett, 1981) and data forecasting (Vaseghi, 2008).

In applied research, measurement errors or data imputation errors frequently occur if the process is not properly monitored or equipment calibration failure, which may result in data point differential often called influential observations or outliers. Influential observation is described as a data point that is far away from most of the data points. Influential observations often alter the performance of the classical methods, such as the nearest mean classifiers (NMC) (Okwonu & Othman, 2012; Skurichina & Duin, 2000) that depends on the sample mean. Hence robust methods are applied to overcome this problem by weighting the data set or deleting the influential observations. However, these procedures often result in information loss. This paper applies the LSA and LPA to obtain robust prediction estimates (Srivastava, 2017) without information loss. We propose two robust classification rules based on the LSA and the LPA estimates. We further apply the predicted estimates from the LSA and LPA to develop four hybrid classifiers. Finally, we compare these proposed methods with the classical nearest mean classifier (NMC) (Okwonu & Othman, 2012; Skurichina & Duin, 2000) and the Bayes classifier.

The comparative classification performance of these methods is also investigated based on the probability of correct classification (PCC) and the percentage performance prediction accuracy (PPPA). These classifiers were adopted to investigate whether horn

measurement can be used to determine cattle gender. We also applied these classifiers to investigate whether their shapes can determine banana variety. This study proposes robust hybrid univariate classifiers that do not expunge outliers, thereby minimizing the loss of vital information. Therefore, the objectives are (1) to minimize the loss of vital information, (2) to minimize the misclassification rate, (3) to derive new hybrid classifiers with robust classification accuracy, (4) to investigate the comparative classification performance of the conventional univariate classifiers and the proposed hybrid classifiers and (5) to investigate the validity of the data dependency theory which states that the performance of any classifier strictly depends on the data structure and sign direction.

The rest of this paper is structured as follows. First, the LSA, LPA, and hybrid methods are explained in Section 2. Then, data collection and analysis are presented in Section 3. Finally, the conclusion follows in Section 4.

MATERIALS AND METHODS

The univariate classifier (UC) and the Bayes classifier (BC) are known classification methods for univariate applications. The UC is based on mean computation, while the BC is designed using the probability concept. In this paper, we will skip the rigors of the formulations and focus on the estimate and plug-in methods.

Linear Prediction Approach (LPA)

The LPA has gained wide coverage to the extent that its coefficients are termed backward and forward autoregression (Eriksson et al., 2019; Mello, 2006; Engle, 1982; Jones, 1978). It has applications in digital signal processing, economics, and many other disciplines (Randall et al., 2020; Tan & Jiang, 2018; Srivastava, 2017; Manolakis & Proakis, 1996; Bultheel & van Barel, 1994). In addition, the LPA produces robust predictive estimates (Srivastava, 2017). We start by defining the dependent and independent random variables to develop the LPA classifier. Let δ denote the dependent random variable and $x_i, i = 1, 2, 3, \dots, k$ be the independent predictor random variables. Suppose:

$$\delta = \theta_0 x^0 + \sum_{i=1}^k \theta_i x_i.$$

The expression can be written as:

$$\delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \dots - \theta_k x_k = \delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \dots - \theta_k x_k = 0.$$

Taking the expectation of the last expression and squaring it, we obtain

$$E|\delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \dots - \theta_k x_k|^2 = 0.$$

It can be expressed as Equation 1:

$$E|\delta|^2 = E|\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k|^2 = |\theta_0|^2 + \left| \sum_{i=1}^k \theta_i x_i \right|^2. \quad (1)$$

It implies that the mean squared error (MSE) is minimized. The focus is to determine the dependent variable δ based on x_i . Then, iterative steps can be introduced on the dependent variable to comply with the predictor variables x_i .

Let $\vartheta = 1$ be a constant, define β as the Hilbert space and φ be the subspace of β , then $\langle \vartheta, x_i, i = 1, 2, 3, \dots, k \rangle$ is the random variable defined on $\varphi \in \beta$ (Bickel & Doksum, 2015; Lindley, 1999). This process can be viewed as a “minimization problem.” Further analysis revealed that $\delta = \langle \delta_i \rangle$ could be paired with $\langle \vartheta, x_i, i = 1, 2, 3, \dots, k \rangle$. The first pairing based on expectation property yields

$$\gamma_{i\delta} = \frac{|\delta, \vartheta|}{|\vartheta|^2} \times |\vartheta| = |\delta, \vartheta| = E(\delta) = \bar{\delta}.$$

It satisfies $E|\delta - \vartheta|^2 = \bar{\delta}$ and $\alpha = \delta - E(\delta)$. The variance of δ , that is δ^2 can be computed in a similar procedure. Suppose there exists a random variable x such that where $\hat{x} = \mu = E(x)$ and the variance of x defined as $var(x) = E|x - \mu|^2$. Based on the projection concept we have

$$\begin{aligned} \gamma_{i\delta} &= \frac{|\delta, \vartheta|}{|\vartheta|^2} \times |\vartheta| + \frac{|\delta, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| = \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| + \frac{|\bar{\delta}, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| \\ &= \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| + \frac{0}{|\varepsilon|^2} \times |\varepsilon| \\ &= \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon|, (|\bar{\delta}, \varepsilon| = 0). \end{aligned}$$

Recall that the covariance between δ and x is denoted as $cov(\delta, x) = (\alpha, \varepsilon)$; it indicates the relationships between the two variables. From the last expression, we obtain $\frac{|\alpha, \varepsilon|}{|\varepsilon|^2} = \rho_{\delta x} \sqrt{\frac{\sigma_\delta^2}{\sigma_x^2}}$. It implies that the prediction can be performed as Equation 2:

$$\widehat{\gamma}_{i\delta} = \bar{\delta} + \rho_{\delta x} \sqrt{\frac{\sigma_\delta^2}{\sigma_x^2}} \times |\varepsilon|, \tag{2}$$

where $\rho_{\delta x}$ denotes the correlation between δ and x , this implies that σ_δ^2 and σ_x^2 are the variance of δ and x , respectively. The analysis indicates that δ can be estimated based on x .

From Equation 2, we can derive the LPA classifier as follows. First, we obtain each group predicted estimate; therefore, Equation 2 can be defined as group predictors such that $k = 1, 2$; then, we restate Equation 2 as Equation 3:

$$\widehat{Y}_{i\delta_k} = \bar{\delta}_k + \rho_{\delta x_k} \sqrt{\frac{\sigma_{\delta_k}^2}{\sigma_{x_k}^2}} \times |\varepsilon_k|. \quad (3)$$

The cutoff mark of Equation 3 is defined in Equation 4:

$$\mathfrak{m} = \frac{\sum_{k=1}^2 \widehat{Y}_{i\delta_k}}{2}. \quad (4)$$

The LPA classifier assigns an object to group one (G_1) if $\widehat{Y}_{i\delta_1} < \mathfrak{m}$. otherwise to group two (G_2) if $\widehat{Y}_{i\delta_1} > \mathfrak{m}$.

The Least Square Approach (LSA)

Prediction by the LSA has received detailed attention in the literature. However, this subsection adopts a brief discussion on its prediction approach. The LSA can be stated as Equation 5:

$$\hat{y} = a + bx, \quad (5)$$

where x denotes the predictor variable, \hat{y} denotes the estimate of the response variable, and b denotes the slope,

$$b = \frac{SS_{xy}}{SS_{xx}}, SS_{xy} = \sum xy - \frac{\sum x \sum y}{k}, SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{k}, \bar{x} = \frac{\sum_{i=1}^k x_i}{k}, \bar{y} = \frac{\sum_{i=1}^k y_i}{k},$$

and the intercept is given as $a = \bar{y} - b\bar{x}$. We apply Equation 5 to derive the LSA classifier by restating Equation 5 as Equation 6:

$$\hat{y}_k = a_k + b_k x_k. \quad (6)$$

Then the cutoff mark of Equation 6 is defined as Equation 7:

$$\beta = \frac{\sum_{k=1}^2 \widehat{y}_k}{2}. \quad (7)$$

The LSA classifies an object (\hat{y}_1) to group one (G_1) if $\hat{y}_1 \geq \beta$, otherwise assigns \hat{y}_1 to group two (G_2) if $\hat{y}_1 < \beta$.

Hybrid Linear Prediction Classifier (HLPC)

From the above discussions, we have given a detailed description of the LPA concerning prediction and classification. The output from Equation 2 is assumed to be a robust estimate to perform group classification for univariate cases. In this subsection, we will invoke the univariate classifier (Huberty & Holmes, 1983), Bayes classifier (Theodoridis

& Koutroumbas, 2009; Ma et al., 2011), the smart univariate classifier (SUC), and smart univariate Bayes classifier (SUBC) to form the four hybrid classifiers categorized as the hybrid linear prediction classifiers (HLPC). We will not undergo derivational details but apply the LPA and LSA predictive values as a plug-in to train these classifiers.

Univariate Linear Predictive Classifier (ULPC)

We apply the estimates from Equation 2 as input to construct the ULPC model and decision boundary as Equations 8 and 9:

$$\bar{q}_k = \frac{\sum_{k=1}^2 (\widehat{\gamma_{i\delta k}})}{n_k}, \tag{8}$$

$$\widehat{\gamma_{i\delta k}} > \frac{\bar{q}_1 + \bar{q}_2}{2}. \tag{9}$$

Therefore Equations 8 and 9 are the univariate linear predictive classifier based on the decision rule from Huberty and Holmes (1983).

Linear Predictive Bayes Classifier (LPBC)

Based on Equation 2 and the concept discussed in Theodoridis and Koutroumbas (2009) and Ma et al. (2011), the LPBC is stated as Equations 10 and 11:

$$P(G_1 | \widehat{\gamma_{i\delta k, p \times 1}}) = \frac{w_1 t_1(\widehat{\gamma_{i\delta k, p \times 1}})}{\sum_{k=1}^2 w_k t_k(\widehat{\gamma_{i\delta k, p \times 1}})} = \left(\sum_{k=1}^2 w_k t_k(\widehat{\gamma_{i\delta k, p \times 1}}) \right)^{-1} w_1 t_1(\widehat{\gamma_{i\delta k, p \times 1}}) \tag{10}$$

$$P(G_2 | \widehat{\gamma_{i\delta k, p \times 1}}) = \frac{w_2 t_2(\widehat{\gamma_{i\delta k, p \times 1}})}{\sum_{k=1}^2 w_k t_k(\widehat{\gamma_{i\delta k, p \times 1}})} = 1 - \left(\sum_{k=1}^2 w_k t_k(\widehat{\gamma_{i\delta k, p \times 1}}) \right)^{-1} w_1 t_1(\widehat{\gamma_{i\delta k, p \times 1}}) \tag{11}$$

Hence, assign $\widehat{\gamma_{i\delta 1}}$ to G_1 if $P(G_1 | \widehat{\gamma_{i\delta 1, p \times 1}}) \geq P(G_2 | \widehat{\gamma_{i\delta 2, p \times 1}})$ otherwise, allocate $\widehat{\gamma_{i\delta 1}}$ to G_2 if

$$P(G_1 | \widehat{\gamma_{i\delta 1, p \times 1}}) < P(G_2 | \widehat{\gamma_{i\delta 2, p \times 1}}).$$

The LPBC applies the Bayes classifier rule to assign an object to the actual group.

Smart Univariate Linear Predictive Classifier (SULPC)

We apply the input from Equation 2 as a plug-in to train the classifier as Equations 12 and 13

$$T_1 = \frac{(\bar{q}_1 - \bar{q}_2)'}{S_{\gamma_{i\delta, p \times 1}}^2} w_1 \gamma_{i\delta 1, p \times 1}, \quad (12)$$

$$T_2 = \frac{(\bar{q}_1 - \bar{q}_2)'}{S_{\gamma_{i\delta, p \times 1}}^2} w_2 \gamma_{i\delta 2, p \times 1}. \quad (13)$$

Then the F -weight (w_k) and the pooled variance ($S_{\gamma_{i\delta, p \times 1}}^2$) based on Equation 2 are stated as

$$w_1 = \frac{\gamma_{i\delta 1, p \times 1}}{Z}, w_2 = \frac{\gamma_{i\delta 2, p \times 1}}{Z}, Z = \gamma_{i\delta 1, p \times 1} \gamma_{i\delta 2, p \times 1}^T, S_{G_1}^2 = \frac{\sum_{i=1}^{G_1} (w_1 \gamma_{i\delta 1, p \times 1} - \bar{q}_1)^2}{n_{G_1} - 1},$$

$$S_{G_2}^2 = \frac{\sum_{i=1}^{G_2} (w_2 \gamma_{i\delta 2, p \times 1} - \bar{q}_2)^2}{n_{G_2} - 1}, S_{\gamma_{i\delta, p \times 1}}^2 = \frac{(n_{G_k} - 1) \sum_{k=1}^2 S_{G_k}^2}{\sum_{k=1}^2 n_{G_k} - 2}.$$

The group evaluation criteria are obtained as Equation 14 to evaluate the performance of this method:

$$T_{t1} = \frac{(\bar{q}_1 - \bar{q}_2)'}{S_{\gamma_{i\delta, p \times 1}}^2} \bar{q}_1, \quad T_{t2} = \frac{(\bar{q}_1 - \bar{q}_2)'}{S_{\gamma_{i\delta, p \times 1}}^2} \bar{q}_2, \quad T_{t1t2} = \frac{T_{t1} + T_{t2}}{2}. \quad (14)$$

Therefore, an object $w_1 \gamma_{i\delta 1, p \times 1}$ is allocated to G_1 if $T_1 \geq T_{t1t2}$; otherwise, assign $w_1 \gamma_{i\delta 1, p \times 1}$ to G_2 if $T_1 \geq T_{t1t2}$.

LPBC/ SULPC

The LPBC/SULPC combines LPBC and SULPC to produce unbiased robust classification results. This combination averts the overfitting problem and upward bias. Overfitting is a process whereby the model predicted value exceeds the given optimal probability of correct classification (PCC).

Hybrid Least Square Classifier (HLSC)

The four methods, i.e., the univariate classifier, Bayes classifier, smart univariate classifier, and smart univariate Bayes classifier discussed in the last subsection, utilize the LSA estimate (Equation 5) to train the different methods. The HLSC consists of the univariate least square classifier (ULSC), least square Bayes classifier (LSBC), the least square smart univariate classifier (LSSUC), and LSBC/LSSUC. Similar plug-in procedures discussed in Equations 8 to 14 are implemented by replacing the LPA estimates with the LSA estimates.

Evaluation Criteria

The evaluation criteria (Huberty & Holmes, 1983) applied in this study are based on Equation 15:

$$C_{\sigma} = 0.5\Phi\left(\frac{\alpha}{2}\right) + 0.5\Phi\left(\frac{\alpha}{2}\right) = \Phi\left(\frac{\alpha}{2}\right) \quad (15)$$

Where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and α is the Mahalanobis distance (Johnson & Wichern, 1992). The probability of correct classification (PCC) denoted by ω , derived from the various methods, is compared with the optimal value (C_{σ}) to investigate the best method. The error of misclassification ($\epsilon = C_{\sigma} - \omega$) can also be applied to determine the robustness of the methods. Equation 16 is called the percentage of performance prediction accuracy (PPPA),

$$\Omega = \left(\frac{\omega}{C_{\sigma}}\right) \times 100 \quad (16)$$

The last expression determines the overall percentage of a correct group membership. It is useful to analyze the performance of the methods at a glance. Equation 16 will be adopted to analyze the comparative performance analysis of these methods.

Data Collection and Analysis

These data sets were collected to investigate the comparative classification performance for the above classifiers. The applications of this study focus on two real data sets from the agricultural sector to determine whether these classifiers could be used to maximally separate different species and varieties of agricultural products. The first data consist of cattle horns measurement in Appendices 1 and 2. The second consists of artificial data on varieties of banana shapes (<https://www.openml.org/d/1460>). The first data set consists of collections of cattle horns for ten months in an abattoir in Abraka, Delta State, Nigeria. This data set consists of two features: horn length and width measured in centimeters (cm) for bull and cow, with 100 instances categorized into two groups. The first group consists of features measured on a bull, while the second group consists of features measured on a cow. This data uses to determine whether the classifiers can accurately predict cattle gender. Appendix 1 consists of the bull data set, categorized as Group one (G_1), and Appendix 2 consists of the cow data set, categorized as Group two (G_2). The banana variety data set originally contains $n = 5,300$ with two attributes. The second group (G_2) consist of $n_2 = 2,376$, we selected $n_1 = 2,376$, hence $n_1 = n_2 = 2,376$, $n = n_1 + n_2$. The details for this data set are contained in (<https://www.openml.org/d/1460>) (KEEL, 2015). The data set was reshuffled into a training set (60%) and a validation set (40%). The mean probability of correct classifications is based on 1000 replications. Both data sets will be analyzed based on the percentage of performance prediction accuracy (PPPA).

RESULTS AND DISCUSSIONS

This section consists of the applications of agricultural production data. The focus is to investigate if these classifiers could be applied to separate animal species or gender and plant varieties into different groups based on measured attributes. This study mimics the univariate measurement for the classification of nanoelectronics and spectroscopy (Leys et al., 2019) and the study by Huberty and Holmes (1983).

Application 1: Cattle Horns Data Set

The results in Table 1 and Figure 1 are based on Equation 16, demonstrating the comparative performance analysis between the classical univariate nearest mean classifier (NMC) and the hybrid NMC based on LSA and LPA predicted estimates. From the analysis, we observed that the LSANMC and LPANMC have higher PPPA values (>95%) than the classical NMC. Furthermore, it shows that the robust predictive estimates of LSA and LPA enhance classification accuracy better than the classical NMC method. The comparative performance analysis is depicted in Figure 1. The findings revealed that the NMC method based on the LSA and LPA is more resistant to influential observations than the classical NMC.

Table 1

Comparative performance of NMC and the hybrid NMC for cattle gender

Classical Methods	PPPA
NMC	64.20
LSANMC	97.00
LPANMC	99.00

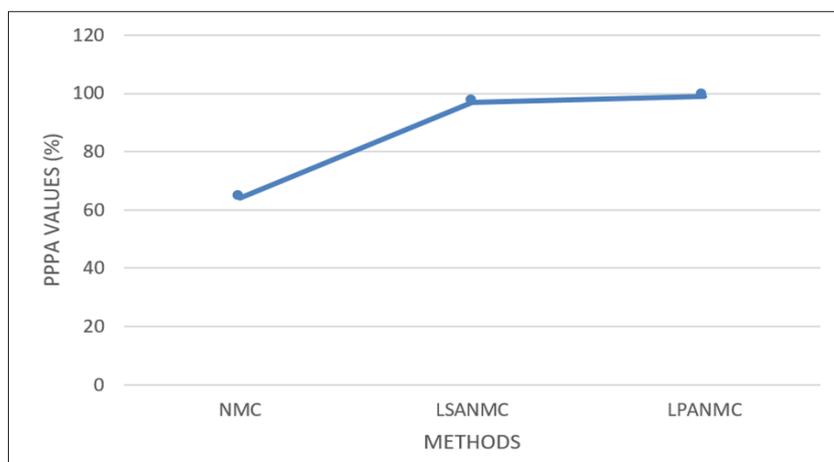


Figure 1. Comparative analysis of the classical NMC and the hybrid NMC for cattle gender

The results in Table 2 are based on Equation 16, demonstrating the comparative performance analysis between the conventional and hybrid methods. The result revealed that all the methods have a very high percentage performance prediction accuracy (PPPA) (Ω). The conventional methods have an average of 95.14% PPPA, while the HLPC and HLSPC have 98.75% and 99.25% PPPA for classifying cattle by gender. Based on the average PPPA for all the methods, the hybrid methods showed more robust classification accuracy than the conventional methods. From this analysis, we remark that the PPPA adopted to analyze the performance of these methods is suitable for the classification task. In Figure 2, the hybrid methods have more robust PPPA values than conventional ones. However, the probability methods (Bayes classifier (BC) and smart univariate Bayes classifier (SUCBC)) demonstrated comparable performance to some of the hybrid methods.

Table 2
Comparative analysis of percentage performance prediction accuracy (PPPA)

Conventional Methods (Average %)		HLPC (%)		HLSPC (%)	
UC	89.54	ULPC	97.00	ULSC	99.00
BC	98.12	LPBC	100.00	LSBC	100.00
SUC	89.05	SULPC	98.00	LSSUC	98.00
SUCBC	98.12	LPBC/ SULPC	100.00	LSBC/ LSSUC	100.00
LPA	99.00				
LSA	97.00				

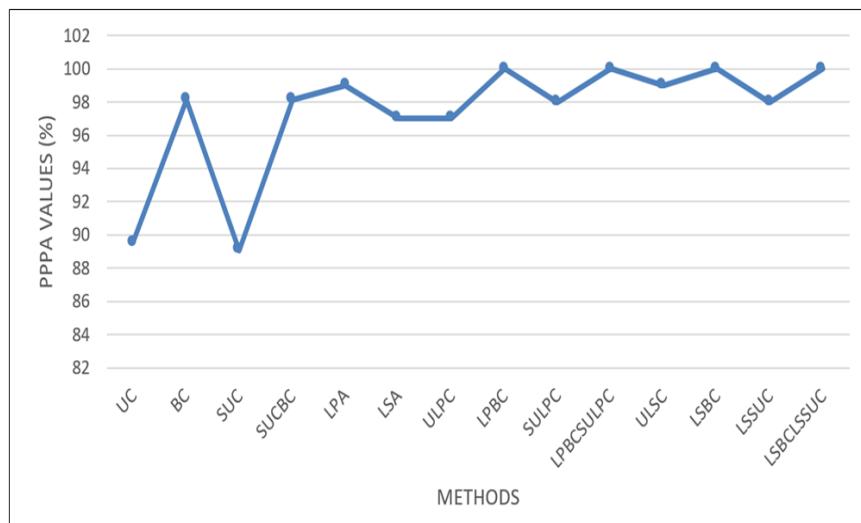


Figure 2. Comparative analysis of PPPA values for cattle gender classification

Application 2: Banana Variety Data Set

The classification results reported in Table 3 show that the classical NMC has better classification performance than the hybrid NMC. However, all the methods have over 80% PPPA, as illustrated in Figure 3.

In comparison to the results reported in Table 1, the analysis in Table 3 showed that the performance of these methods is data-dependent. The former (Table 1) showed an upward trend, while the latter (Table 3) showed vice versa. The implication of the comparative analysis indicated that the performance of any classifier depends on the data set. It affirmed the data dependency theory, which states that the performance of any classification method depends strictly on the data structure and sign direction. The data structure can be continuous or discrete.

Table 3

Comparative performance of NMC and the hybrid NMC for a banana variety

Classical Methods	PPPA
NMC	98.14
LSANMC	84.89
LPANMC	84.89

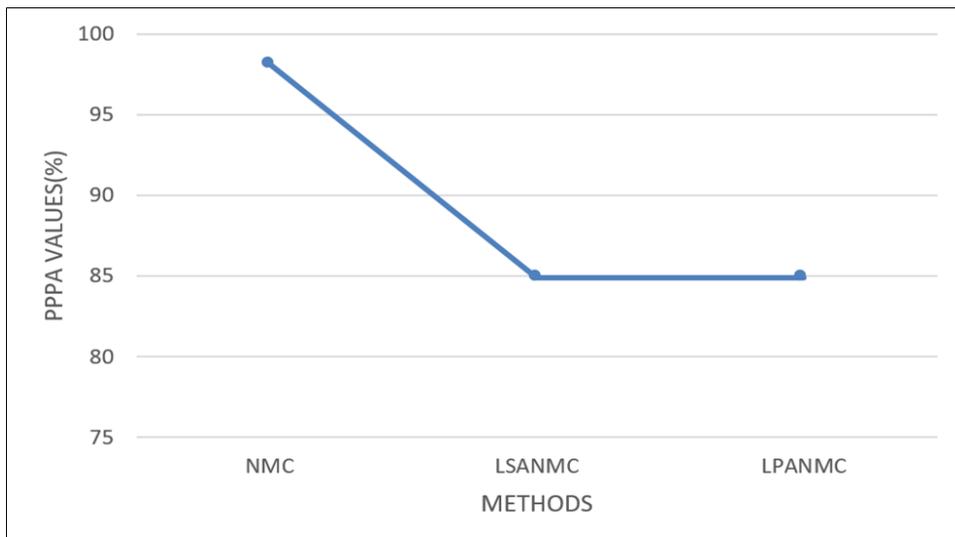


Figure 3. Comparative analysis of the classical NMC and the hybrid NMC for a banana variety

Table 4

Comparative analysis of percentage performance prediction accuracy (PPPA)

Conventional Methods (%)		HLPC (%)		HLSPC (%)	
UC	97.85	ULPC	88.81	ULSC	88.81
BC	99.93	LPBC	93.51	LSBC	93.51
SUC	99.98	SULPC	89.85	LSSUC	89.85
SUCBC	99.96	LPBC/SULPC	93.51	LSBC/LSSUC	93.51
LPA	93.51				
LSA	93.51				

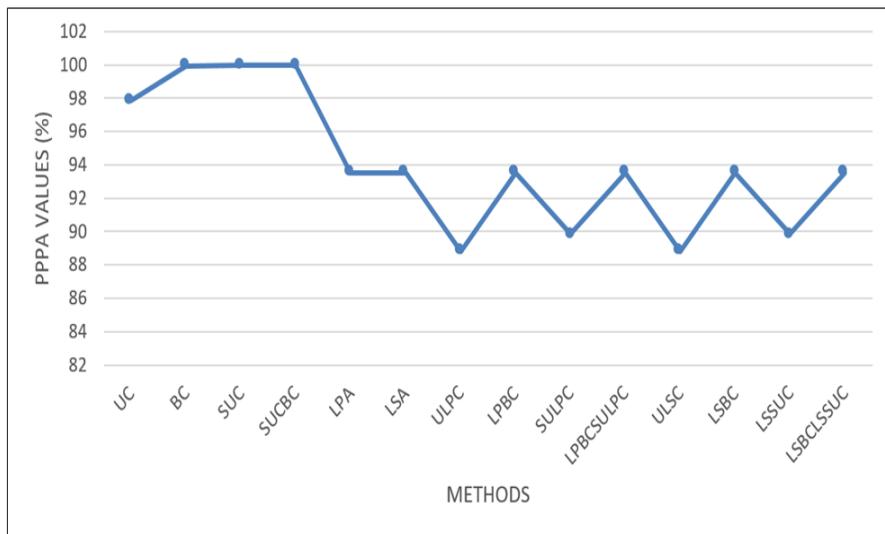


Figure 4. Comparative analysis of PPPA values for banana variety classification by shapes

The comparative percentage performance prediction accuracy based on this data set is reported in Table 4. The result revealed that all the methods have over 88% correct group membership prediction. This result demonstrated that these methods are capable of classifying banana varieties based on shapes with a minimum PPPA of 88.81% and a maximum PPPA of 99.98%. The conventional methods have an average PPPA of 97.46% banana variety classification by shapes, while the hybrid methods HLPC and HLSPC have 91.42% average PPPA, respectively. The hybrid methods based on LPA (HLPC) and LSA (HLSPC) have similar results for this data set. This unique performance showed that the conventional classifiers demonstrated more robust classification accuracy than the hybrid methods. Figure 4 shows that the conventional methods have superior performance over the

hybrid methods. However, the hybrid probability methods (LPBC, LSBC, LPBC/SULPC, LSBC/LSSUC) have the same PPPA values as the LPA and LSA methods.

The cattle data set and the classifiers demonstrated that cattle gender could be determined by horn length and base width measurements. The study also indicated that shapes could classify the banana variety. The 90% average PPPA classification accuracy for all the methods based on the two data sets showed that these methods are robust. The analysis in Figures 1 and 2 imply that the hybrid methods showed superior performance over the conventional methods, while in Figures 3 and 4, the conventional methods altered the superior performance of the hybrid methods except for the probability-based hybrid methods. From this analysis, we observed the upward and downward trends as depicted in Figures 1 to 4, which affirmed the data dependency theory. We also observed that the LSA and LPA showed consistent performance in classifying cattle gender and banana variety. This consistent robust performance was also observed in the hybrid classifiers. The strength of the data dependency theory on the classification methods was obvious in the two data sets, as shown in the upward and downward trends depicted in Figures 1 to 4. The limitations of these classifiers are based on the validity of the data dependency theory.

CONCLUSION

We have proposed the LPA and LSA classification rules and four hybrid classifiers. The evaluation criteria were also established and reinforced into PPPA for easy and fast analysis. The performance of these classifiers is demonstrated using agricultural produce data. The first data set applied to test the performance comparison of these methods was obtained by measuring the length and base width of cattle horns. The second data set consists of two classes of banana variety. The result affirmed that we could apply the proposed LPA, LSA, and the hybrid classifiers to robustly classify cattle into gender based on horn and width length measurement and classify banana variety based on shapes. The investigated LPA and LSA techniques showed comparable classification accuracy with PPPA of over 90%. The analysis revealed that the proposed and hybrid classifiers are robust enough to perform classification based on these data sets. These techniques generally showed varying high percentage performance prediction accuracy based on the data sets. The results demonstrated that these methods could be applied as alternative classifiers to perform classification tasks. We remark that the results reported in this paper affirmed the effects of data dependency theory on classification methods. We look forward to extending these classifiers to multi-dimensional applications in the future.

ACKNOWLEDGEMENT

The authors want to express our gratitude to Universiti Utara Malaysia and Delta State University for supporting this research work. In addition, the authors acknowledged Dr. Okwonu and Ms. Blessing. for contributing the cattle data set for this study. However, this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Almetwally, E. M., & Almongy, H. M. (2018). Comparison between M-estimation, S-estimation, and MM estimation methods of robust estimation with application and simulation censoring. *International Journal of Mathematical Archive*, 9(11), 1-9.
- Atal, B. S. (2006). The history of linear prediction. *IEEE Signal Processing Magazine*, 23(2), 154-161.
- Bickel, P. J., & Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics* (Vol. 1, 2nd Ed). Chapman and Hall/CRC.
- Bultheel, A., & van Barel, M. (1994). Linear prediction: Mathematics and engineering. *Bulletin of the Belgian Mathematical Society*, 1, 1-58.
- Campbell, N. A., Lopuhaä, H. P., & Rousseeuw, P. J. (1999). On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine*, 17(23), 2685-2695. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981215\)17:23<2685::AID-SIM35>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0258(19981215)17:23<2685::AID-SIM35>3.0.CO;2-W)
- Chen, C., & Liu, C. J. (2012). The application of total least squares method in data fitting of speed radar. *Applied Mechanics and Materials*, 203, 69-75. <https://doi.org/10.4028/www.scientific.net/amm.203.69>
- Croux, C., Rousseeuw, P. J., & Hossjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association*, 89(428), 1271-1281. <https://doi.org/10.2307/2290990>
- Dobler, P. C. (2002). Mathematical statistics: Basic ideas and selected topics. *The American Statistician*, 56(4), 332-332. <https://doi.org/10.1198/tas.2002.s204>
- Drygas, H. (2011). On the relationship between the method of least squares and Gram-Schmidt orthogonalization. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 15(1), 3-13. <https://doi.org/10.12697/ACUTM.2011.15.01>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007. <https://doi.org/10.2307/1912773>
- Eriksson, A., Preve, D., & Yu, J. (2019). Forecasting realized volatility using a nonnegative semiparametric model. *Journal of Risk and Financial Management*, 12(3), Article 139. <https://doi.org/10.3390/jrfm12030139>
- Girshin, S. S., Kuznetsov, E. A., & Petrova, E. V. (2016, May). Application of least square method for heat balance equation solving of overhead line conductors in case of natural convection. In *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)* (pp. 1-5). IEEE Publishing. <https://doi.org/10.1109/ICIEAM.2016.7911417>.

- Gupta, A. K., & Govindarajulu, Z. (1973). Some new classification rules for c univariate normal populations. *Canadian Journal of Statistics*, 1(1-2), 139-157. <https://doi.org/10.2307/3314996>
- Hariato, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi algoritma Naïve Bayes classifier untuk mendeteksi anomaly dengan univariate fitur selection [Naïve Bayes algorithm optimization classifier to detect anomalies with univariate feature selection]. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 40-49. <https://doi.org/10.29408/edumatic.v4i2.2433>
- Hasselmann, K., & Barnett, T. P. (1981). Techniques of linear prediction for systems with periodic statistics. *Journal of Atmospheric Sciences*, 38(10), 2275-2283. [https://doi.org/10.1175/1520-0469\(1981\)038<2275:TOLPFS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<2275:TOLPFS>2.0.CO;2)
- He, Z., Zuo, R., Zhang, D., Ni, P., Han, K., Xue, Z., Wang, J., & Xu, D. (2021). A least squares method for identification of unknown groundwater pollution source. *Hydrology Research*, 52(2), 450-460. <https://doi.org/10.2166/nh.2021.088>
- Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant. *WIREs Computational Statistics*, 2(1), 36-43. <https://doi.org/10.1002/wics.61>
- Hubert, M., Debruyne, M., & Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *WIREs Computational Statistics*, 10(3), Article e1421. <https://doi.org/10.1002/wics.1421>
- Huberty, C. J., & Holmes, S. E. (1983). Two-group comparisons and univariate classification. *Educational and Psychological Measurement*, 43(1), 15-26. <https://doi.org/10.1177/001316448304300103>
- Jaeger, B. (2006). The method of least squares. In A. A. Lazakidou (Ed.), *Handbook of Research on Informatics in Healthcare and Biomedicine* (pp. 181-185). IGI Global. <https://doi.org/10.4018/978-1-59140-982-3.ch023>
- Johnson, R. A., & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis* (3rd Ed.). Prentice-Hall.
- Jones, R. H. (1978). Multivariate autoregression estimation using residuals. In *Applied Time Series Analysis I*, (pp. 139-162). Academic Press. <https://doi.org/10.1016/B978-0-12-257250-0.50009-X>
- Karimi-Bidhendi, S., Munshi, F., & Munshi, A. (2018). Scalable classification of univariate and multivariate time series. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1598-1605). IEEE Publishing. <https://doi.org/10.1109/BigData.2018.8621889>
- KEEL. (2015). *Banana*. OpenML. <https://www.openml.org/d/1460>
- Kern, M. (2016). *Numerical Methods for Inverse Problems*. John Wiley & Sons.
- Kordestani, M., Hassanvand, F., Samimi, Y., & Shahriari, H. (2020). Monitoring multivariate simple linear profiles using robust estimators. *Communications in Statistics - Theory and Methods*, 49(12), 2964-2989. <https://doi.org/10.1080/03610926.2019.1584314>
- Koubaa, Y. (2006). Application of least-squares techniques for induction motor parameters estimation. *Mathematical and Computer Modelling of Dynamical Systems*, 12(4), 363-375. <https://doi.org/10.1080/13873950500064103>
- Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, 32(1), 1-10. <https://doi.org/10.5334/irsp.289>

- Lindley, D. V. (1999). Introduction to the practice of statistics, by David S. Moore and George P. McCabe. Pp. 825 (with appendices and CD-ROM).£ 27.95. 1999. ISBN 0 7167 3502 4 (WH Freeman). *The Mathematical Gazette*, 83(497), 374-375. <https://doi.org/10.2307/3619120>
- Ma, D., Wei, W., Hu, H., & Guan, J. (2011). The application of Bayesian classification theories in distance education system. *International Journal of Modern Education and Computer Science*, 3(4), 9-16.
- Manolakis, D. G., & Proakis, J. G. (1996). *Digital Signal Processing. Principles, Algorithm, and Applications* (4th Ed.). Prentice-Hall International Inc.
- Marple, S. L., & Carey, W. M. (1989). Digital spectral analysis with applications. *The Journal of the Acoustical Society of America*, 86, Article 2043. <https://doi.org/10.1121/1.398548>
- Mello, L. (2006). *Linear Predictive Coding as an Estimator of Volatility*. arXiv e-Print. <https://doi.org/10.48550/arXiv.cs/0607107>
- Miller, S. J. (2006). *The method of least squares*. Williams College. https://web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/MethodLeastSquares.pdf
- Ogundokun, R. O., Lukman, A. F., Kibria, G. B., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, 5, 543-548. <https://doi.org/10.1016/j.idm.2020.08.003>
- Okwonu F. Z., & Othman, A. R. (2012). A model classification technique for linear discriminant analysis for two groups. *International Journal of Computer Science Issues (IJCSI)*, 9(3), 125-128.
- Olarenwaju, B. A., & Harrison, I. U. (2020). Modeling of COVID-19 cases of selected states in Nigeria using linear and non-linear prediction models. *Journal of Computer Sciences Institute*, 17, 390-395.
- Penenberg, D. N. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd edn, vols I and II PJ Bickel and KA Doksum, 2015 Boca Raton, Chapman and Hall–CRC xxii+ 548 pp., \$99.95 (vol. I); 438 pp., \$99.95 (vol. II) ISBN 978-1-498-72380-0. *Journal of the Royal Statistical Society Serives*, 179(4), 1128-1129.
- Randall, R. B., Antoni, J., & Borghesani, P. (2020). Applied digital signal processing. In R. Allemang & P. Avitabile (Eds), *Handbook of Experimental Structural Dynamics* (pp.1-81). Springer. https://doi.org/10.1007/978-1-4939-6503-8_6-1
- Skurichina, M., & Duin, R. P. (2000). Boosting in linear discriminant analysis. In *International Workshop on Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science*, (Vol. 1857, pp. 190-199). Springer. https://doi.org/10.1007/3-540-45014-9_18
- Song, K., Wang, N., & Wang, H. (2020). A metric learning-based univariate time series classification method. *Information*, 11(6), Article 288. <https://doi.org/10.3390/info11060288>
- Srivastava, S. (2017). *Fundamentals of linear prediction*. The Institute for Signal and Information Processing. https://www.isip.piconepress.com/courses/msstate/ece_7000_speech/lectures/current/lecture_03/paper/paper.pdf
- Sun, W., Zuo, F., Dong, A., & Zhou, L. (2015). Application of least square curve fitting algorithm based on LabVIEW in pressure detection system. In *2015 International Conference on Applied Science and Engineering Innovation* (pp. 39-43). Atlantis Press.

- Tan, L., & Jiang, J. (2018). *Digital Signal Processing: Fundamentals and Applications*. Academic Press.
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. In C. C. Aggarwal (Ed.), *Data Classification: Algorithms and Applications* (pp. 1-28). Chapman and Hall/CRC. <https://doi.org/10.1201/b17320>
- Theodoridis, S., & Koutroumbas, K. (2009). Classifiers based on Bayes decision theory. In *Pattern recognition* (pp. 13-28). Academia Press.
- Vaseghi, S. V. (2008). *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons.
- Verardi, V., & McCathie, A. (2012). The S-estimator of multivariate location and scatter in stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, 12(2), 299-307. <https://doi.org/10.1177/1536867X1201200208>
- Yao, T. T., Bai, Z. J., Jin, X. Q., & Zhao, Z. (2020). A geometric Gauss-Newton method for least squares inverse eigenvalue problems. *BIT Numerical Mathematics*, 60(3), 825-852. <https://doi.org/10.1007/s10543-019-00798-9>
- Ye, N. (2020). Naïve Bayes classifier. In *Data Mining* (pp. 1-6). CRC Press. <https://doi.org/10.1201/b15288-5>

Appendix 1*Length and base width measurements of bull horns (group one)*

Horn length (x) (cm)	Horn width (y) (cm)						
17.6	8.6	23.6	10.4	17.6	8.6	23.6	10.4
21.5	9.5	22.7	10.9	21.5	9.5	22.7	10.9
21	7.8	24.2	11.9	21	7.8	24.2	11.9
17.2	6.4	22.4	11	17.2	6.4	22.4	11
16.5	8.2	19.8	8.6	16.5	8.2	19.8	8.6
21.9	8.7	21.2	7.4	21.9	8.7	21.2	7.4
18.1	9.8	20.3	6	18.1	9.8	20.3	6
11.7	8.1	19	5.7	11.7	8.1	19	5.7
15.4	7.2	18.7	5	15.4	7.2	18.7	5
18.2	8.5	19.4	5.3	18.2	8.5	19.4	5.3
17.8	7.4	17.8	4.7	17.8	7.4	17.8	4.7
16.5	6.3	19.8	5.4	16.5	6.3	19.8	5.4
21.5	8.8	17.6	4.2	21.5	8.8	17.6	4.2
22	10.1	18.6	5.1	22	10.1	18.6	5.1
20.1	9	21	6.5	20.1	9	21	6.5
21	10.1	17.7	4.8	21	10.1	17.7	4.8
17.3	7.2	20.2	6	17.3	7.2	20.2	6
19.2	9.4	19.6	5.6	19.2	9.4	19.6	5.6
20.1	10	19.1	5.1	20.1	10	19.1	5.1
18.2	7.4	17.9	4.6	18.2	7.4	17.9	4.6
18	8.2	18.3	5	18	8.2	18.3	5
17.6	7	17.9	4.6	17.6	7	17.9	4.6
18.2	7.3	17.7	4.8	18.2	7.3	17.7	4.8
16.5	9.2	20.2	6	16.5	9.2	20.2	6
21.5	11.6	19.6	5.6	21.5	11.6	19.6	5.6

Appendix 2

Length and base width measurements of cow horns (group two)

Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)
23.3	10.3	29	12.7	23.3	10.3	29	12.7
24.5	8.8	30.1	10.9	24.5	8.8	30.1	10.9
25.6	10.4	30.6	17.1	25.6	10.4	30.6	17.1
24.9	9.2	34	11.2	24.9	9.2	34	11.2
29.3	13.1	26.2	12	29.3	13.1	26.2	12
26.2	9.2	30.1	11.6	26.2	9.2	30.1	11.6
25.9	9	29.5	11.3	25.9	9	29.5	11.3
25	8.9	27.4	10	25	8.9	27.4	10
26.5	10.5	25.9	9.8	26.5	10.5	25.9	9.8
20.3	9.1	28.3	11	20.3	9.1	28.3	11
23.7	10	29	11.2	23.7	10	29	11.2
21.9	9.5	30.2	12	21.9	9.5	30.2	12
27.2	10.7	31	12.4	27.2	10.7	31	12.4
28.1	12.2	27.5	10.8	28.1	12.2	27.5	10.8
26.7	11.9	26.8	10.2	26.7	11.9	26.8	10.2
27.4	11.3	29.6	11.6	27.4	11.3	29.6	11.6
29.1	17.4	28.1	10.8	29.1	17.4	28.1	10.8
27.5	12.6	30	11.8	27.5	12.6	30	11.8
29	12.8	27.4	10.2	29	12.8	27.4	10.2
27.3	11.2	28.8	9.8	27.3	11.2	28.8	9.8
21.9	11.1	28.6	10.3	21.9	11.1	28.6	10.3
23.7	10.4	31.2	12.6	23.7	10.4	31.2	12.6
22.8	13.8	29.6	11.6	22.8	13.8	29.6	11.6
24.9	11	28.1	10.8	24.9	11	28.1	10.8
26.2	12.3	30	11.8	26.2	12.3	30	11.8

Significant Factors in Agile Software Development of Effort Estimation

Pantjawati Sudarmaningtyas^{1,2} and Rozlina Mohamed^{1*}

¹Faculty of Computing, Universiti Malaysia Pahang, 26300 UMP, Gambang, Kuantan, Pahang, Malaysia

²Department of Information System, Universitas Dinamika, 60298 Surabaya, Jawa Timur, Indonesia

ABSTRACT

The Agile effort estimation involves project-related and people-related factors. This research objective is to find the factors that influence Agile effort estimation significantly through path analysis using a structural equation model. This research built an agile effort estimation path coefficient model from six constructs from theories and previous studies. Project-related factors represent by requirement and design implementation constructs. People-related factors are measured by the construct of experience, knowledge, and technical ability. The last construct is the effort itself. SmartPLS is employed for the confirmatory composite analysis and the structural model assessment. The confirmatory composite analysis indicated that all constructs are reliable and valid. Furthermore, the structural model assessment found that all factors of project-related constructs have a positive relationship and significant influence, showing a coefficient path value of 59.1% between requirement and design implementation constructs. All constructs represent people-related factors indicated by the coefficient path value of 67% between experience and knowledge, 42.6% between experience and technical ability, and 54.4% between knowledge and technical ability. In addition, all constructs proved influential simultaneously to effort by

31.1%. Positively contribute provided by requirement, experience, and technology's ability. Significantly influenced provided by constructs of the developer's knowledge and technical ability. The largest effect is given by technical ability, knowledge, and experience on medium and small scales. Contrarily, both constructs from project-related effects can be negligible because there was no influence. Based on the result,

ARTICLE INFO

Article history:

Received: 26 November 2021

Accepted: 14 March 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.30>

E-mail addresses:

pantja@dinamika.ac.id (Pantjawati Sudarmaningtyas)

rozlina@ump.edu.my (Rozlina Mohamed)

*Corresponding author

this study concludes that the significant factors in Agile effort estimation are technical ability, knowledge, and experience.

Keywords: Agile methodology, effort estimation, path analysis, people-related factors, project-related factors, structural equation model

INTRODUCTION

One of the crucial parts of a software project is effort estimation. The activity estimates the effort necessary for developing the software product in either man-days or man-hours. In addition, the effort is a key factor that serves as a basis for calculating the cost and schedule needed for completing the project (Bloch et al., 2012). Generally, an effort is formed by a combination of people and time, which calculates the number of productive working hours needed to complete a job. Man-hours, man-days, man-months, or man-years are typically used to express the units of effort (Trendowicz & Jeffery, 2014).

In contrast, agile software development methodology is more emphasis on coding, shorter delivery cycles, and several iterations to complete the software development. Therefore, the effort estimation in this methodology is crucial because delivery time and project velocity are calculated based on the estimation results. Another characteristic of Agile methodology is the collaborative and cooperative approach between all stakeholders, actively involving users and empowering the team to make decisions (Project-Management.com, 2019). Agile's estimation process is divided into two phases: early estimation and iterative estimation. Early estimation was used to get the initial scope just enough to describe the entire software project. Meanwhile, iterative estimation conducted at the start of an iteration is to anticipate new or change requirements.

The top three methods currently used in agile are machine learning, expert judgment, and algorithmic. The Planning Poker technique is a part of the expert judgment method most implemented in Agile. Although this technique is in accordance with Agile characteristics, the result has shown biased value (Sudarmaningtyas & Mohamed, 2021). In addition, these methods heavily rely on the estimator experience and rarely involve other factors in the process of estimation.

The Agile estimation identified factors that are critical in determining software effort. In addition, people factors are also important because the estimation process involves a team of developers from different disciplines (Munialo & Muketha, 2016). Therefore, this study is conducted to find the significant factors that influence effort estimation in Agile.

Considering the importance of estimation efforts in Agile, expected the significant factors produced in this research can contribute to improving the existing Agile effort estimation method. In addition, implementation of those factors expected creates a better estimate of effort and increases accuracy while reducing the technicality and time required.

This paper organizes into five sections. Section 1 contains the research background, followed by describing related works. Section 3 discusses the research methodology that involves the development of the path coefficient model, a questionnaire based on our research questions and hypothesis, and examining and assessing the structural model. Section 4 is the result and discussion of our findings, and the last section is wrapped up our conclusions.

RELATED WORKS

Agile software development (ASD) methods are lightweight methods that focus on simplicity, speed, self-organizing teams, and involving the customer as part of the team. Simplicity and speed can be achieved through more straightforward design and iterative development cycles. Each development cycle emphasizes delivering a demonstrable working product that focuses on the main functions. The customer involved in each development cycle is like a team member so that the requirement customer, who often changes, can be anticipated quickly by the team (Abrahamsson et al., 2002; Bourque & Fairley, 2014).

ASD is based on an iterative and incremental development model that promotes rapid response to changes and focuses on customer satisfaction, timely and continuous delivery, informal methods and minimal planning (Fernandez-Diego et al., 2020). ASD methods have several advantages, such as delivering working software faster, dealing with changing user requirements, and promoting better working relationships among all stakeholders (Zhang et al., 2010). Those advantages can significantly reduce the apparent overhead associated with heavyweight, plan-based methods used in large-scale software development projects and resolve issues of slow execution on concurrent engineering (Boehm, 2006; Bourque & Fairley, 2014).

Scrum, Xtreme Programming (XP), Test Driven Development (TDD), Agile Unified Process (AUP), Kanban, and Distributed Agile Software Development (DASD) are ASD dynamic methods available for developing a software product. Each method has unique attributes and qualities, so choosing the proper method is critical and should be based on the project requirements (Fernandez-Diego et al., 2020; Prakash & Viswanathan, 2017).

Effort estimation is a crucial part of ASD for three reasons: 1) maximize project velocity; 2) optimize individual developer effort across multiple projects; 3) optimal scheduling to achieve global efficiencies (Malgonde & Chari, 2019). However, most of the effort estimation method used in ASD is the expert judgment method, especially the Planning Poker technique (Sudarmaningtyas & Mohamed, 2021).

Following the characteristics of ASD, effort estimation is usually conducted by structured group consensus and lightweight approaches. In addition, it is also based on the user story that delivers in one or two sentences and contains pieces of functionality or

feature software worth for the user (Trendowicz & Jeffery, 2014). Commonly, estimates are conducted periodically, internally generated, and must be involved stakeholders to compromise, review, and reach an agreement concerning the number of resources and time to finish the projects (Bourque & Fairley, 2014). Providing an approximation of the resources needed to complete a project, especially the delivery of products or services in accordance with the specified characteristics of functional and non-functional, is the objective of the estimation effort (Project Management Institute, 2017).

Instead of giving a single value, the estimation value is better expressed in intervals because the estimator has confidence in the possibility that the actual efforts will be within range (Jørgensen, 2016). In addition, it could also provide three estimates: best-case (optimistic), normal-case (most likely), and worst-case (pessimistic) scenarios. To obtain best-case and worst-case scenarios, multiply percentages according to the organization's norms and confidence level in the opportunity in question with normal-case (Chemuturi, 2009).

The Agile effort estimation is influenced by factors related to the project and factors related to people. The factors related to the project include the type of project, quality requirements, hardware and software requirements, ease of operation, complexity, data transaction, and multiple sites. People-related factors consist of communication skills, familiarity with a team, managerial Skills, security, working time, experience with the previous project, and technical ability (Popli & Chauhan, 2014). Accuracy in the agile effort estimation relied heavily on the expertise of professionals in software development. Thereby expert groups' estimates could diminish the optimism bias arising from group discussion (Lenarduzzi et al., 2015; López-Martínez et al., 2018; Mahnič & Hovelja, 2012).

METHODOLOGY

Development of Path Coefficient Model

In the Agile effort estimation, the project factors are essential to determine. In addition, people factors are also important because the estimation process involves a team of developers from different disciplines (Munialo & Muketha, 2016; Popli & Chauhan, 2014). People-related factors represented each team's expertise, experience, domain knowledge, and technical ability (Popli & Chauhan, 2014). For better results, estimation effort should include observance of various factors derived from software development methodology. For that reason, project-related factors acquired from an Agile development methodology consist of two activities which are: requirement engineering and design implementation (Chemuturi, 2009; Sommerville, 2011). The mapping factors are based on the theoretical aspect of the previous research described in Table 1.

Table 1
Mapping factors based on the theoretical aspect

People Factors	Agile Development Methodology	
	Requirement Engineering	Design Implementation
Experience	The quality requirements	Ease of operation
Domain knowledge		Complexity
Technical ability		Data transaction

This study constructs a relation of factors grounded in previous literature studies, and the graph consists of three parts, as depicted in Figure 1. The first part, indicated with the orange dash line, represents project-related factors derived from the Agile development methodology column in Table 1, comprising two variables. The second part, the area surrounded by the blue dash line, corresponds with the people factors column, containing three variables. Finally, the last part reflects the effort itself. Survey results will confirm the relation of factors in this graph to achieve this research objective.

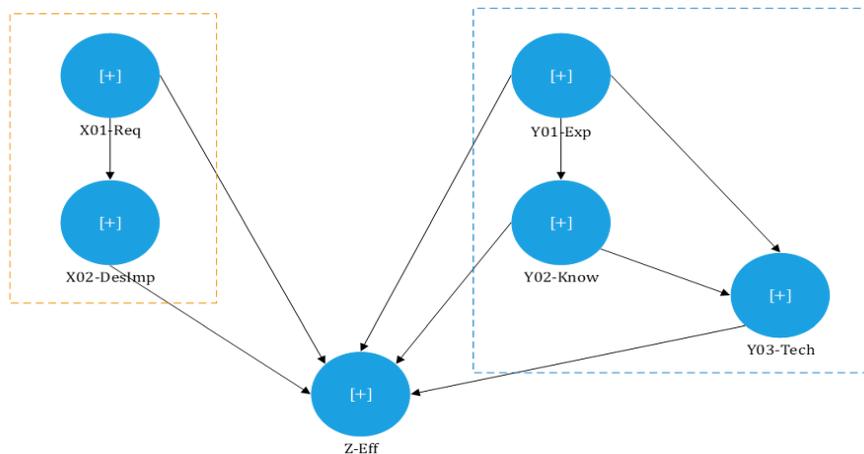


Figure 1. The factors relation grounded in previous literature studies

Project-related factors comprise the requirement and design implementation variables, symbolized by X01-Req and X02-DesImp, respectively. People-related factors, indicated by Y01-Exp, Y02-Know, and Y03-Tech, are represented by experience, knowledge domain, and technical ability. In addition, the effort constructed is denoted by Z-Eff. Based on theory, variable in the project-related factors has a correlation, where the requirement influences

design implementation. On the other hand, variable in the people-related factors has a relation that assumes that developers' experience would leverage their knowledge and technical ability. In addition, technical ability is also affected by knowledge. Therefore, the effort is influenced by the requirement, design implementation, experience, knowledge, and technical ability.

This study proves assumptions on this model through nine hypotheses constructed based on research questions emerge. Table 2 reveals the research questions derived from this model and the hypotheses related to the research questions. The first column contains research questions about the relationship of any constructs, while the next column explains hypotheses related to those research questions.

Table 2
Research question lead hypotheses

Research Question	Hypothesis
RQ1: Do requirements influence the design implementation?	H1: The requirement construct is positively related to the design implementation construct.
RQ2: Do requirements affect effort estimation?	H2: The requirement construct is positively related to the effort construct.
RQ3: Does design implementation affect the effort?	H3: The construct of design implementation is positively related to the effort construct.
RQ4: Does experience have an impact on a developer's knowledge domain?	H4: The experience construct is positively related to the knowledge domain construct.
RQ5: Does experience influence a developer's technical ability?	H5: The experience construct is positively related to the technical ability construct.
RQ6: Do experience influence effort estimation?	H6: The experience construct is positively related to the effort construct.
RQ7: Does the knowledge affect the developer's technical ability?	H7: The knowledge domain construct has positive relationships with the technical ability construct.
RQ8: Does knowledge have an impact on effort estimation?	H8: The knowledge domain construct has positive relationships with the effort construct.
RQ9: Does the technical ability influence effort estimation?	H9: The construct of technical ability has a positive relationship with effort.

The relationship and influence between constructs are captured in Table 3. The relation notation between two constructs is written in the first column, the relation direction in the second column, and the associated hypotheses number in the last column.

Table 3
Structural relationships and hypotheses

Structural Relationships	Direction	Hypothesis
X01-Req -> X02-DesImp	+	H1
X01-Req -> Z-Eff	+	H2
X02-DesImp -> Z-Eff	+	H3
Y01-Ex -> Y02-Know	+	H4
Y01-Ex -> Y03-Tech	+	H5
Y01-Ex -> Z-Eff	+	H6
Y02-Know -> Y03-Tech	+	H7
Y02-Know -> Z-Eff	+	H8
Y03-Tech -> Z-Eff	+	H9

The first column in Table 3 contains a notation of relations between two constructs, and the second column contains the relation direction. For example, the structural relationship denoted X01-Req -> X02-DesImp indicates that the requirement construct influences the design implementation construct. When X01-Req positively influences X02-DesImp, the direction column fills with a plus sign (+). In contrast, the negative direction relationship signaled the negative impact between the two constructs.

Questionnaire

To confirm the relations of those factors conducted by the survey through distributed questionnaire was developed based on six constructs in Figure 1, and several indicators measure each construct. The association between factor, construct, and indicator is revealed in Table 4.

Table 4
Factor, construct, and indicator relationship

Factors Derived from Previous Literature Studies	Construct	Indicator	Description
Requirement Engineering	Requirement	Functional requirement	Measure the quality requirements
		Non-functional requirement	

Table 4 (Continue)

Factors Derived from Previous Literature Studies	Construct	Indicator	Description
Design Implementation	Design Implementation	User interface	Assess the ease of operation
		Software and hardware interfaces	
		Diversity of technology	Represent the complexity
		The sophisticated or novelty technology	
		Coding complexity	
		Database size	
Database complexity			

Construct is a representative of a conceptual definition built from a set of indicators. An indicator is in the form of a single variable used in conjunction with one or more variables to form a composite measure, where each indicator has attributes. Attributes embedded in indicators are complexity and size because these attributes mostly used agile effort estimation in the last three years (Fernandez-Diego et al., 2020; Sudarmaningtyas & Mohamed, 2021). All constructs, indicators, and attributes used in the questionnaire are briefly described in Table 5, and the questionnaire used to collect data is entirely served in Appendix 1.

Table 5
Questionnaire constructs and indicators

Construct	Scales Type	Indicator	Description
X01-Req (Requirement)	1:Never (0%) 2: Occasionally (30%) 3:Sometimes (50%) 4:Normally (80%) 5:Always (100%)	X1: Functional requirement Complexity	Frequency of the complexity of <i>functional requirements</i> considered by developers in estimating efforts.
		X2: Non-functional requirement Complexity	Frequency of the complexity of <i>non-functional requirements</i> considered by developers in estimating efforts.

Table 5 (Continue)

Construct	Scales Type	Indicator	Description		
X02-DesImp (Design Implementation)	1:Never (0%) 2: Occasionally (30%) 3:Sometimes (50%) 4:Normally (80%) 5:Always (100%)	X3: User interface Complexity	In effort estimation, directly or indirectly, consider the complexity of the software user interface		
		X4: Software and hardware interfaces Complexity	In effort estimation, directly or indirectly, consider the complexity of interfaces between software and hardware		
		X5: Diversity of technology	In effort estimation, directly or indirectly, consider the diversity technology that uses		
		X6: The sophisticated or novelty technology	In effort estimation, directly or indirectly, consider the sophisticated or novelty technology that uses		
		X7: Coding Complexity	In effort estimation, directly or indirectly, consider the difficulty level of coding		
		X8: Database Size	In effort estimation, directly or indirectly, consider the database size		
		X9: Database Complexity	In effort estimation, directly or indirectly, consider the database complexity		
		Y01-Exp (Experience)	1: None; 2: Low; 3: Fair; 4: High; 5: Very High.	Y1: Job experience	Adequate job experience can affect a person produce more accurate estimates
				Y2: Effort estimation experience	Experience in estimating effort may affect one individual to produce a more accurate estimate.
Y02-Know (Domain Knowledge)	1: None; 2: Low; 3: Fair; 4: High; 5: Very High.	Y3: Similar project	Frequently being involved in similar projects can affect a person to produce more accurate estimates.		
		Y4: Good track record	A good track record in previous projects can affect a person to produce more accurate estimates.		
Y03-Tech (Technical Ability)	1: None; 2: Low; 3: Fair; 4: High; 5: Very High.	Y5: Technical ability	Developers who have technical expertise can produce more precise estimates.		
		Y6: Involve in many projects	Involving in many projects can affect a person producing more accurate estimates.		

Table 5 (Continue)

Construct	Scales Type	Indicator	Description
Z-Eff (Effort)	0: No; 1: Yes.	Z1: Guess/intuition	Mechanism of estimating effort just by guessing or based on intuition.
		Z2: Experience	Mechanism of estimating effort based on experience.
		Z3: Track record	Mechanism of estimating effort based on previous projects' track record.
		Z4: Attributes	Mechanism of estimating effort using specific attributes that are related to the project.

Requirement constructs have two indicators because software system requirements are generally classified as functional or non-functional. The indicators are used to measure design implementation, classified by the interface, technology, developing programs, and database. The chosen indicator represents an important design implementation process because most agile methods users do not require detailed design documentation. Interface indicator denoted by the user interface and hardware-software interface. Technology is represented by indicators of diversity and sophistication or novelty. Finally, coding is an indicator for developing a program, while the size and complexity indicators represent the database (Hamouda, 2014; Khatri et al., 2016; Pasuksmit et al., 2021; Rosa et al., 2021; Sommerville, 2011; Yuliansyah et al., 2018).

The developer's experience, domain knowledge about the project, and technical ability are important aspects of software project development experience measured by job experience and effort estimation experience. In addition, domain knowledge is measured by being frequently involved in similar projects and having a good track record in previous projects. Finally, technical ability is measured by the developer's technical ability and involvement in many projects (Adnan & Afzal, 2017; Fernandez-Diego et al., 2020; Popli & Chauhan, 2014).

How to perform the effort estimation is an indicator of effort. Performing effort estimation through guess or intuition is the first indicator. The second indicator performs effort estimation based on experience. The third estimation effort is based on the track record in previous projects, and the last uses attributes-related projects to estimate the effort (Popli & Chauhan, 2013).

Respondents

Agile effort estimation is usually done through a group discussion that empowers the developers' team to decide the estimated effort (Sudarmaningtyas & Mohamed, 2020). Therefore, this study is targeting developers as respondents. In addition, estimation efforts

conducted by the developers have higher accuracy than beginner's estimates (Lenarduzzi et al., 2015; López-Martínez et al., 2018; López-Martínez et al., 2017). The expertise of professionals in software development is the primary aspect that influences estimation accuracy through Planning Poker. Therefore, the estimation conducted by expert groups could be diminished the optimism bias arising from group discussion (Lenarduzzi et al., 2015; López-Martínez et al., 2018; Mahnič & Hovelja, 2012).

The sample size was determined by G*Power software because the previous studies did not specifically mention the number of respondents. Our sample size is calculated with criteria t-test linear multiple regression with a significant error of 5%, effect size 0.35, and 5 predictors. With those criteria, the minimal sample size that G*Power suggests is 40. The source data in this study were collected from 41 developers as respondents, fulfilling the minimal sample size. Respondent demographics are classified by gender and three aspects, as revealed in Table 6.

Table 6
Respondent demographics

Aspect		Male	Female	Total
Job Experience	> 10 years	5	3	8
	6–10 years	6	1	7
	3–5 years	10	4	14
	< 3 years	12	0	12
Total		33	8	41
Project Experience	> 10 projects	16	4	20
	6–10 projects	7	0	7
	3–5 projects	7	3	10
	< 3 projects	3	1	4
Total		33	8	41
Conduct Effort Estimation	Always (100%)	11	3	14
	Normally (80%)	17	3	20
	Sometimes (50%)	5	2	7
	Occasionally (30%)	0	0	0
	Never (0%)	0	0	0
Total		33	8	41

Table 6 shows that as much as 80% of respondents are dominated by males, while 20% are female. In addition, 71% of respondents have job experience of more than three years, and 66% have been involved in more than six projects. Most respondents (83%) have experience in effort estimation, and 95% of respondents believe that effort estimation

result is more accurate when attributes associated with the project are considered during estimation—raw data from the survey is contained in Appendix 2.

Examine and Confirm the Structural Model

The structural model was examined and confirmed by executing a confirmatory composite analysis (CCA) by evaluating path coefficients, construct reliability, and construct validity. The indicator loadings measure Path Coefficients. Construct Reliability can be assessed using the traditional Cronbach's alpha and composite reliability approach. However, composite reliability is the preferred reliability metric for structural equation modeling (SEM) statistical techniques. Finally, construct Validity is evaluated by examining convergent and discriminant validity (Hair et al., 2020).

The path coefficient examined by the values of indicator loadings should be at least 0.50 and ideally 0.708 or higher. The reliability of constructs is suggested to be above 0.70 for both Cronbach's alpha and composite reliability. Convergent validity is measured by the average variance extracted (AVE) with a value of at least 0.5. Discriminant validity is evidenced when a reflective construct has the strongest relationships with its indicators compared with any other construct in the path model. Discriminant validity is evaluated through the result of the Fornell-Larcker criterion. On this criterion, the shared variance within the constructs should be larger than the shared variance between the constructs (Bourque & Fairley, 2014).

Assess the Structural Model

The structural model, known as the inner model, was assessed by evaluating the size and significance of the structural path relationships, assess R^2 , examining the f^2 effect size, and evaluating the predictive relevance based on Q^2 . Evaluating the size and significance of the structural path relationships is required to examine whether the determined hypotheses gain empirical support that, indicated by t values, exceed +/- 1.645. The assess R^2 measures the variance of endogenous variables exhibited X02-DesImp, Y02-Know, Y03-Tech, and Z-Eff (Hair et al., 2020).

The f^2 statistics indicate the relative strength of each independent variable in predicting the dependent variable. The f^2 was evaluated by observing the change in R^2 when each independent variable was excluded (Equation 1). The f^2 value of 0.02, 0.15 and 0.35 represents small, medium, and large effects (Hair et al., 2020).

$$f^2 = \frac{R_{included}^2 - R_{excluded}^2}{1 - R_{included}^2} \quad (1)$$

By evaluating the Q^2 , the positive value indicates that the model can accurately predict the data points as reflective indicators of endogenous variables (Hair et al., 2020).

RESULTS AND DISCUSSION

Confirmatory Composite Analysis

The preliminary path coefficient model to find significant factors in Agile effort estimation (AEE) has six constructs and 19 indicators, as revealed in Figure 2. Two constructs in project-related factors are measured by 11 indicators, where requirement constructs have two indicators, while nine support the design implementation construct. Each construction represents a people-related factor measured by two indicators, and the rest contributes to the efforts' construct. We implement the preliminary path coefficient model in SmartPLS.

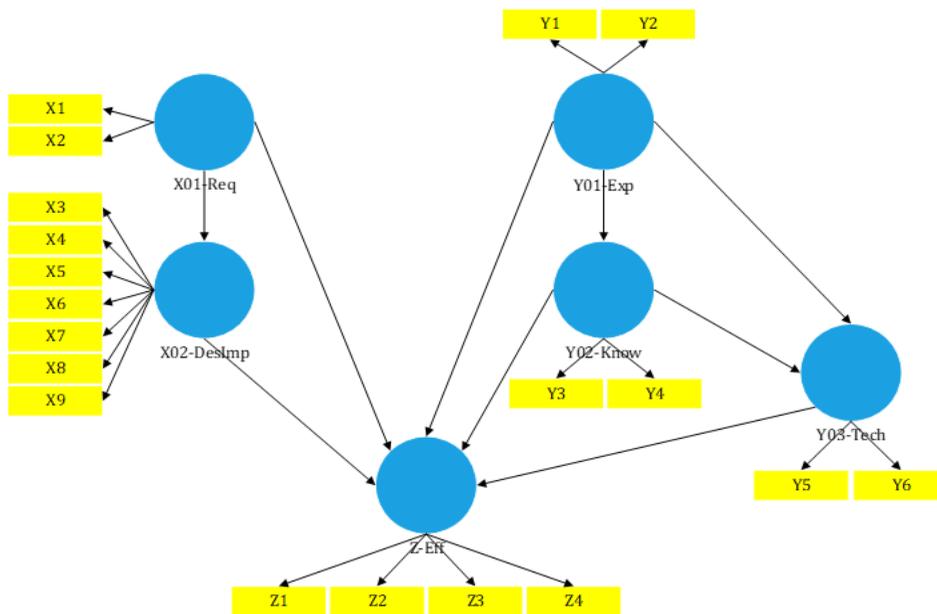


Figure 2. Preliminary path coefficient model

The result of outer loading in the first running found that indicators X7 and Z1 in X02-DesImp and Z-Eff constructs have insignificant t Statistic. Consequently, those indicators were omitted from the preliminary model. Furthermore, the X02-DesImp construct AVE value is lower than requisite in the second execution. Therefore, X8 as the indicator with the lowest outer loading is deleted. Thus, the X02-DesImp construct is expected to reach the value of the required AVE. Finally, all constructs have achieved the requisite AVE value in the third running, and the confirmed path coefficient model is depicted in Figure 3.

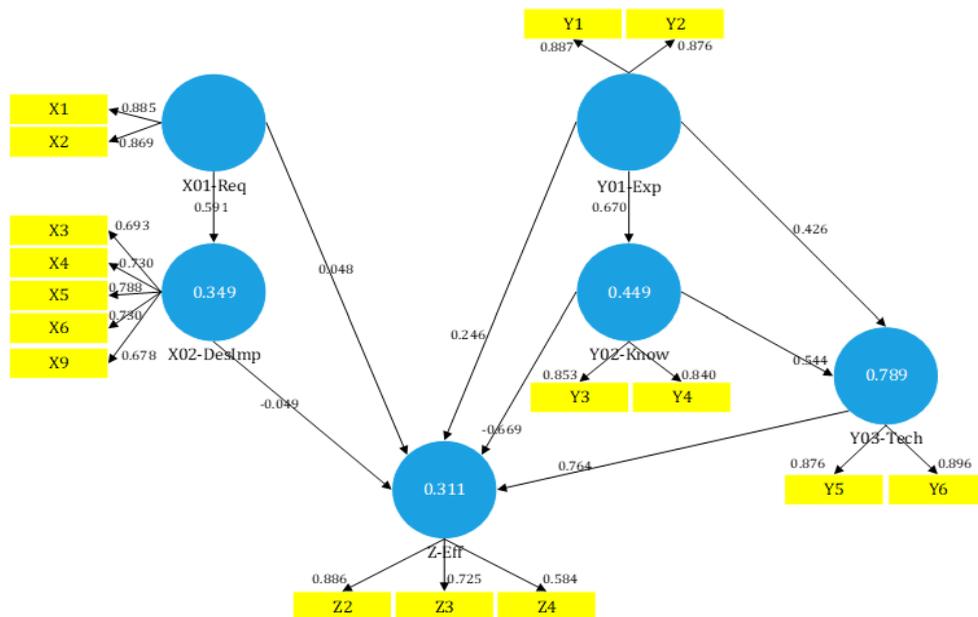


Figure 3. Confirmed path coefficient model

In the confirmed path coefficient model, almost all indicators have ideal factor loadings values, except X3, X9, and Z4. Although the factors are not ideal, the third indicator complied with the rule as the load factor value is above 0.5. Therefore, the third indicator has fulfilled the rule because the loading factor value is above 0.5.

Constructs that represent project-related factors, both indicators in the constructs of Requirement (X01-Req) can be measured significantly with the loadings of functional requirement (X1) is 88.5% and non-functional requirement (X2) is 86.9%. In addition, the construct of Design and Implementation (X02-DesImp) influenced significantly worth 69.3% of the user-interface complexity (X3), 73% of the complexity of the software-hardware interface (X4), 78.8% by a variety of technologies (X5), 73% of sophisticated/novelty of the technology (X6), and 67.8% of complexity database (X9). However, the Design and Implementation constructs are insignificantly influenced by indicators of difficulty of coding (X7) and database size (X8).

In contrast to project-related factors, all indicators on each construct representing the people-related factors can significantly influence because they have loadings factor values above the ideal value. In addition, the Effort (Z-Eff) construct has the same situation as the X02-DesImp construct, where not all indicators are significantly influential. The indicators that significantly influence Z-Eff are Z2, Z3, and Z4, with loadings factor values worth 88.6%, 72.5%, and 58.4%, while indicator Z1 (guessing/intuition) proved no effect. A conventional reliability assessment with Cronbach's alpha indicates that four constructs

have good reliability, while the others are classified as acceptable. Likewise, composite reliability, the preferred reliability metric in sem, reflects good reliability because the value is above 0.70 for all constructs. Therefore, all constructs in this model are reliable and valid, as shown in Table 7.

Convergent validity assessed by AVE demonstrates good validity because the AVE value is above 0.5 for all constructs. In addition, discriminant validity assessed by Fornell-Larcker also indicates good validity. It is supported that the value of the reflective construct with its indicators is higher than any other construct in the path model.

Table 7
Results of reliability assessment

Construct	Reliability Assessment		Validity Assessment	
	Cronbach's Alpha	Composite Reliability	AVE	Fornell-Larcker
X01-Req	0.700**	0.869**	0.769	0.877***
X02-DesImp	0.776**	0.847**	0.525	0.725***
Y01-Exp	0.713**	0.874**	0.777	0.881***
Y02-Know	0.604*	0.835**	0.716	0.846***
Y03-Tech	0.726**	0.879**	0.784	0.886***
Z-Eff	0.576*	0.781**	0.550	0.742***

Note: Acceptable; **Good; ***The value of the reflective construct with its indicators is higher than any other construct.

The measurement quality of the CCA result shows that all constructs in the path coefficient model can be confirmed. This result also assures that the confirmed path coefficient model's six reflectively measured composite constructs are reliable and valid.

The Structural Model Assessment

Estimates of path coefficients and significance of the structural path relationships are conducted to examine whether there is empirical support for the pre-determined hypotheses. Five hypotheses (H1, H4, H5, H7, H9) are empirically supported because they have consistent and significant values with hypothesized ones. Although H2 and H6 have a positive direction but are insignificant; in contrast, H8 has a significant influence although a negative tendency. In addition, H3 has inconsistent signs and is insignificant. Path coefficients and significance testing as the result of SmartPLS revealed in Figure 3 are presented in Table 8.

The path coefficient values and *t* Statistics indicate that the Requirement construct is positively related to and significantly influences the Design and implementation construct. Thereby, the H1 hypothesis is accepted. Thus, an increase of 55.9% in the construct of

Design and implementation for each increase occurred in the Requirement construct. On the other hand, the Requirement construct is positively related to the Effort construct, although the influence is insignificant. Therefore, it means the H2 hypothesis is also acceptable.

Table 8
Path coefficients and significance testing

Hypothesis	Structural Relationships	Path Coefficients	t Statistic
H1	X01-Req -> X02-DesImp	0.591	4.912**
H2	X01-Req -> Z-Eff	0.048	0.184
H3	X02-DesImp -> Z-Eff	-0.049	0.203
H4	Y01-Exp -> Y02-Know	0.670	4.926**
H5	Y01-Exp -> Y03-Tech	0.426	3.686**
H6	Y01-Exp -> Z-Eff	0.246	0.774
H7	Y02-Know -> Y03-Tech	0.544	5.034**
H8	Y02-Know -> Z-Eff	-0.669	2.173*
H9	Y03-Tech -> Z-Eff	0.764	2.511*

Note: * $p < 0.05$; ** $p < 0.01$

Design and Implementation construct are not positively related, and influences are insignificant to the Effort construct. In other words, every increase in Design and Implementation construct will decrease the Effort construct by 4.8%. Due to that reason, the H3 hypothesis is rejected.

The Experience construct has positively related to the constructs of Knowledge, Technical Ability, and Effort. This construct significantly influences the Knowledge construct at 67%, while the Technical Ability construct is influenced by 42.6%. Nevertheless, this construct is insignificant to the Effort construct. Therefore, based on the statement, the hypotheses of H4, H5, and H6 are acceptable.

Knowledge constructs significantly influence the technical ability and Effort constructs, respectively, reaching 54.4% and 66.9%. However, the Knowledge construct contrasts with the other constructs because it has negatively related to the Effort construct. Therefore, hypothesis H7 is acceptable nevertheless rejects hypothesis H8.

Hypothesis H9 is acceptable because the Technical Ability construct has positively associated and significantly influenced the Effort construct. Every improvement in the Technical Ability constructs increases the Effort construct by 76.4%.

Observing R^2 values shows the ability of exogenous constructs to clarify endogenous constructs. The exogenous construct X01-Req gives moderate influences of 34.9% to endogenous construct X02-DesImp, while exogenous construct Y01-Exp gives strong

influences of 44.9% to endogenous construct Y02-Know. Endogenous variable Y03-Tech got strong influence simultaneously from exogenous constructs Y01-Exp and endogenous construct Y02-Know as 78.9%. The last endogenous construct is Z-Eff that simultaneously influenced by constructs X01-Req, X02-DesImp, Y01-Exp, Y02-Know, and Y03-Tech by 31.1%. It indicates that the Effort construct was also influenced by another construct not mentioned in this study.

The interest of this study is in the effect size of endogenous construct Z-Eff. Construct Y01-Exp contributes a small size effect to Z-Eff by 0.028. On the other hand, the constructs of Y02-Know and Y03-Tech give a medium effect size to the Z-Eff. In addition, the effect of X01-Req and X02-DesImp is considered negligible because it did not affect the Z-Eff construct. Constructs that account for most of the variance in Z-Eff are Y03-Tech followed by Y02-Know with effect sizes f^2 of 0.164 and 0.155. The results of examining the f^2 effect size are completely served in Table 9.

The evaluation revealed that all endogenous constructs have Q^2 values: X02-DesImp of 0.148, Y02-Know has 0.279, Y03-Tech by 0.576, and Z-Eff is worth 0.102. This evaluation indicated that all Q^2 values are above 0.0, which means the confirmed path coefficient model can provide relevant predicting for all endogenous constructs.

Table 9
Examine the f^2 effect size

	R^2 for Z-Eff	f^2 Effect Size
R2 Includes all variable	0.311	
R2 Excludes X01-Req	0.309	0.003
R2 Excludes X02-DesImp	0.310	0.001
R2 Excludes Y01-Exp	0.292	0.028
R2 Excludes Y02-Know	0.204	0.155
R2 Excludes Y03-Tech	0.198	0.164

CONCLUSION

This study found that the project-related factors, represented by requirement and design implementation, and people-related factors, which consist of experience, knowledge, and technical ability, proved influential simultaneously to effort by 31.1%. This outcome supports and aligns with Popli and Chauhan's (2014) research result and Munialo and Muketha's (2016). Although all factors simultaneously influence effort, not all factors have a positive relationship and significant influence. According to our acceptable hypotheses, requirement, experience, and technology's ability positively contribute to Agile effort estimation. In addition, Agile effort estimation is significantly influenced by the developer's knowledge and technical ability.

The assessment of the people-related factors provides effect size to effort on a small scale and medium scale, where technical ability (Y03-Tech) gives the largest effect size, followed by knowledge (Y02-know) and experience (Y03-Exp). Contrarily, the effects of project-related can be negligible because there was no influence from both constructs of their representatives. It is consistent with the previous results, stating that requirements, designs and implementation do not significantly influence the effort. Therefore, this study concludes that people-related factors, especially technical ability, knowledge, and experience, are significant in Agile effort estimation.

Our conclusion is supported by the research finding of Asnawi et al. (2012) that software developers' involvement is the top factor in Agile methods. In addition, our acceptable hypotheses strengthened the conclusions research of Ramessur and Nagowah (2020) that stated the two most impacting factors in Agile projects are staff experience and technical ability and complexity of requirements.

Further research is built more accurate Agile effort estimation by implementing those significant factors. In addition, exploring the other factors that influence agile effort estimation is also important because, based on our findings, the simultaneous influence on project-related and people-related factors is 31.1%. This value convinces us that there are still many other indicators that future researchers can explore.

ACKNOWLEDGMENT

The work is supported in part by UMP-UIAM-UITM under Grant Nos. RDU200755 titled "Design and Event Management Prioritization Framework Exploiting Fuzzy Ahp and Fuzzy Topsis Approach."

REFERENCES

- Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2002). *Agile software development methods: Review and analysis*. arXiv Preprint. <http://arxiv.org/abs/1709.08439>
- Adnan, M., & Afzal, M. (2017). Ontology based multiagent effort estimation system for scrum agile method. *IEEE Access*, 5, 25993-26005. <https://doi.org/10.1109/ACCESS.2017.2771257>
- Asnawi, A. L., Gravell, A. M., & Wills, G. B. (2012). Factor analysis: Investigating important aspects for agile adoption in Malaysia. In *2012 Agile India* (pp. 60-63). IEEE Publishing. <https://doi.org/10.1109/AgileIndia.2012.13>
- Bloch, M., Blumberg, S., & Laartz, J. (2012). *Delivering Large Scale IT.pdf*. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/delivering-large-scale-it-projects-on-time-on-budget-and-on-value>
- Boehm, B. (2006). A view of 20th and 21st century software engineering. In *Proceeding of the 28th International Conference on Software Engineering - ICSE '06* (pp. 12-29). ACM Publishing. <https://doi.org/10.1145/1134285.1134288>

- Bourque, P., & Fairley, R. E. (2014). *Guide to the software engineering - Body of knowledge*. IEEE Computer Society.
- Chemuturi, M. (2009). *Software estimation best practices, tools & techniques: A complete guide for software project estimators*. J. Ross Publishing.
- Fernandez-Diego, M., Mendez, E. R., Gonzalez-Ladron-De-Guevara, F., Abrahao, S., & Insfran, E. (2020). An update on effort estimation in agile software development: A systematic literature review. *IEEE Access*, 8, 166768-166800. <https://doi.org/10.1109/ACCESS.2020.3021664>
- Hair, J. F., Page, M., & Brunsveld, N. (2020). *Essentials of business research methods* (4th Ed.). Routledge.
- Hamouda, A. E. D. (2014). Using agile story points as an estimation technique in CMMI organizations. In *2014 Agile Conference* (pp. 16-23). IEEE Publishing. <https://doi.org/10.1109/AGILE.2014.11>
- Jørgensen, M. (2016). Unit effects in software project effort estimation: Work-hours gives lower effort estimates than workdays. *Journal of Systems and Software*, 117, 274-281. <https://doi.org/10.1016/j.jss.2016.03.048>
- Khatri, S. K., Malhotra, S., & Johri, P. (2016). Use case point estimation technique in software development. In *2016 5th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO)* (pp. 123-128). IEEE Publishing. <https://doi.org/10.1109/ICRITO.2016.7784938>
- Lenarduzzi, V., Lunesu, I., Matta, M., & Taibi, D. (2015). Functional size measures and effort estimation in agile development: A replicated study. In *International Conference on Agile Software Development* (pp. 105-116). Springer. https://doi.org/10.1007/978-3-319-18612-2_9
- López-Martínez, J., Ramírez-Noriega, A., Juárez-Ramírez, R., Licea, G., & Martínez-Ramírez, Y. (2017). Analysis of Planning poker factors between university and enterprise. In *2017 5th International Conference in Software Engineering Research and Innovation (CONISOFT)* (pp. 54-60). IEEE Publishing. <https://doi.org/10.1109/CONISOFT.2017.00014>
- López-Martínez, J., Ramírez-Noriega, A., Juárez-Ramírez, R., Licea, G., & Jiménez, S. (2018). User stories complexity estimation using Bayesian networks for inexperienced developers. *Cluster Computing*, 21(1), 715-728. <https://doi.org/10.1007/s10586-017-0996-z>
- Mahnič, V., & Hovelja, T. (2012). On using planning poker for estimating user stories. *Journal of Systems and Software*, 85(9), 2086-2095. <https://doi.org/10.1016/j.jss.2012.04.005>
- Malgonde, O., & Chari, K. (2019). An ensemble-based model for predicting agile software development effort. *Empirical Software Engineering*, 24, 1017-1055. <https://doi.org/10.1007/s10664-018-9647-0>
- Munialo, S. W., & Muketha, G. M. (2016). A review of agile software effort estimation methods. *International Journal of Computer Applications Technology and Research*, 5(9), 612-618.
- Pasuksmit, J., Thongtanunam, P., & Karunasekera, S. (2021). Towards just-enough documentation for agile effort estimation: What information should be documented? IN *2021 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 114-125). IEEE Publishing. <https://doi.org/10.1109/ICSME52107.2021.00017>
- Popli, R., & Chauhan, N. (2013). A sprint-point based estimation technique in Scrum. In *2013 International Conference on Information Systems and Computer Networks* (pp. 98-103). IEEE Publishing. <https://doi.org/10.1109/ICISCON.2013.6524182>

- Popli, R., & Chauhan, N. (2014). Agile estimation using people and project related factors. In *2014 International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 564-569). IEEE Publishing. <https://doi.org/10.1109/IndiaCom.2014.6828023>
- Prakash, B., & Viswanathan, V. (2017). A survey on software estimation techniques in traditional and agile development models. *Indonesian Journal of Electrical Engineering and Computer Science*, 7(3), 867-876. <https://doi.org/10.11591/ijeecs.v7.i3.pp867-876>
- Project-Management.com. (2019). *10 key principles of agile software development*. Project Management.Com. <https://project-management.com/10-key-principles-of-agile-software-development/>
- Project Management Institute. (2017). *A guide to the project management body of knowledge (PMBOK® Guide)* (Sixth Edit). Project Management Institute, Inc. <https://www.pmi.org/>
- Ramessur, M. A., & Nagowah, S. D. (2020). Factors affecting sprint effort estimation. In *Advances in Intelligent Systems and Computing* (Vol. 1089, pp. 507-518). Springer. https://doi.org/10.1007/978-981-15-1483-8_43
- Rosa, W., Clark, B. K., Madachy, R., & Boehm, B. (2021). Empirical effort and schedule estimation models for agile processes in the US DoD. In *IEEE Transactions on Software Engineering* (pp. 1-1). IEEE Publishing. <https://doi.org/10.1109/TSE.2021.3080666>
- Sommerville, I. (2011). *Software engineering* (9th Ed.). Addison-Wesley.
- Sudarmaningtyas, P., & Mohamed, R. (2021). A review article on software effort estimation in agile methodology. *Pertanika Journal of Science and Technology*, 29(2), 837-861. <https://doi.org/10.47836/pjst.29.2.08>
- Sudarmaningtyas, P., & Mohamed, R. B. (2020). Extended planning poker: A proposed model. In *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 179-184). IEEE Publishing. <https://doi.org/10.1109/ICITACEE50144.2020.9239165>
- Trendowicz, A., & Jeffery, R. (2014). *Software project effort estimation: foundation and best practice guidelines for success*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-03629-8>
- Yuliansyah, H., Qudsiah, S. N., Zahrotun, L., & Arfiani, I. (2018). Implementation of use case point as software effort estimation in Scrum Framework. *IOP Conference Series: Materials Science and Engineering*, 403, Article 012085. <https://doi.org/10.1088/1757-899X/403/1/012085>
- Zhang, X., Hu, T., Dai, H., & Li, X. (2010). Software development methodologies, trends, and implications. *Information Technology Journal*, 9(8), 173-178.

Appendix 1

Questionnaire

EFFORT ESTIMATION QUESTIONNAIRE

Effort estimation is a crucial part of the software development process because effort estimation accuracy can help determine the software delivery time to customers. Besides that, the estimation effort can be used to calculate the cost of developing software.

Trendowich (2014) defines effort as a combination of person and time, representing the amount of fully productive work time one person would need to complete a certain work.

This questionnaire aims to determine the developer's point of view about the factors contributing to making an effort estimation and the attributes that influence effort estimation accuracy.

This questionnaire consists of three (3) parts. The first part contains questions about personal data. The second part contains factors that influence the estimation value, and the third part contains parameters that make the estimates more accurate.

To achieve these objectives, please fill out this questionnaire based on your experiences. Researchers would like to thank you for your participation.

* Required

1. Email *

PART 1. PERSONAL DATA

Your personal data is protected, not published, and used only in this research.

2. Name *

3. Gender *

Mark only one oval.

Male

Female

4. *Period of employment (year) **

Mark only one oval.

- < 3
- 3 - 5
- 6 - 10
- > 10

5. *Number of involvement in software development projects. **

Mark only one oval.

- < 3 projects
- 3 - 5 projects
- 6 - 10 projects
- > 10 projects

PART 2.
EFFORT
ESTIMATION
ATTRIBUTES

Commonly, software development projects are done by a team. Any members of the team have a specific role, task, and responsibility. The team's estimation determines the completion time of a project to the tasks under their responsibility. Fill in the following statements based on your perspective and experience while conducting effort estimation for each task.

6. *Every involved in a project, I ... **

Mark only one oval.

- 1 2 3 4 5
-
- Never done effort estimation Always doing the effort estimation

7. *I conduct effort estimation through ... **

Check all that apply.

- Just guessing or based on intuition*
- Based on experience*
- Based on the track record of the previous project*
- Based on specific attributes that related to the project*

8. *Effort estimation more accurate when based on attributes associated with the project. **

Mark only one oval.

- 1 2 3 4 5
-
- Disagree Strongly Agree

Significant Factors in Agile Effort Estimation

9. *Directly or indirectly, I consider the following attributes to create an effort estimation.**

Mark only one oval per row.

	Never (0%)	Occasionally (30%)	Sometimes (50%)	Normally (80%)	Always (100%)
<i>The complexity of software user-interface</i>	<input type="radio"/>				
<i>The complexity of interfaces between software and hardware</i>	<input type="radio"/>				
<i>Variety of technologies that use</i>	<input type="radio"/>				
<i>The use of sophisticated/novelty technologies</i>	<input type="radio"/>				
<i>Functional requirements complexity</i>	<input type="radio"/>				
<i>Non-functional requirements complexity</i>	<input type="radio"/>				
<i>The difficulty level of coding</i>	<input type="radio"/>				
<i>Database size</i>	<input type="radio"/>				
<i>Database complexity</i>	<input type="radio"/>				

10. *Write three (3) attributes that you normally or always use in effort estimation but not yet mentioned in the previous point.*

11. *I generally expressed the value of effort estimation in the form of ... **

Mark only one oval.

- Story points that stated the effort thru the Fibonacci numbers series (1, 2, 3, 5, 8, ...).*
- T-shirt sizing, the effort size stated small, medium, or large labels that indicate level complexity.*
- Time buckets, the effort estimation expressed in days and hours.*

**PART 3.
ACCURACY
ATTRIBUTES**

Several factors can influence the accuracy of a person in estimates. This section used to know factors affecting someone in making accurate estimates.

12. *In my experience, the following attributes can affect a person to produce more accurate estimates. **

Mark only one oval per row.

	No Impact	Low Impact	Fair Impact	High Impact	Very High Impact
<i>Application experience and familiarity.</i>	<input type="radio"/>				
<i>The technical ability of the developer.</i>	<input type="radio"/>				
<i>Effort estimation experience.</i>	<input type="radio"/>				
<i>Involving in many projects.</i>	<input type="radio"/>				
<i>Adequate job experience.</i>	<input type="radio"/>				
<i>A good track record on previous projects.</i>	<input type="radio"/>				

13. *Write three (3) attributes that influence a person to produce a more accurate estimation but not yet mentioned in the previous point.*

This content is neither created nor endorsed by Google.

Appendix 2

Survey result

ID RESPONDEN	Personal Data			Requirements		Design Implementation							Experience		Domain Knowledge		Technical Ability		Effort				
	Gender	Job Experience (years)	Project Experience (projects)	Functional Requirements Complexity	Non-Functional Requirements Complexity	User Interface Complexity	Software and Hardware Interfaces Complexity	Diversity of Technology	The Sophisticated or Novelty Technology	Coding Complexity	Database Size	Database Complexity	Job Experience	Effort Estimation Experience	Similar Project	Good Track Record	Technical Ability	Involve in Many Projects	Guess/Intuition	Experience	Track Record	Attributes	
	X3	X4	X5	X6	X7	X8	X9	Y1	Y2	Y3	Y4	Y5	Y6	Z1	Z2	Z3	Z4						
R_01	F	4	2	4	3	3	2	3	5	3	2	2	4	2	4	4	3	1	0	0	0	0	0
R_02	M	1	2	3	3	1	2	3	2	1	1	1	4	4	5	3	5	0	1	1	1	1	1
R_03	F	4	4	5	3	4	5	4	5	4	4	5	5	5	5	5	5	0	1	1	1	1	1
R_04	M	2	4	5	4	3	5	5	5	3	4	4	5	5	5	5	5	0	1	1	1	1	1
R_05	M	1	4	4	3	4	3	4	4	3	4	4	4	3	4	3	4	0	1	1	0	1	1
R_06	M	1	2	4	2	1	2	3	5	1	1	1	3	5	4	4	5	0	1	1	1	1	1
R_07	M	4	4	5	4	1	5	5	3	1	5	5	5	5	5	3	5	0	1	1	1	1	1
R_08	M	4	4	3	2	2	3	3	3	2	3	3	3	3	3	3	3	0	1	1	1	1	1
R_09	F	3	1	3	3	4	5	4	3	4	4	3	5	4	5	5	5	0	1	1	1	1	1
R_10	F	2	2	5	4	4	5	5	4	4	5	5	5	5	5	5	5	0	1	1	1	1	1
R_11	M	3	4	5	5	5	4	4	5	3	4	4	5	5	5	3	5	0	1	1	1	1	1

ID RESPONDEN	Personal Data			Requirements		Design Implementation							Experience		Domain Knowledge		Technical Ability		Effort			
	Gender	Job Experience (years)	Project Experience (projects)	Functional Requirements Complexity	Non-functional Requirements Complexity	User interface Complexity	Software and Hardware Interfaces Complexity	Diversity of Technology	The Sophisticated or Novelty Technology	Coding Complexity	Database Size	Database Complexity	Job Experience	Effort Estimation Experience	Similar Project	Good Track Record	Technical Ability	Involve in Many Projects	Guess/Intuition	Experience	Track Record	Attributes
				X1	X2	X3	X4	X5	X6	X7	X8	X9	Y1	Y2	Y3	Y4	Y5	Y6	Z1	Z2	Z3	Z4
R_12	M	2	4	4	2	3	4	4	3	5	2	4	3	5	4	3	4	2	1	1	0	1
R_13	M	1	2	4	3	4	4	4	3	3	3	3	5	5	5	5	5	5	0	0	0	1
R_14	M	2	3	5	4	5	5	5	5	4	3	3	4	4	4	4	5	4	0	1	1	1
R_15	F	2	4	4	4	4	3	4	3	5	4	4	5	4	4	5	5	4	0	1	1	1
R_16	M	2	4	5	3	5	5	3	3	3	3	3	4	4	4	4	5	3	0	1	1	0
R_17	M	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	0	1	1	1
R_18	F	4	4	5	5	5	5	4	4	4	3	4	4	5	5	4	5	4	0	1	0	1
R_19	M	1	3	4	4	4	4	3	5	3	3	4	4	4	3	4	4	3	0	1	1	1
R_20	M	2	4	5	4	5	3	4	4	5	4	4	5	5	5	4	4	5	0	1	1	1
R_21	M	2	3	4	3	4	2	3	3	4	2	3	4	3	5	4	5	5	0	1	1	1
R_22	M	3	4	5	4	4	2	5	4	5	5	5	5	5	5	5	5	5	0	1	1	1

Significant Factors in Agile Effort Estimation

ID RESPONDEN	Personal Data			Requirements		Design Implementation							Experience		Domain Knowledge		Technical Ability		Effort			
	Gender	Job Experience (years)	Project Experience (projects)	Functional Requirements Complexity	Non-functional Requirements Complexity	User interface Complexity	Software and Hardware Interfaces Complexity	Diversity of Technology	The Sophisticated or Novelty Technology	Coding Complexity	Database Size	Database Complexity	Job Experience	Effort Estimation Experience	Similar Project	Good Track Record	Technical Ability	Involve in Many Projects	Guess/Intuition	Experience	Track Record	Attributes
	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y1	Y2	Y3	Y4	Y5	Y6	Z1	Z2	Z3	Z4			
R_23	M	1	3	4	4	3	5	3	3	3	3	4	4	4	3	1	0	0	0	0	0	
R_24	M	1	2	4	4	4	5	3	4	3	4	4	4	4	5	0	1	1	1	1	1	
R_25	M	4	4	5	5	5	5	5	5	5	5	5	5	5	5	0	1	1	1	1	0	
R_26	M	4	4	5	4	4	4	3	3	3	3	5	5	5	5	0	1	1	1	1	0	
R_27	M	4	4	5	4	4	3	3	3	3	4	5	4	5	4	0	0	0	0	0	1	
R_28	M	3	4	5	5	4	5	5	5	5	5	5	4	5	5	1	1	1	1	1	1	
R_29	M	3	4	5	5	5	5	4	4	4	3	4	3	5	3	0	1	1	1	1	0	
R_30	M	3	4	5	4	4	5	2	5	5	4	5	4	4	4	0	0	1	0	1	0	
R_31	M	1	3	4	4	4	4	4	4	4	3	4	4	4	3	0	1	0	1	0	0	
R_32	M	1	1	5	5	5	5	5	5	5	2	2	2	2	2	0	1	0	1	0	0	
R_33	M	2	2	4	3	3	5	4	4	4	4	4	4	4	4	0	0	1	0	1	0	
R_34	M	3	4	4	3	4	5	4	4	4	4	5	4	5	5	1	1	1	1	1	0	

ID RESPONDEN	Personal Data			Requirements		Design Implementation							Experience		Domain Knowledge		Technical Ability		Effort			
	Gender	Job Experience (years)	Project Experience (projects)	Functional Requirements Complexity	Non-functional Requirements Complexity	User interface Complexity	Software and Hardware Interfaces Complexity	Diversity of Technology	The Sophisticated or Novelty Technology	Coding Complexity	Database Size	Database Complexity	Job Experience	Effort Estimation Experience	Similar Project	Good Track Record	Technical Ability	Involve in Many Projects	Guess/Intuition	Experience	Track Record	Attributes
				X1	X2	X3	X4	X5	X6	X7	X8	X9	Y1	Y2	Y3	Y4	Y5	Y6	Z1	Z2	Z3	Z4
R_35	M	1	3	3	3	4	3	5	4	5	5	5	5	5	4	4	5	4	1	0	0	0
R_36	M	2	2	5	3	3	3	5	5	5	3	5	4	4	4	5	5	4	0	0	0	1
R_37	F	2	2	5	3	5	4	5	5	3	1	3	2	3	4	2	3	2	0	0	0	1
R_38	F	2	4	5	2	4	1	4	5	4	2	1	2	3	5	5	5	3	0	0	1	0
R_39	M	2	2	4	3	4	4	5	4	4	3	5	3	3	5	3	4	4	0	0	0	1
R_40	M	2	3	5	4	5	5	4	4	3	3	3	3	5	5	3	5	4	1	1	1	1
R_41	M	1	1	5	4	5	4	4	5	5	4	4	4	4	4	4	4	3	1	0	0	0

Transformer Population Failure Rate State Distribution, Maintenance Cost and Preventive Frequency Study Based on Markov Model

Nor Shafiqin Shariffuddin^{1,2}, Norhafiz Azis^{1,3*}, Jasronita Jasni¹, Mohd Zainal Abidin Ab Kadir¹, Muhammad Sharil Yahaya⁴ and Mohd Aizam Talib⁵

¹Advanced Lightning, Power and Energy Research Centre (ALPER), Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Electrical Technology Section, Universiti Kuala Lumpur British Malaysian Institute, 53100 Gombak, Selangor, Malaysia

³Institute of Nanoscience and Nanotechnology Technology (ION2), Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

⁴Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

⁵TNB Research Sdn Bhd, No. 1, Lorong Ayer Itam, Kawasan Institut Penyelidikan, 43000 Kajang, Selangor, Malaysia

ABSTRACT

This work investigates the state distributions of failure rate, performance curve, maintenance cost and preventive frequency of the transformer population through the Markov Model (MM). The condition parameters data of the oil samples known as Oil Quality Analysis (OQA), Dissolved Gas Analysis (DGA), Furanic Compounds Analysis (FCA) and age were analyzed from 370 distribution transformers. This work utilized the

computed failure rate prediction model of the transformer population based on MM using the nonlinear minimization technique. First, the transition probabilities for each state were adjusted based on pre-determined maintenance repair rates of 10%, 20%, and 30%. Next, the failure rate state distributions and performance curves at various states were analyzed. Finally, the maintenance costs and preventive maintenance frequency were estimated utilizing the proposed maintenance policy models and the failure

ARTICLE INFO

Article history:

Received: 3 December 2021

Accepted: 9 March 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.31>

E-mail addresses:

snorshafiqin@gmail.com (Nor Shafiqin Shariffuddin)

norhafiz@upm.edu.my (Norhafiz Azis)

jas@upm.edu.my (Jasronita Jasni)

mzk@upm.edu.my (Mohd Zainal Abidin Ab Kadir)

sharil@utem.edu.my (Muhammad Sharil Yahaya)

aizam.talib@tnb.com.my (Mohd Aizam Talib)

*Corresponding author

rate state probabilities. The result reveals that the transition from state 2 to state 1 with a 30% pre-determined maintenance repair rate can provide an average reduction of failure rate up to 11%. Based on the failure rate state probability, an average increment of maintenance cost from RM 18.32 million to RM 251.87 million will be incurred over 30 years. In total, 85% of the transformer population must undergo maintenance every nine months to avoid reaching very poor conditions.

Keywords: Failure rate, health index, maintenance cost, maintenance policy model, Markov model, preventive maintenance frequency, state distribution, transformer

INTRODUCTION

Transformers are critical components of power systems whereby their failure can significantly impact end users and utilities, such as high replacement costs, revenue losses and customer inconvenience. Maintenance is one of the key approaches to ensuring the reliable operation of transformers (Islam et al., 2017; Tenbohlen et al., 2011). Time-based maintenance is among the common approach that utilities have utilized. Nowadays, several utilities have started to adopt condition-based maintenance to optimize the overall investments and, at the same time, maintain the safe and reliable operation of transformers (Shariffuddin et al., 2021a).

Condition-based management can improve the efficiency of asset maintenance practices, hence lowering associated costs through comprehensive analysis that acts as a preventive strategy against underlying faults (Jahromi et al., 2009). Health Index (HI) is a condition-based management concept that has been adopted as an effective tool to evaluate transformer conditions. HI considers multiple condition parameter data and employs a variety of criteria to determine the condition of transformers that may be inaccessible via individual measurement techniques. The HI is calculated using a scoring method whereby the condition parameter data is classified using weighting and ranking approaches. In accordance with HI, the condition is classified into several categories as defined by Naderian et al. (2008).

Aside from HI, the failure rate is another key factor in examining transformers' reliability, which can also be used to determine the optimized maintenance strategies (Jürgensen et al., 2016a). However, since transformers have a long lifespan, obtaining data for in-service failures is challenging (Ghazali et al., 2015; Jürgensen, 2018). Several studies have implemented techniques such as the proportional hazard model, Bayesian updating scheme and linear interpolation between different inspection outcomes to estimate the failures data based on the available information (Brown et al., 2018; Jürgensen et al., 2016a; Lindquist et al., 2005). One of the biggest challenges in failure rate modeling is the lack of historical failure records (Jurgensen et al., 2018; Lindquist et al., 2005). One

unique approach to determining the failure rate can be carried out based on the condition of the assets (Jürgensen, 2016). Currently, the study to predict the failure rate of transformers is still limited. Some studies mainly focus on single-time condition parameter data and employ statistical data-driven approaches to model the failure rate (Jürgensen, 2016; Jürgensen et al., 2016b). The Markov model (MM) is identified as one of the approaches for predicting the failure rate of transformers. MM has been widely implemented in civil engineering to forecast the states of bridges, pavements, stormwater piping components and steel hydraulic structures (Riveros & Arredondo, 2010; Camahan et al., 1987; Micevski et al., 2002; Stevens et al., 2020). Additionally, it is also used in electrical equipment such as switchgear and transformers (Hamoud & Yiu, 2020; Hoskins et al., 1999). Recently, several studies have been conducted to determine the transformer population's transition probabilities for condition state prediction using the MM (Selva et al., 2018; Yahaya et al., 2017). MM can also be used to analyze the effect of asset maintenance, repair, and replacement on asset state distribution (Camahan et al., 1987; Cesare & Al, 1992; Li et al., 2016). A number of studies have examined the ideal maintenance strategy for healthcare centers using the MM (Carnero & Gómez, 2017; González-Domínguez et al., 2020). Other studies conducted by Yahaya et al. (2018a, 2018b) have employed the updated MM with designated maintenance policy to determine the maintenance and maintenance cost analysis based on the condition state distribution of the transformer population. Currently, the utilization of failure rate based on HI to estimate the maintenance cost and preventive frequency of the transformer population is yet to be explored.

This study examined the effect of pre-determined maintenance policy on the state distributions of failure rate, performance curve, maintenance cost and preventive frequency of the transformer population based on MM. The previous 370 distribution transformers rated at 33/11 kV and 30 MVA with condition parameters data of the oil samples known as Oil Quality Analysis (OQA), Dissolved Gas Analysis (DGA), Furanic Compounds Analysis (FCA) and age are utilized (Shariffuddin et al., 2021b). It is a continuation of a previous work by Shariffuddin et al. (2021b), in which the failure rate state distribution and failure rate performance curve with and without pre-determined maintenance repair rates are compared and analyzed. Finally, the maintenance cost analysis and preventive maintenance frequency based on a pre-determined maintenance policy model are conducted.

METHODS

Markov Model Overview

A Markov process is a memoryless process that predicts future states based on the current state. A Markov process can be modeled as a Markov chain composed of elements P_{ij} , where P_{ij} denotes the probability of a condition transitioning from state i to state j (Borovkov, 2003; McDonald, 2004). This study interpreted the P_{ij} as a probability of equipment decaying

from state i to j at a specific time interval. Two assumptions were taken to simplify the model in this study. First, the future failure rate model was evaluated using natural and monotonic distributions, where the states of transformer failure rate either stayed the same state or changed to the next state. Second, in each row of the MM transition matrix, the probability summation was set to one. In total, five P_{ij} terms were required to construct the transition matrix of MM, as per Equation 1:

$$P = \begin{bmatrix} P_{11} & 1 - P_{11} & 0 & 0 & 0 \\ 0 & P_{22} & 1 - P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & 1 - P_{33} & 0 \\ 0 & 0 & 0 & P_{44} & 1 - P_{44} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [1]$$

The final state, P_{55} , was set to 1, which indicated that all transformers reached and remained in the last state. This final state is known as the absorbing state because once entered, and there is no way to exit without intervention activities such as component repair or replacement. Finally, the transition probabilities matrix was determined using a nonlinear optimization technique to determine the values of four parameters, P_{11} , P_{22} , P_{33} , and P_{44} , by minimizing the sum of the absolute differences between computed and predicted failure rate data as described in Equation 2.

$$\min \sum_{t=1}^N |A(t) - B(t, P)| \quad [2]$$

where N denotes the number of years in each zone, P denotes the transition probabilities (P_{11} , P_{22} , P_{33} , P_{44}), $A(t)$ denotes the computed failure rate at time t , and $B(t, P)$ denotes the MM-predicted failure rate at time t . After generating the transition matrix, the prediction of the future failure rate state in year t can be determined based on Equation 3:

$$F_t = F_0 \times P^t \quad [3]$$

where t denotes the interval number, F_0 denotes the initial state, and P denotes the transition probability matrix.

Updating Markov Model

Apart from forecasting purposes, MM can also be used to aid utilities in planning and implementing maintenance strategies. A pre-determined maintenance repair rate is considered one of the maintenance strategies that can be accomplished by updating the MM. Through the application of maintenance policy in the form of a matrix, the updated transition matrix, P_R , with a pre-determined repair rate was determined; thus, Equation

1 becomes Equation 4. The $P_{\text{repair rate}}$ can be introduced at any state of the failure rate, depending on the maintenance policy. Equation 4 shows an example of an updated MM transition matrix with a pre-determined maintenance repair rate at state 2.

$$P_R = \begin{bmatrix} P_{11} & 1 - P_{11} & 0 & 0 & 0 \\ P_{\text{repair rate}} & (1 - P_{\text{repair rate}}) * P_{22} & (1 - P_{\text{repair rate}}) * 1 - P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & 1 - P_{33} & 0 \\ 0 & 0 & 0 & P_{44} & 1 - P_{44} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [4]$$

Next, the transformer population’s distribution and failure rate trends were evaluated once the transition matrix was updated. The transformer population distribution can be calculated using Equation 5:

$$S_{(t+1)} = S_{(t)} \times P_R \quad [5]$$

$S_{(t+1)}$ represents the next failure rate at the specified interval, $S_{(t)}$ represents the current failure rate, and P_R represents the updated transition matrix.

Next, the failure rate trend was modeled based on Equation 6:

$$FR_{(t+1)} = FR_{(t)} \times P_R \times M^T \quad [6]$$

Where $FR_{(t+1)}$ and $FR_{(t)}$ represent the next failure rate at the specified interval and current failure rate, P_R represents the updated transition matrix, and M^T represents the matrix transform of the failure rate state scales where $R = [1.59 \ 1.99 \ 2.69 \ 3.59 \ 5.60]$ as defined by Shariffuddin et al. (2021b).

Maintenance Policy and Cost

Maintenance is one of the most essential tasks in an electrical power system to ensure the transformers are safe to operate and reduce unscheduled interruptions. The maintenance policy for the transformer can be determined by the MM based on the future state probability of the transformer, and it is represented in the matrix form. The restoration process can be modeled by permitting transformer current states to revert to the previous states through suitable maintenance techniques such as Condition Assessment (CA), minor work that involves Corrective Maintenance (CM) and major work, which consists of refurbishment and replacement. This study determined the maintenance cost based on several assumptions related to the maintenance policy model. First, the CA was performed for transformers in all states and was carried out annually to monitor the transformer. Second, CM was performed on all transformers in state 4 to enhance the failure rate in states 3, 2 and 1, while major work was performed on all transformers in state 5 to enhance the failure rate

in states 4, 3, 2 and 1. Finally, the 10% pre-determined maintenance policy was assigned to each state to simulate the impact of the maintenance cost of the transformer. In this study, the maintenance actions, activities, and estimated cost breakdowns were obtained based on Yahaya et al. (2018b).

An example of a 10% pre-determined maintenance policy model, **M** at state 4, can be seen in Equation 7:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [7]$$

The maintenance policy model was combined with the normal deterioration, and it is known as **PM**. This combination was used to determine the future state probability at a specific year, *t*, based on MM iteration formulation as shown in Equation 8:

$$F_t = F_0 \times (PM)^t \quad [8]$$

Next, the estimated maintenance cost, **C**, of each future state in each year, *t*, was calculated based on Equation 9:

$$Expected\ Cost = F_0 \times (PM)^t \times C_t \quad [9]$$

Maintenance Policy and Preventive Maintenance Frequency

Preventive maintenance is a programmed maintenance activity, such as routine inspections, oil filtration and replacement of transformer components to reduce the probability of failure (Carnero & Gómez, 2017). Therefore, the frequency of maintenance was established considering the asset reliability requirements. The maintenance frequency was determined in this study by providing several assumptions on the maintenance policy model. In order to prevent the failure rate from reaching state 5, preventive maintenance activities were conducted to improve the failure rate to state 2 after the maintenance, and the maximum permissible state of degradation was state 4. All transformers were set as high importance in the electrical power system network. Since there was a limitation on the maintenance record of the transformer under study, the maintenance frequency matrix and transformer frequency of maintenance tests were carried out based on NETA MTS-2011 (White & Widup, 2014). The intervention states, conditions, and the frequency of maintenance tests for the transformer can be seen in Table 1.

Table 1

Intervention state, condition, and frequency of maintenance test

Intervention State	Condition	Frequency of maintenance test (in Months)
1, 2	Good	9
3	Average	6
4	Poor	3

The corresponding maintenance policy model matrix assumed that preventative maintenance prevented the failure rate from reaching state 5 and subsequently restored the failure rate to state 2 after the maintenance work was carried out. Therefore, Equation 10 was expressed as follows:

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad [10]$$

The combination of the maintenance policy model with the normal deterioration is known as **PS**. Based on MM iteration formulation, this combination was utilized to determine the future state probability at a specific year, t , as indicated in Equation 11:

$$F_t = F_0 \times (PS)^t \quad [11]$$

Next, the model was evaluated to assess the relative maintenance frequency of the transformers. Since each state j is associated with the maintenance frequency, thus the expected maintenance frequency of each future state for each year, t was based on the intervention state and frequency of the maintenance test, as shown in Table 1.

RESULTS AND DISCUSSION

Updated Markov Model

The distribution of future state probability for the transformer population in each state based on MM for 30 years is shown in Figure 1. It was plotted using the transition matrices for zones 1 and 2, which were calculated based on Equations 12 and 13.

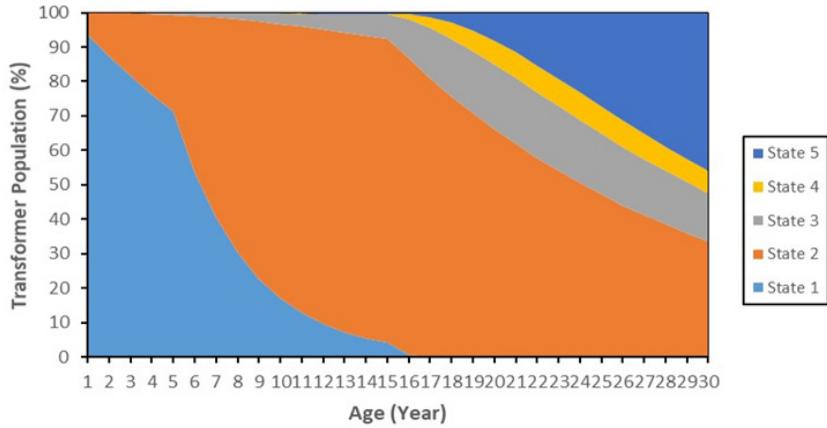


Figure 1. Transformer population distribution in each of the states

$$P = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [12]$$

$$P = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [13]$$

The transformer populations in state 1, state 2, state 3, state 4 and state 5 are 82.07%, 17.69%, 0.18%, 0.07% and 0% during the first five years as shown in Figure 1. At age of 30 years, the transformer populations in state 1, state 2, state 3, state 4 and state 5 change to 0.00%, 33.63%, 13.94%, 6.52% and 45.92%.

The effect of 10%, 20%, and 30% pre-determined maintenance repair rates on the transformer population by updated transition matrices in Equations 12 and 13 were also carried out. For example, the updated transition matrices with 10% pre-determined maintenance repair rates for zone 1 and zone 2, P_R , are shown in Equations 14 and 15:

$$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0.1 & 0.8910 & 0.0090 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [14]$$

$$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0.1 & 0.8910 & 0.0090 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad [15]$$

Based on Equations 5 and 6, the transformer population distribution in each state and the failure rate performance curve were re-plotted using the MM algorithm. Figure 2 depicts the transformer population distribution in each state with a 10% pre-determined maintenance repair rate performed in state 2. During the first 5 years, the transformer populations in state 1, state 2, state 3, state 4 and state 5 are 84.24%, 15.55%, 0.14%, 0.06% and 0%. The transformer populations in state 1, state 2, state 3, state 4 and state 5 change to 4.73%, 37.97%, 13.14%, 5.98% and 38.18% at age of 30 years.

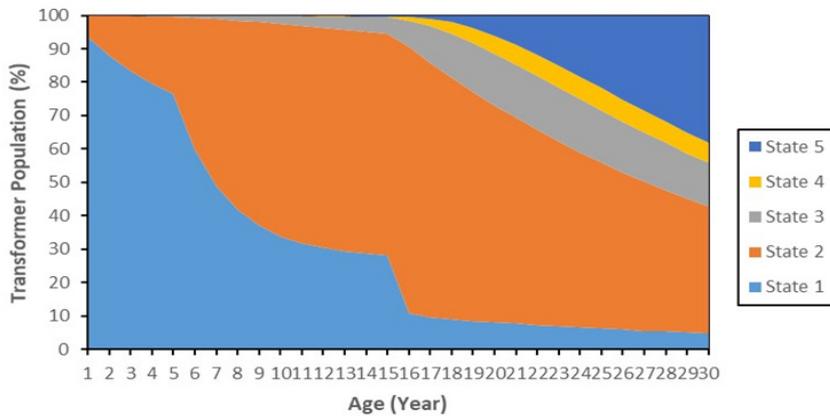


Figure 2. Transformer population distribution in each of the states with a 10% pre-determined maintenance repair rate performed at state 2

The failure rate performance curve with a 10% pre-determined maintenance repair rate performed at state 2 can be seen in Figure 3. It is observed that both actual and predicted failure rate increases with the increment of age. MM’s predicted failure rate performance curve closely follows the actual failure rate performance curve. It is found that the failure rate performance curve decreases by 4.77% once a 10% pre-determined maintenance repair rate is implemented for the transition from state 2 to state 1.

A sensitivity study on various state transitions was carried out to examine the effect of updating transition matrices on the failure rate reduction. The updated transition matrices for each state with a 10% pre-determined maintenance repair rate are shown in Table 2.

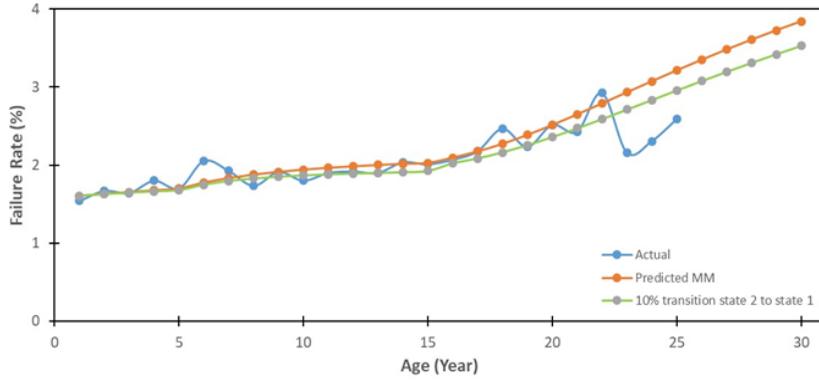


Figure 3. Failure rate performance curves with a 10% pre-determined maintenance repair rate performed at state 2

Table 2

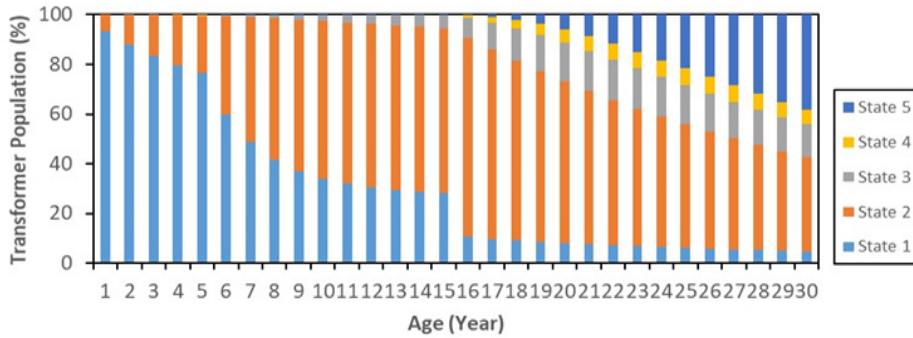
Updated transition matrices with a 10% pre-determined maintenance repair rate for each state

Transition State	Updated Transition Matrices for Zone 1					Updated Transition Matrices for Zone 2				
	1					2				
2 to 1	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0.1 & 0.8910 & 0.0090 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$					$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0.1 & 0.8910 & 0.0090 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$				
3 to 2	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0.1 & 0.4944 & 0.4056 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$					$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0.1 & 0.8910 & 0.0090 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$				
3 to 1	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0.1 & 0 & 0.4944 & 0.4056 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$					$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0.1 & 0 & 0.8910 & 0.0090 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$				
4 to 3	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0.1 & 0.8910 & 0.0090 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$					$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0.1 & 0.4503 & 0.4500 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$				
4 to 2	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0.1 & 0 & 0.8910 & 0.0090 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$					$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0.1 & 0 & 0.4503 & 0.4500 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$				

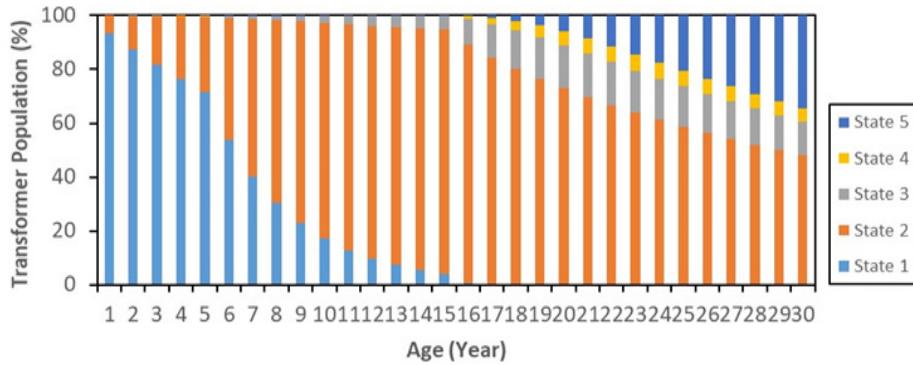
Table 2 (Continue)

Transition State	Updated Transition Matrices for Zone 1	Updated Transition Matrices for Zone 2
4 to 1	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0.1 & 0 & 0 & 0.8910 & 0.0090 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0.1 & 0 & 0 & 0.4503 & 0.4500 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
5 to 4	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0 & 0.1 & 0.9 \end{bmatrix}$	$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0 & 0.1 & 0.9 \end{bmatrix}$
5 to 3	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0 & 0.1 & 0 & 0.9 \end{bmatrix}$	$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0 & 0.1 & 0 & 0.9 \end{bmatrix}$
5 to 2	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0 & 0.1 & 0 & 0 & 0.9 \end{bmatrix}$	$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0 & 0.1 & 0 & 0 & 0.9 \end{bmatrix}$
5 to 1	$P_R = \begin{bmatrix} 0.9348 & 0.0652 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.5493 & 0.4507 & 0 \\ 0 & 0 & 0 & 0.9900 & 0.0100 \\ 0.1 & 0 & 0 & 0 & 0.9 \end{bmatrix}$	$P_R = \begin{bmatrix} 0.7515 & 0.2485 & 0 & 0 & 0 \\ 0 & 0.9900 & 0.0100 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0100 & 0 \\ 0 & 0 & 0 & 0.5003 & 0.4997 \\ 0.1 & 0 & 0 & 0 & 0.9 \end{bmatrix}$

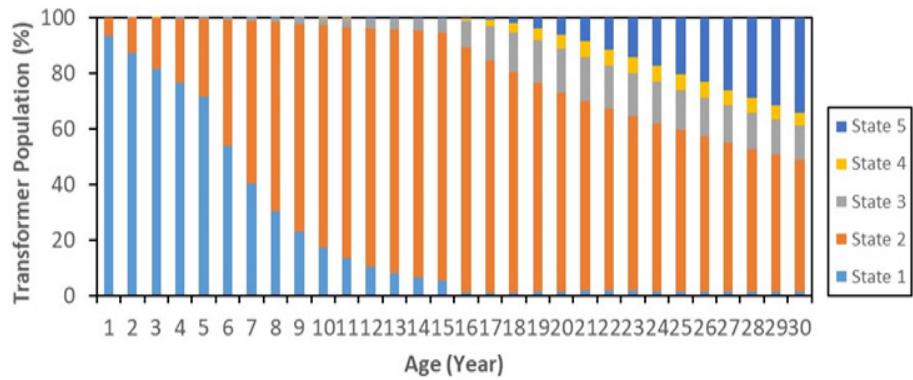
Figure 4 illustrates the transformer population distribution based on various state transitions utilizing a 10% pre-determined maintenance repair rate. An obvious improvement in transformer distribution is observed once the pre-determined maintenance repair rate is implemented during the transition from state 2 to state 1. There is a 46.18% increment in the transformer population in state 1. The transformer population in states 2, 3, 4 and 5 decrease by 9.48%, 16.21%, 16.27% and 19.69% respectively.



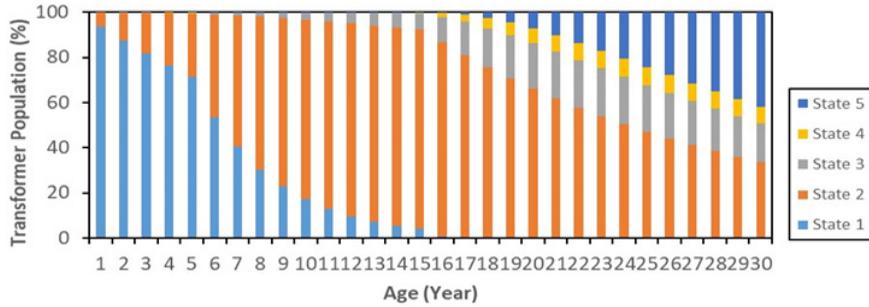
(a)



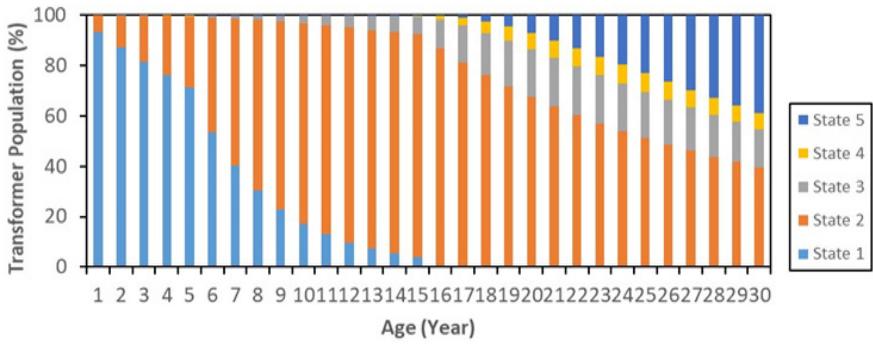
(b)



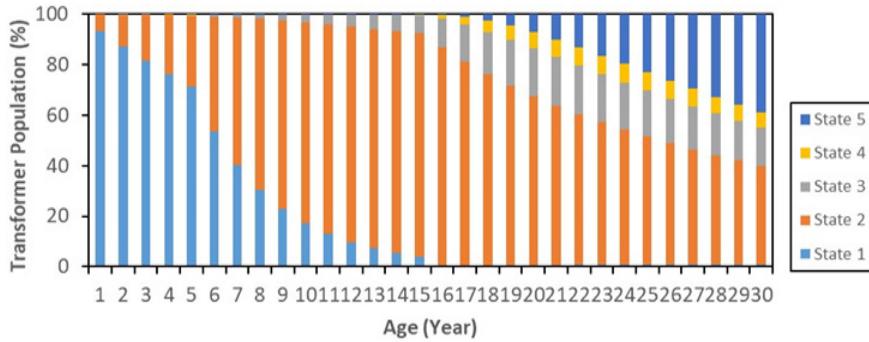
(c)



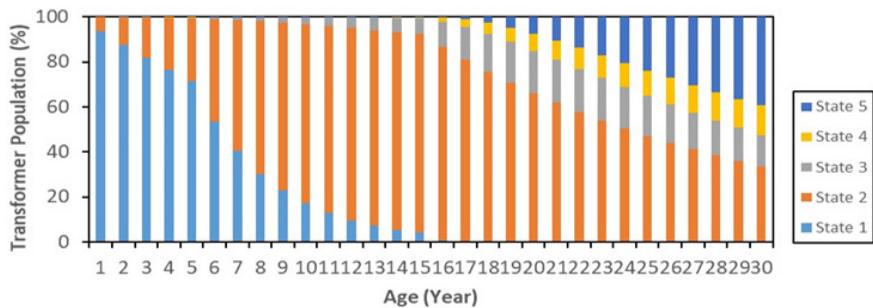
(d)



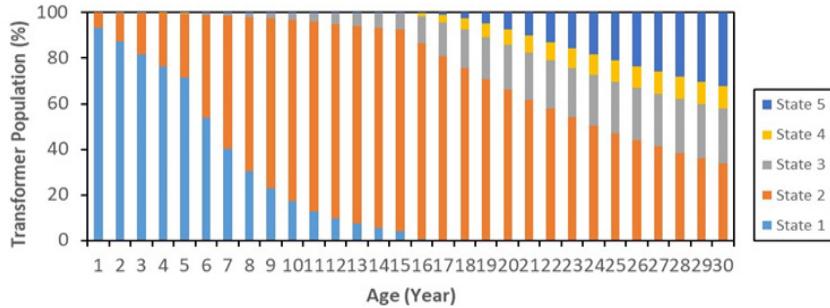
(e)



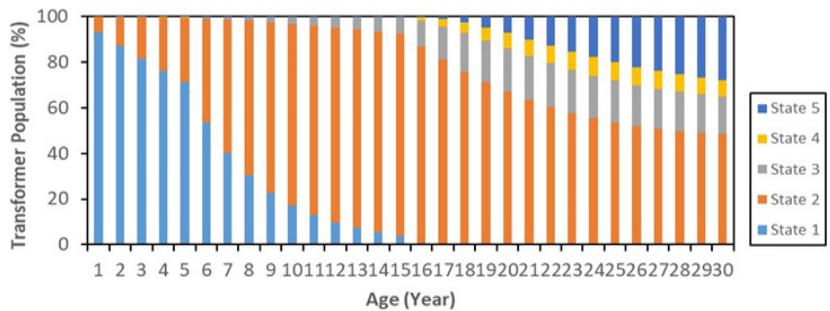
(f)



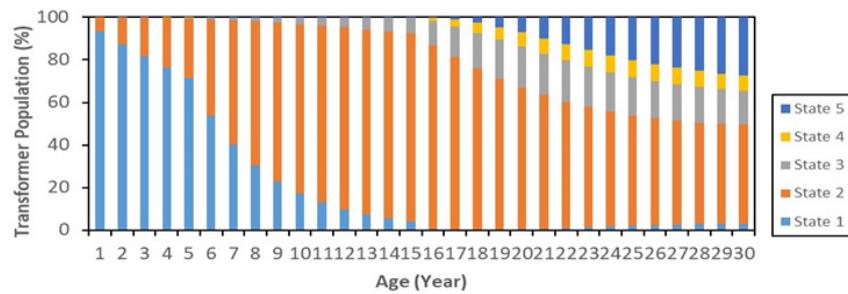
(g)



(h)



(i)



(j)

Figure 4. The transformer population distribution utilizing a 10% pre-determined maintenance repair rate perform at state transitions of (a) state 2 to state 1; (b) state 3 to state 2; (c) state 3 to state 1; (d) state 4 to state 3; (e) state 4 to state 2; (f) state 4 to state 1; (g) state 5 to state 4; (h) state 5 to state 3; (i) state 5 to state 2; (j) state 5 to state 1

The failure rate performance curves with a 10% pre-determined maintenance repair rate on different state transitions are shown in Figure 5. During the first 5 years, there was no significant impact on the failure rate reduction among the various states. However, for the next 5 years, it is observed that the transition from state 2 to state 1 significantly impacts the failure rate reduction up to 4.77%. In addition, throughout the prediction period, it is found that significant reductions in failure rate occur from the age of 9 to 18 years.

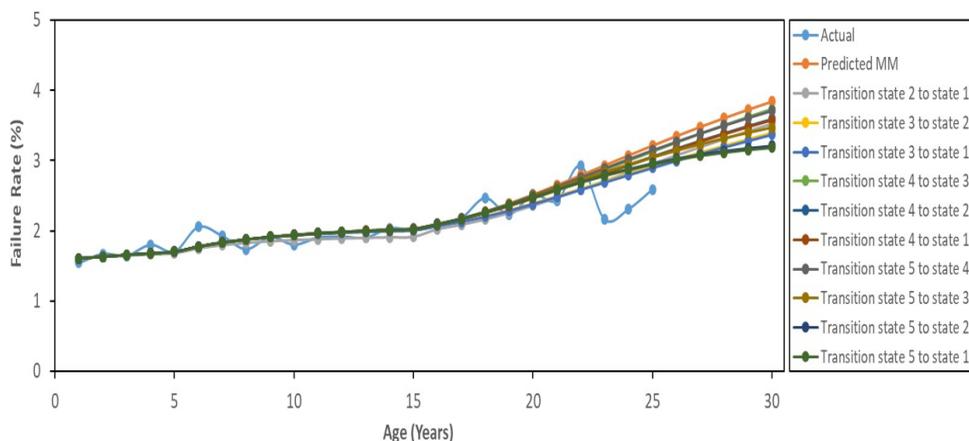


Figure 5. Failure rate performance curves with a 10% pre-determined maintenance repair rate performed at different states

The same adjustment of transition matrices as Table 2 was performed for maintenance repair rates of 20% and 30%. The failure rate performance curve and distribution of the transformer population were determined using the same methodology. The failure rate state distributions and trends of the failure rate performance curves are similar for 10%, 20%, and 30% pre-determined maintenance repair rates. For all rates, the transition from state 2 to state 1 significantly affects the failure rate state distribution and results in the highest failure rate reduction in the performance curve. The comparison of the failure rate performance curves for various pre-determined maintenance repair rates is shown in Figure 6. The pre-determined maintenance repair rate of 30% has the highest impact on the failure rate state distribution of the transformer population.

According to the case study, it is found that updating transition matrices with selected pre-determined maintenance repair rates at various states could reduce the failure rate. The comparison of average failure rate reduction with 10%, 20% and 30% pre-determined maintenance repair rates performed at all states for 30 years can be seen in Figure 7. The transition from state 2 to state 1 with 10%, 20% and 30% pre-determined maintenance repair rates have the highest impact on the average failure rate reductions with percentages of 4.77%, 8.47% and 11.47%, respectively. It is followed by a transition from state 3 to

state 1 with percentages of 4.00%, 6.75% and 8.24%, respectively. The lowest impact on the average failure rate reductions with 10 %, 20% and 30% pre-determined maintenance repair rates occur at the transition from state 5 to state 4 with percentages of 0.79%, 1.42% and 1.92%, respectively. It is followed by a transition from state 4 to state 3 with percentages of 0.99%, 1.98% and 2.95%, respectively. It is observed that the changes for each failure rate state distribution are influenced not only by the percentage of the pre-determined maintenance repair rate but also by the transition of failure rate state improvement in the MM.

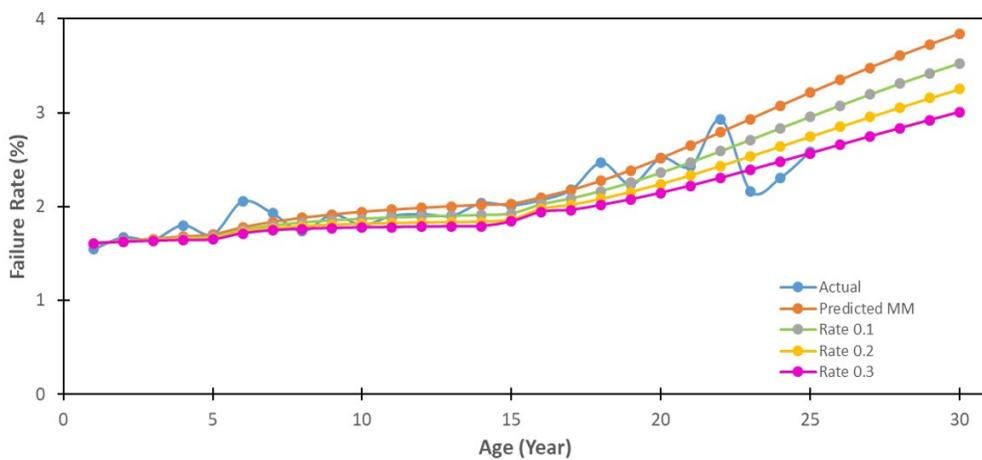


Figure 6. Failure rate performance curves with 10%, 20% and 30% pre-determined maintenance repair rates

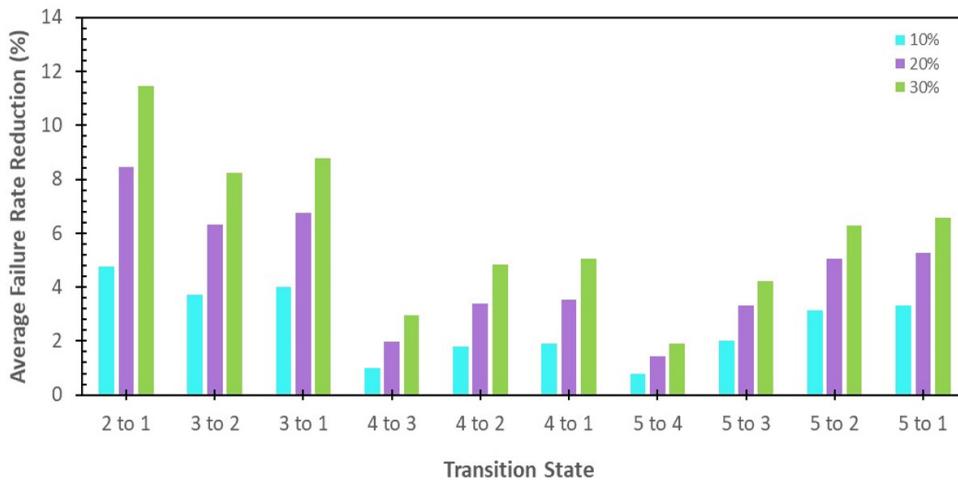
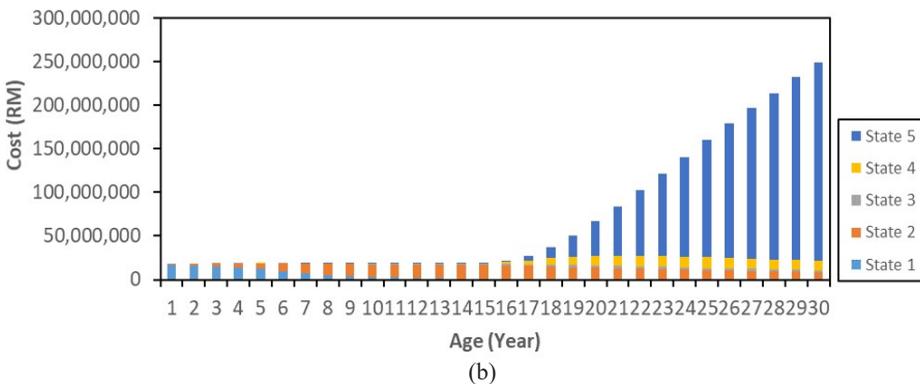
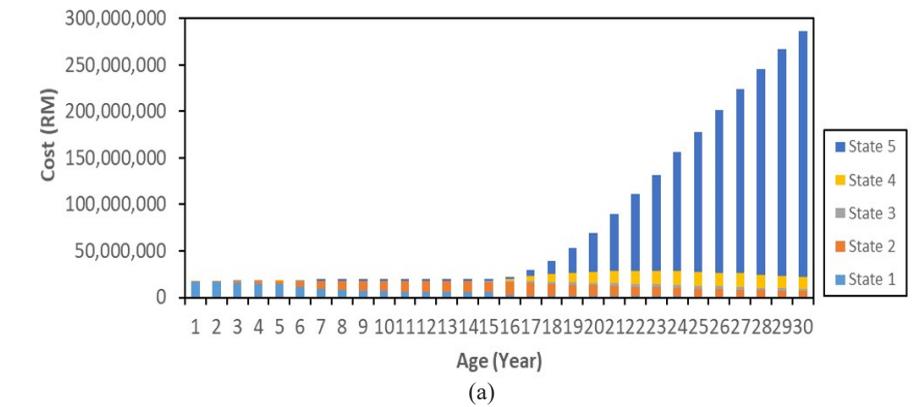


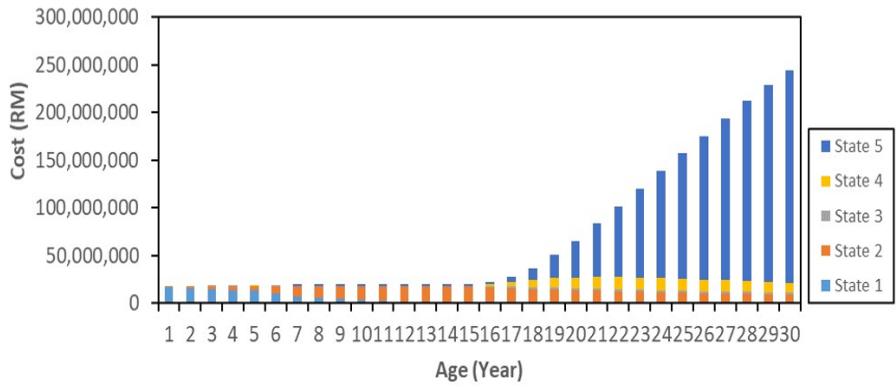
Figure 7. Comparison of failure rate reduction for all states

Maintenance Policy and Cost Study

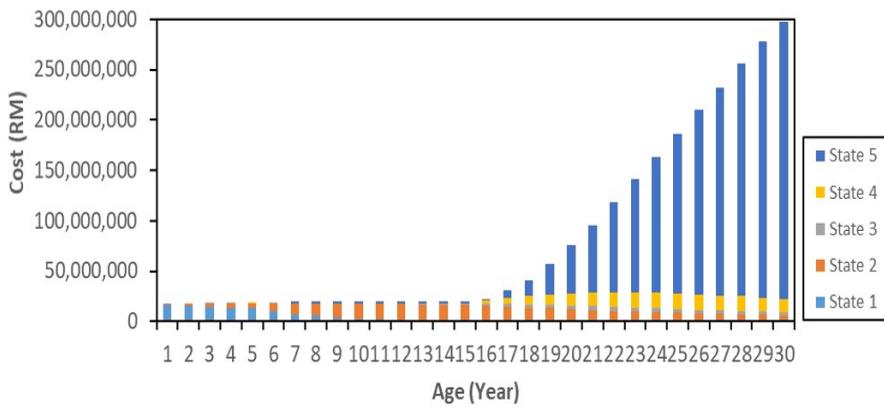
The analysis of maintenance costs was performed using the transition matrices computed from the generated MPM, as shown in Equations 12 and 13. In addition, as indicated in Equation 8, the future state distributions were updated depending on the maintenance policy model and the combined effect of deterioration. Finally, the cost of maintenance for each of the state distributions was plotted using Equation 9 and the cost estimation defined by Yahaya et al. (2018b).

The 10% pre-determined maintenance policy was assigned to each state to simulate the impact of the maintenance cost of the transformer. The distribution of transformer maintenance costs by the state is shown in Figure 8. The costs in states 4 and 5 contribute the most to the total maintenance cost. Over the 30 years, there has been no significant increment in the maintenance costs for states 1, 2, and 3. However, it is observed that the estimated costs for the transformer gradually increase, and there is a significant increment in the cost in state 5 once the age reaches 17 years. Furthermore, the total maintenance costs rise by an average of RM 18.32 million to RM 251.87 million.

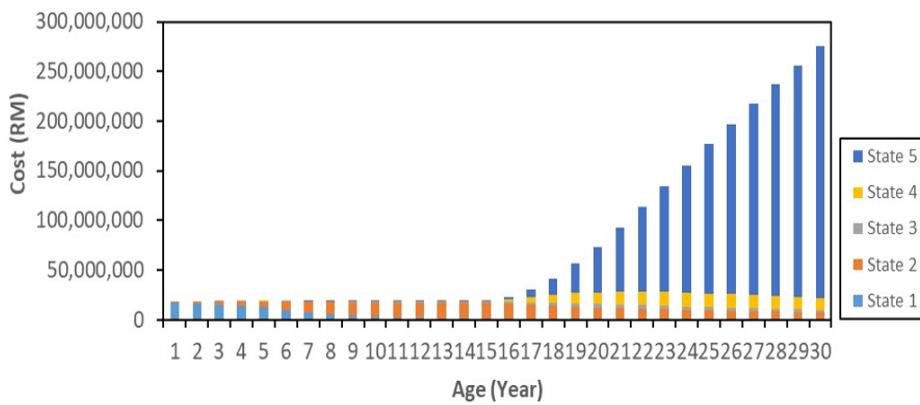




(c)

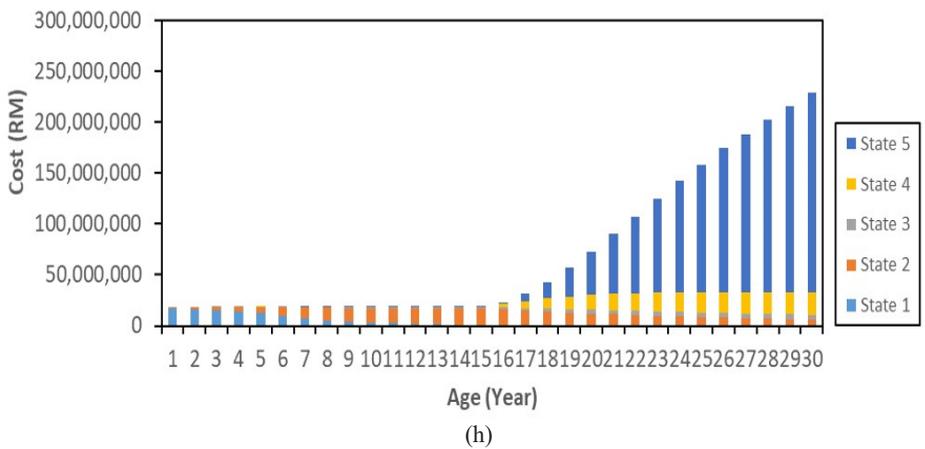
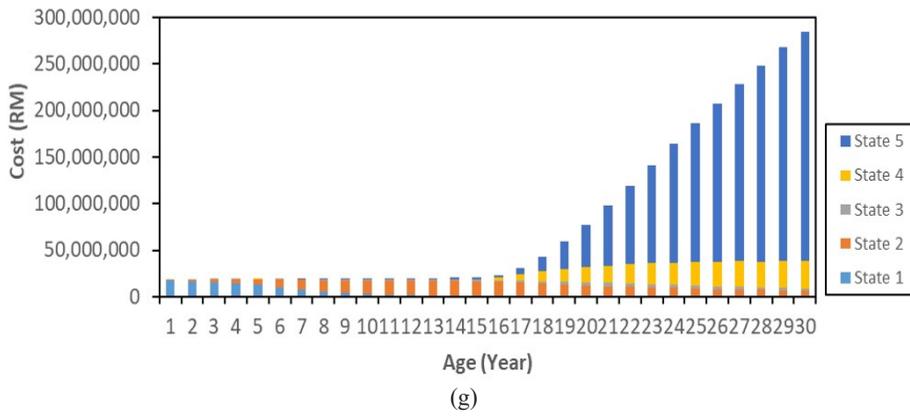
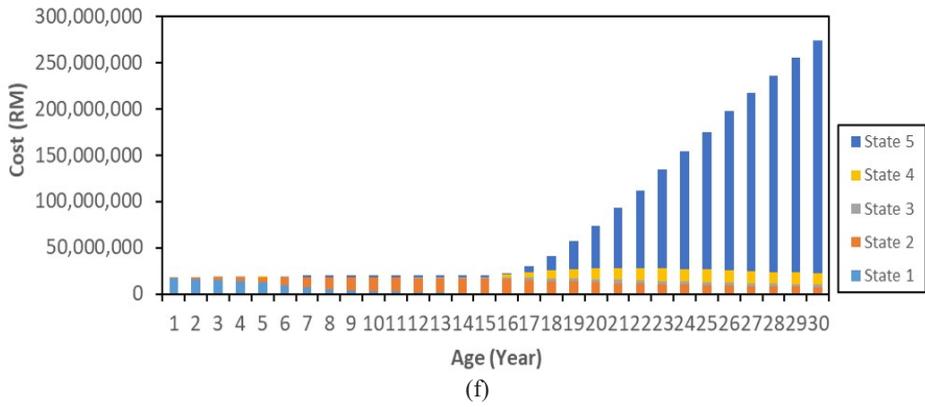


(d)



(e)

Transformer Population Failure Rate Study Based on Markov Model



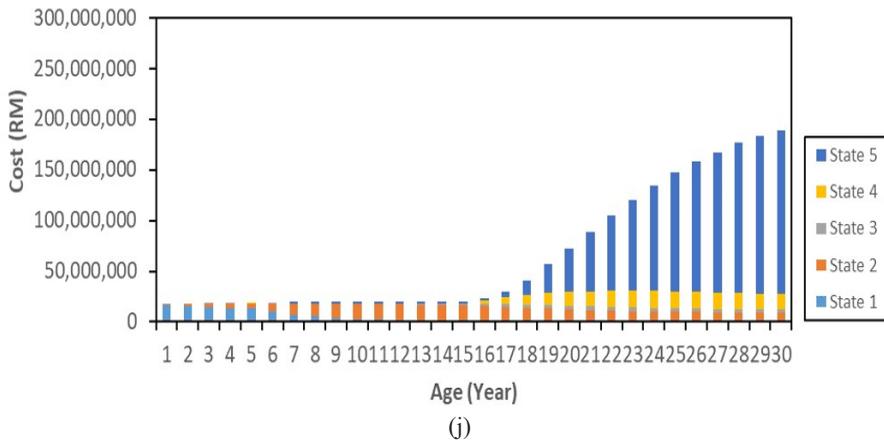
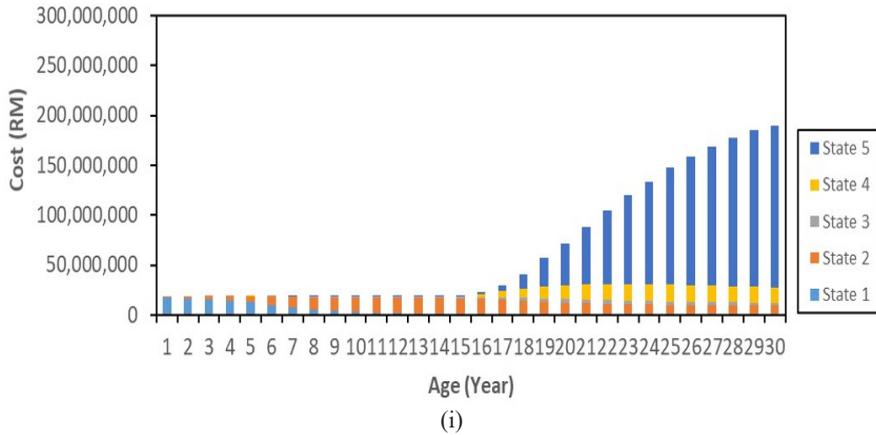


Figure 8. The transformer maintenance cost distribution with a 10% pre-determined maintenance policy calibrated at (a) state 2 to state 1; (b) state 3 to state 2; (c) state 3 to state 1; (d) state 4 to state 3; (e) state 4 to state 2; (f) state 4 to state 1; (g) state 5 to state 4; (h) state 5 to state 3; (i) state 5 to state 2; (j) state 5 to state 1

Maintenance Policy and Preventive Maintenance Frequency

The maintenance frequency analysis was performed using the transition matrices obtained from the generated MPM, as illustrated in Equations 12 and 13. The future state distributions are updated based on the maintenance policy model and the combined effect of deterioration, as shown in Equation 11. The frequency of maintenance was calculated for each state distribution using Equation 11 and the frequency of maintenance test as specified in Table 1.

The distribution of transformer population in each maintenance frequency over the 30 years of prediction is presented in Figure 9. The maintenance frequency for every nine months contributes the highest portion of the maintenance frequency. The maintenance

frequency for every six months contributes a lesser portion of the maintenance frequency than every nine months. The maintenance frequency for every three months contributes the lowest portion of the maintenance frequency. In practice, the maintenance frequency for every three months is quite difficult. A comprehensive examination and expert judgment are required to obtain necessary information, such as transformer reliability requirements. Nonetheless, this study can help utilities plan for maintenance by adjusting the maintenance policy and frequency of maintenance to optimize the asset management of the transformer population.

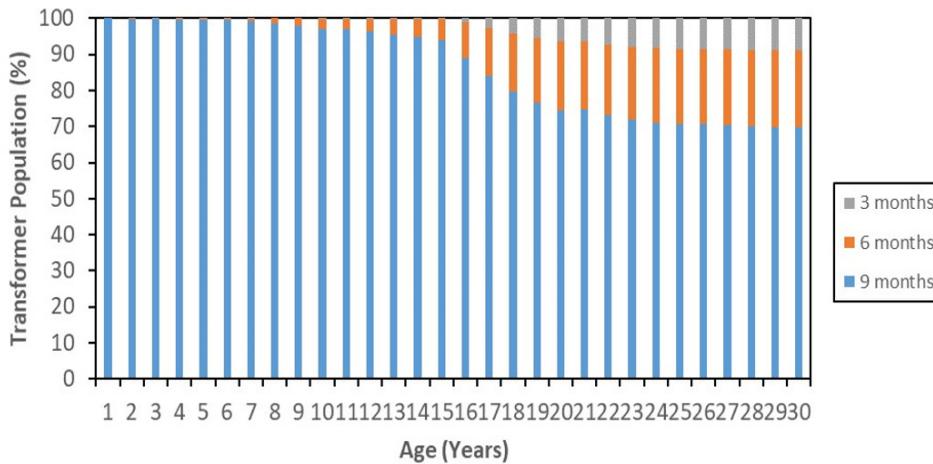


Figure 9. Distribution of transformer population according to maintenance frequency

The estimation of the maintenance frequency by months for 370 transformers is shown in Figure 10. It is observed that 316 transformers would be required to be maintained every nine months, 41 transformers every six months and 13 transformers every three months. The maintenance frequency for every nine months contributes the highest portion of the maintenance frequency since 85% of the transformer population is in states 1 and 2.

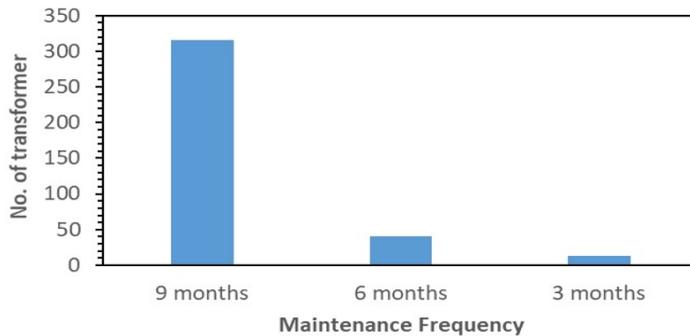


Figure 10. Estimated maintenance frequency for the transformer

CONCLUSION

It is found that MM can be used to aid utilities in planning and to implement maintenance strategies. Pre-determined maintenance repair rate and policy model is considered one of the maintenance strategies that can be carried out by updating the MM. The effect of various pre-determined maintenance repair rates and maintenance cost studies on the failure rate based on MM are investigated in this study. It is found that the updated transition matrices with selected pre-determined maintenance repair rates at various states can reduce the failure rate. The results of this study show that the adjustment of transition matrices from state 2 to state 1 has the highest impact on the failure rate reduction. The pre-determined maintenance repair rate of 30% has the highest impact on the transformer population's failure rate state distribution and failure rate performance curve. Over the 30 years of forecasting, it is observed that the estimated cost for the transformer population gradually increases, and there is an apparent increment of costs in state 5 once the age reaches 17 years. Moreover, the total maintenance costs are expected to rise by an average of RM 18.32 million to RM 251.87 million. Most transformers require maintenance every nine months to prevent the transformer from degrading to state 5, which is very poor. Overall, the findings of this study could help utilities plan and optimize their asset investments.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to the Ministry of Higher Education for the funding provided for this study under the FRGS scheme of FRGS/1/2019/TK07/UPM/02/3 (03-01-19-2071FR).

REFERENCES

- Borovkov, K. (2003). Markov chains. In *Elements of Stochastic Modelling* (pp. 75-128). World Scientific. https://doi.org/10.1142/9789812779199_0003
- Brown, R. E., Member, S., Frimpong, G., Member, S., & Willis, H. L. (2018). Failure rate modeling using equipment inspection data. *IEEE Transactions on Power Systems*, 19(2), 782-787. <https://doi.org/10.1109/TPWRS.2004.825824>
- Camahan, J. V., Davis, W. J., Shahin, M. Y., Keane, P. L., & Wu, M. I. (1987). Optimal maintenance decisions for pavement management. *Journal of Transportation Engineering*, 113(5), 554-572. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1987\)113:5\(554\)](https://doi.org/10.1061/(ASCE)0733-947X(1987)113:5(554))
- Carnero, M. C., & Gómez, A. (2017). Maintenance strategy selection in electric power distribution systems. *Energy*, 129, 255-272. <https://doi.org/10.1016/j.energy.2017.04.100>
- Cesare, M. A., & Al, E. (1992). Modelling bridge deterioration using Markov Chains. *Journal of Transportation Engineering*, 118(6), 820-833. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1992\)118](https://doi.org/10.1061/(ASCE)0733-947X(1992)118)

- Ghazali, Y. Z. Y., Talib, M. A., & Soosai, A. M. (2015, June 15-18). TNB approach on managing asset retirement for distribution transformers. In *23rd International Conference on Electricity Distribution* (pp. 1-5). Lyon, France.
- González-Domínguez, J., Sánchez-Barroso, G., & García-Sanz-Calcedo, J. (2020). Scheduling of preventive maintenance in healthcare buildings using Markov chain. *Applied Sciences*, *10*(15), Article 5263. <https://doi.org/10.3390/APP10155263>
- Hamoud, G. A., & Yiu, C. (2020). Assessment of spare parts for system components using a Markov model. *IEEE Transactions on Power Systems*, *35*(4), 3114-3121. <https://doi.org/10.1109/TPWRS.2020.2966913>
- Hoskins, R. P., Strbac, G., & Brint, A. T. (1999). Modelling the degradation of condition indices. *IEEE Proceedings: Generation, Transmission and Distribution*, *146*(4), 386-392. <https://doi.org/10.1049/ip-gtd:19990063>
- Islam, M. M., Lee, G., & Hettiwatte, S. N. (2017). Application of a general regression neural network for health index calculation of power transformers. *International Journal of Electrical Power and Energy Systems*, *93*, 308-315. <https://doi.org/10.1016/j.ijepes.2017.06.008>
- Jahromi, A., Piercy, R., Cress, S., Service, J., & Fan, W. (2009). An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine*, *25*(2), 20-34. <https://doi.org/10.1109/MEI.2009.4802595>
- Jürgensen, J. H. (2016). *Condition-based failure rate modelling for individual components in the power system* (Licentiate dissertation). KTH Royal Institute of Technology, Sweden. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A931186&dswid=1085>
- Jürgensen, J. H. (2018). *Individual failure rate modelling and exploratory failure data analysis for power system components* (Ph.D. dissertation). KTH Royal Institute of Technology, Sweden. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1251662&dswid=-8289>
- Jurgensen, J. H., Brodersson, A. L., Nordstrom, L., & Hilber, P. (2018). Impact assessment of remote control and preventive maintenance on the failure rate of a disconnector population. *IEEE Transactions on Power Delivery*, *33*(4), 1501-1509. <https://doi.org/10.1109/TPWRD.2017.2710482>
- Jürgensen, J. H., Nordstrom, L., & Hilber, P. (2016a). A review and discussion of failure rate heterogeneity in power system reliability assessment. In *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (pp. 1-8). IEEE Publishing. <https://doi.org/https://doi.org/10.1109/PMAPS.2016.7764078>
- Jürgensen, J. H., Nordström, L., & Hilber, P. (2016b). Individual failure rates for transformers within a population based on diagnostic measures. *Electric Power Systems Research*, *141*, 354-362. <https://doi.org/10.1016/j.epsr.2016.08.015>
- Li, L., Li, F., Chen, Z., & Sun, L. (2016). Use of Markov chain model based on actual repair status to predict bridge deterioration in Shanghai, China. *Transportation Research Record*, *2550*, 106-114. <https://doi.org/10.3141/2550-14>
- Lindquist, T. M., Bertling, L., & Eriksson, R. (2005). Estimation of disconnector contact condition for modelling the effect of maintenance and ageing. In *2005 IEEE Russia Power Tech* (pp. 1-7). IEEE Publishing. <https://doi.org/10.1109/PTC.2005.4524406>

- McDonald, D. (2004). *Elements of applied probability for engineering, mathematics and systems science*. World Scientific. <https://doi.org/10.1142/5456>
- Micevski, T., Kuczera, G., & Coombes, P. (2002). Markov model for storm water pipe deterioration. *Journal of Infrastructure Systems*, 8(2), 49-56. [https://doi.org/10.1061/\(asce\)1076-0342\(2002\)8:2\(49\)](https://doi.org/10.1061/(asce)1076-0342(2002)8:2(49))
- Naderian, A., Cress, S., Piercy, R., Wang, F., & Service, J. (2008). An approach to determine the health index of power transformers. In *Conference Record of IEEE International Symposium on Electrical Insulation* (pp. 192-196). IEEE Publishing. <https://doi.org/10.1109/ELINSL.2008.4570308>
- Riveros, G. A., & Arredondo, E. (2010). *Guide to the development of a deterioration rate curve using condition state inspection data*. US Army Corps of Engineering. <https://apps.dtic.mil/sti/pdfs/ADA527894.pdf>
- Selva, A. M., Azis, N., Yahaya, M. S., Ab Kadir, M. Z. A., Jasni, J., Ghazali, Y. Z. Y., & Talib, M. A. (2018). Application of markov model to estimate individual condition parameters for transformers. *Energies*, 11(8), Article 2114. <https://doi.org/10.3390/en11082114>
- Shariffuddin, N. S., Azis, N., Selva, A. M., Yahaya, M. S., Jasni, J., & Talib, M. A. (2021a). Investigation on the relationship between failure rates and health index of distribution transformer population. In *Proceedings of the IEEE International Conference on Properties and Applications of Dielectric Materials* (pp. 119-122). IEEE Publishing. <https://doi.org/10.1109/ICPADM49635.2021.9493977>
- Shariffuddin, N. S., Azis, N., Selva, A. M., Yahaya, M. S., Jasni, J., Kadir, M. Z. A. A., & Talib, A. M. (2021b). *Failure Rate Estimation for Transformer Population based on Health Index through Markov Model Approach*. 29(4), 3029-3042. <https://doi.org/10.47836/pjst.29.4.42>
- Stevens, N. A., Lydon, M., Marshall, A. H., & Taylor, S. (2020). Identification of bridge key performance indicators using survival analysis for future network-wide structural health monitoring. *Sensors (Switzerland)*, 20(23), 1-15. <https://doi.org/10.3390/s20236894>
- Tenbohlen, S., Vahidi, F., Gebauer, J., & Gmbh, M. R. (2011, August 22-26). Assessment of power transformer reliability. In *17th International Symposium on High Voltage Engineering* (pp. 1-6). Hannover, Germany.
- White, J. R., & Widup, R. (2014). Factors to consider when determining maintenance intervals. *IEEE Transactions on Industry Applications*, 50(1), 188-194. <https://doi.org/10.1109/TIA.2013.2288229>
- Yahaya, M. S., Azis, N., Kadir, M. Z. A. A., Jasni, J., Hairi, M. H., & Talib, M. A. (2017). Estimation of transformers health index based on the markov chain. *Energies*, 10(11), 1-11. <https://doi.org/10.3390/en10111824>
- Yahaya, M. S., Azis, N., Selva, A. M., Kadir, M. Z. A. A., Jasni, J., Hairi, M. H., Ghazali, Y. Z. Y., & Talib, M. A. (2018a). Effect of pre-determined maintenance repair rates on the health index state distribution and performance condition curve based on the Markov Prediction Model for sustainable transformers asset management strategies. *Sustainability (Switzerland)*, 10(10), Article 3399. <https://doi.org/10.3390/su10103399>
- Yahaya, M. S., Azis, N., Selva, A. M., Kadir, M. Z. A. A., Jasni, J., Kadim, E. J., Hairi, M. H., & Ghazali, Y. Z. Y. (2018b). A maintenance cost study of transformers based on markov model utilizing frequency of transition approach. *Energies*, 11(8), Article 2006. <https://doi.org/10.3390/en11082006>

MOFTI: Mobile Self-Service Framework to Uplift Customer Experience for the Telecommunications Industry in Malaysia

Zainab Abu Bakar¹, Hazrina Sofian² and Nazean Jomhari^{2*}

¹*Department of Information Systems, Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*

²*Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*

ABSTRACT

The new era of digital technologies has caused customers to require more efficient, easier access and immediate services. The first concern in the telecommunications industry is unpleasant conventional interactions (live agents) and interactive voice responses. The second area of concern is the operational change due to the MCO, which has impacted customer experience at physical counters. MCO imposes physical distancing at physical counters, with customers required to be at least one metre apart. Some businesses have imposed long queues to avoid congestion within a limited space, forcing customers to queue outside business spaces: perhaps along the road, in front of other premises, and under uncertain weather conditions. The shortened working hours during the MCO have amplified the number of unresolved issues among customers and long queues. This research aims to formulate a proposed framework for uplifting customer experience while managing the restrictions people face due to the MCO. The proposed solution is based on a literature review, reviewing existing mobile apps, and analysing data collected in 2020 through surveys and interviews. The evaluation results show that the proposed solution can be a guide for telecommunications service providers to enhance the customer's experience of their self-service platforms while managing the restrictions people face due to the MCO.

The novelty of this research is aligned with the Industrial Revolution 4.0 to reduce the workforce through the automation of business processes to meet the demands of the future economy.

Keywords: Automation of business process, customer experience, customer loyalty, digital contact centre, human restriction, telecommunications customer services

ARTICLE INFO

Article history:

Received: 6 December 2021

Accepted: 29 March 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.32>

E-mail addresses:

wqc180008@siswa.um.edu.my (Zainab Abu Bakar)

hazrina@um.edu.my (Hazrina Sofian)

nazean@um.edu.my (Nazean Jomhari)

*Corresponding author

INTRODUCTION

According to the Industry Performance Report (IPR) produced by the Malaysian Communications and Multimedia Commission (MCMC), broadband subscriptions in Malaysia almost doubled over a recent five-year period to reach 39.4 million in 2018. With the 4G and 5G network expansions in December 2018, the demand for continuous connectivity and mobile technology is growing among Malaysians. Statistics show that mobile broadband subscriptions grew by 4.3% from 2016 to 2017. In those years, network coverage improvements and attractive pricing plans from service providers were growth factors. Since then, mobile devices have continuously made consumers' lifestyles more convenient through e-commerce applications, navigation applications, entertainment applications, and many more. Furthermore, with the implementation of the Internet of Things (IoT), big data, and artificial intelligence (AI), digital services are evolving and impacting everyone.

Moreover, digital advertising expenditure (ADEX), which provides widespread advertising coverage through thousands of digital publishers, is becoming more well-known because of its cost-effectiveness, global reach, increased audience size, and multi-platform use (IPR, 2018). IPR (2018) also mentioned that the smartphone market share is growing due to strong customer demand, aggressive marketing by vendors, and a wide selection of devices available at various prices worldwide. The annual MCMC survey in 2018 indicates that smartphones remain the most common device with which to access the internet (93.1%), compared to laptops (44.2%), desktops 28.1%, and tablets (20.4%).

Irfan (2016) mentioned that mobile phone services in the telecommunications industry are growing, with more than 1.7 billion subscribers worldwide and about 80% of the global population as potential targets. It shows that users are more attached to the phone than other devices. Telecommunications services continue to be the main revenue makers, while digital platforms are becoming new sources of revenue for the telecommunications industry. Telecommunications service providers in Malaysia, such as TM, TIME, Celcom, Maxis, and Digi, are focusing on retaining and making additional revenue from existing subscribers to meet the demands of the future economy. Additional revenue can be made by moving beyond their traditional business through the automation of their business processes. One way to automate business processes is to find ways for customers, especially digital natives, to help themselves through online self-service technology (SST). Online SST can take the form of web-based or mobile applications. Since the mobile SST is the focus of this research, this paper showcases the mobile devices and online SST provided by telecommunications service providers in Malaysia.

Currently, service providers in Malaysia have equipped mobile devices and online SST with various customer features, such as the capability to sign-up, check account status, check for updates, obtain technical assistance, perform billing-related transactions, and

receive the latest promotions on a single platform. According to Salomann et al. (2006), self-service can be clustered into two central themes, customers and technology. Lu et al. (2015) pointed out that with the development of customer capability in technology and mobile self-service, customers have personalized service experience based on their specific knowledge and skills.

Zhou et al. (2010) suggested using Platform-as-a-Service (PaaS), whereby services are more targeted, focused, and personalized to maintain customer loyalty and increase an organization's revenue. On the other hand, Riahi and Ristock (2014) proposed conventional interaction, whereby a contact centre provides a centralized customer service handled by live agents who, for example, answer telephone calls, respond to emails, and conduct live chats, and make outbound calls.

Interactive voice response (IVR) technology is another type of conventional interaction that can facilitate the tasks of live agents. IVR can also assist customer service agents by providing customer solutions, repeating announcements, and routing to the main menu selection if the customer selects the wrong option. In addition, IVR shortens the time taken for customers to obtain solutions by reducing the burden on the customer service agent. For example, IVR obtains customer information such as the account number before control is handed to a live agent. Although these advantages show that IVR can assist live agents, it has also been indicated that IVR is not a sufficient contact centre solution for most customer concerns.

Flaws also exist with live agents and IVR in that customers and live agents have to endure several. For example, they must (1) undergo repetitive processes; (2) spend time calling and waiting; and (3) face inadequate, incorrect, and/or inconsistent information. In addition, al-Mashraie et al. (2020) mentioned other drawbacks, such as call disconnection and the frequency with which customers have to obtain solutions, which can impact the customer's experience of the customer services.

According to Fang et al. (2020), implementing a lockdown policy effectively reduces the spread of the COVID-19 virus due to population flows. At the early stage of the MCO implementation, all Malaysians were ordered to stay at home, except for essential workers in the health services and those employed by water, electricity, telecommunications, and food supply companies. The government also shortened business hours, imposed physical distancing, and promoted workplace hygiene. Telecommunications service providers have complied with these orders in different ways. For instance, Maxis has limited the number of customers allowed to enter its premises at one time and practises strict social distancing.

Meanwhile, TM has reduced its work schedule to operate only three days a week, with a maximum of three staff and five customers at a time. Furthermore, TM only performs installation and restoration for appointments made before the MCO, except in the Enhanced Movement Control Order (EMCO) areas. This situation impacts the customer's

experience as it leads to frustration and dissatisfaction with the inefficiency of the service since connectivity is critical during the MCO. Furthermore, the situation has worsened because several services can only be performed at service centres. These include service subscriptions (in addition to the online subscriptions channel at www.unifi.com.my), account terminations, collections, and replacements of SIM cards for Unifi Mobile, and device collections for Unifi Air and CPE replacement.

The first concern in the telecommunications industry is the unpleasant conventional interactions with live agents and IVR. According to Rawson et al. (2013), organizations must identify the appropriate metrics and develop software that emphasizes customer experience. The second concern is the operational changes due to the MCO, which has increased the time to resolve customer complaints like network issues and call disconnections. Digitization in today's world has shaped customer services from face-to-face or voice interaction into digital interaction (Bhale & Bedi, 2021). Digital interaction has been crucial since 2020 because the COVID-19 pandemic has increased people's reliance on internet services due to the MCO. People are forced to work, study, and even shop through online platforms from home (MCMC, 2020). Realizing these areas of concern, the first objective of this research is to analyse customer preference and experience when using the mobile self-service features provided by telecommunications service providers in Malaysia while managing the restrictions people face due to the MCO. The second objective of this research is to propose a framework for overcoming the unpleasant conventional methods since there are flaws and limitations with conventional interaction (live agents) and IVR as well as with the operational changes due to the MCO, which has impacted customer experience at physical counters. The third research objective is to evaluate Malaysia's proposed self-service framework providers. Subsequently, the proposed framework can be used as a guideline for the software developers to enhance or develop a new self-service mobile app for telecommunication industries in Malaysia.

RELATED WORK

People rely on technology as a driving force for advancing economic systems and their quality of life. The COVID-19 pandemic has enhanced the demand for mobile app services due to their ease of use, privacy, security, usefulness, previous experience, technology competency, lifestyle, trust, and credibility (Salam et al., 2021). In this study, a literature review is carried out to review the existing research on mobile self-service and customer experience. Second, this research reviews existing mobile applications, specifically those used by the telecommunications industry in Malaysia. Current researchers on mobile self-service and customer experience have highlighted four main issues: customer preference, service quality, perspectives of self-service, and the impact on operational change due to the MCO.

Customer Preference

According to Al-Mashraie et al. (2020), telecommunications service providers secure customer loyalty and attract new customers by increasing their market share, offering a variety of plans, bundling services to meet customer demands, and providing discounts. To stay competitive in the market, service providers must understand their customer's behaviours and preferences before they can subsequently provide a tailored service for each customer. Xu et al. (2015) mentioned that the growth of smartphones and mobile applications had captured the attention of businesses. However, businesses must understand user behaviours regarding their initial acceptance and continued use of mobile applications. Wang et al. (2019) described the adoption of inertia in marketing. Inertia is a physical term that describes the tendency of a physical object to resist change. In marketing, consumer inertia is some customers' tendency to continue using the same product when superior options exist. According to the Forbes website, marketing inertia can be driven by customer habits, brand loyalty, switching cost and many more. One of the reasons why a consumer continues using the same service can be the inertia strategy taken by the company in setting prices after analysing their competitors. Thus, service providers can increase levels of inertia to sustain long-term positive relationships with consumers.

Service Quality

Al-Mashraie et al. (2020) stated that service quality is one of the most influential factors affecting customers. The quality of a mobile service can influence customers to continue using a particular service. Hence, customers usually evaluate the service quality to decide if it is worth continuing after the trial (Wang et al., 2019) to continuously improve its mobile app service quality to sustain market growth. Improving the store service quality is also important in making it comparable with the quality of the other major applications' stores to convince the customers. Businesses need to emphasize the benefits, quality, aesthetic appeal, and enjoyment these mobile applications can deliver to encourage customers to use them during their marketing campaigns (Xu et al., 2020).

However, according to Riahi and Ristock (2014), conventional interaction through live agents or IVR is not a sufficient contact centre solution for most customer concerns. Both types of interaction have inherent limitations, such as repetition in the call handling process and time spent (calling and waiting). At the same time, the information provided is often inadequate, incorrect, and/or inconsistent. IVR also takes a long time to reach a live agent, and the process is longer if the customer selects the wrong option. Al-Mashraie et al. (2020) mentioned that conventional interaction has issues such as call disconnections and the length of time spent over the phone, which impact customer experience. In addition, it may mean customers tend to change to a different telecommunications service provider.

Perspectives of Self-Service

Beatson et al. (2006) suggested the key attributes for SST that need to align with a personal service. Identifying service attributes, such as the ease of using SST facilities, saving time, low risk, and customization, allows service managers to focus on service performance drivers to provide the best customer service. Hence, customer service staff can train customers to use the available SST. By providing this training, customers will consider the customer service staff agent's burden and simultaneously improve the customer's satisfaction and commitment to solving the issues independently.

Salomann et al. (2006) stated that self-service is important for customer relationships. The acceptance factors related to SST are ease of use, usefulness, risk, and the need for interaction with personnel. These factors increase customer satisfaction, word of mouth, and repeat purchase intentions towards the services while avoiding complaints. However, a lack of human interaction may become a challenge for companies that require high levels of attention and integration with persona design (i.e., fictional characters). When designing an SST, the listed characteristics include a name, age, gender, job title, accent, and sense of humour to create an imaginary person.

Barua et al. (2018) interpreted the perceived reliability of SST as having significant effects on perceived risk, technology trust, and customer satisfaction. The two important predictors under perceived reliability are perceived security and perceived control. When designing and implementing SST, these must be considered to moderate the risk, satisfy existing users, and attract new customers. Meanwhile, Yang et al. (2014) reviewed perceived usefulness (PU) and perceived ease of use (PEU) in developing SST for mobile devices. The study concluded that PEU is more important than PU in driving the adoption of new SST. However, PU is more important than PEU when mobile services are those of a new brand.

Impact on Operational Change

The COVID-19 outbreak has caused changes in the operating hours of businesses in Malaysia. However, according to Fang et al. (2020), the imposition of an extensive lockdown policy effectively reduces the spread of the virus due to population flows. In addition, restrictions on human mobility and the enhancement of social distancing policies have significantly reduced the infection rate.

In previous research, Aziz et al. (2020) mentioned that at the early stage of the MCO implementation, all Malaysians were instructed primarily to stay at home and prohibited from attending large gatherings. All facilities were closed except primary and essential services such as health, water, electricity, telecommunications, and food supply companies. In phase 4 of the MCO, the government allowed certain businesses to open and the movement of people if mandatory standard operating procedures were followed, including movement restrictions and social distancing, while no mass gatherings were allowed.

Table 1 describes the focus of this research, mobile self-service and customer experience while highlighting four main issues: customer preference, service quality, perspectives of self-service, and the impact on operational change due to the MCO.

Table 1
The focus of research on mobile self-service and customer experience

The focus of the research	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Service quality	/		/	/					/					
Customer experience	/	/	/		/	/	/	/		/	/	/		
Self-service perspective						/	/			/	/	/		
Impact on operation change due to MCO													/	/

Note:
R1 refers to the article by Xu et al. (2015)
R2 refers to the article by Al-Mashraie et al. (2020)
R3 refers to the article by Wang et al. (2019)
R4 refers to the article by Wang et al. (2016)
R5 refers to the article by Riahi and Ristock (2014)
R6 refers to the article by De Leon et al. (2020)
R7 refers to article by Salomann et al. (2006)
R8 refers to the article by Rawson et al. (2013)
R9 refers to the article by Basole and Karla (2012)
R10 refers to the article by Beatson et al. (2006)
R11 refers to the article by Barua et al. (2018)
R12 refers to the article by Yang et al. (2014)
R13 refers to an article by Aziz et al. (2020)
R14 refers to the article by Fang et al. (2020)

Table 1 describes how, in recent years, most researchers have focused on enhancing app services, improving customer experience, and enabling SST for mobile apps. However, with the commencement of the COVID-19 pandemic, researchers drastically shifted their focus onto enabling mobile apps to accommodate human restrictions due to the MCO.

Physical distancing, whereby customers need to be 1 metre apart, has become an issue, as the current physical business spaces are insufficient in size. Some businesses have limited their number of daily customers, leading to unresolved issues among customers. Meanwhile, some businesses have imposed low queues, whereby customers must queue outside the business spaces: perhaps along the road or in front of other premises, and uncertain weather. The MCO has also meant limits on working hours, amplified the number of unresolved issues and long queues.

Aziz et al. (2020) and Fang et al. (2020) focused merely on human restrictions due to COVID-19. However, as seen in the research trends since 2012, researchers have focused

on enhancing app services or enabling SST. They have mostly incorporated customer experience into their research. It is to ensure the uplifting of the user experience. Thus, this research aims to focus on how mobile apps have been enabled to accommodate human restrictions due to the MCO while uplifting the user experience. Table 2 compares Live Agent, IVR, Physical Counter and SST.

Table 2
The comparison between live agent, IVR, physical counter and SST

Platform	Strength	Weakness
Conventional interactions (live agents)	1) Offer flexibility and perform human interaction	1) Undergo repetitive processes 2) Spend time calling and waiting 3) Face inadequate, incorrect, and/or conflicting 4) Call disconnected
IVR	1) Facilitate the tasks of live agents, for instance, providing solutions to customers, repeating announcements, and routing to the main menu selection if the customer selects the wrong option. 2) IVR shortens the time taken for customers to obtain solutions by reducing the burden on the customer service agent. For example, IVR obtains customer information such as the account number before control is handed to a live agent.	1) Undergo repetitive processes 2) Spend time calling and waiting 3) Face inadequate, incorrect, and/or conflicting information. 4) Call disconnected
Physical counters (e.g., branches, outlets or kiosks)	1) Offer flexibility and perform human interaction	1) MCO rules imposed physical distancing at physical counters, such as long queues, shortened working hours and promoted workplace hygiene 2) Transactions are limited. Self-service kiosks are pre-programmed to execute commands only to a certain extent. Complicated transactions are usually not supported

Table 2 (Continue)

Platform	Strength	Weakness
SST (e.g., Apps or Portal)	1) Made consumers' lifestyles more convenient through e-commerce applications, navigation applications, entertainment applications, and many more.	1) Lack of human interaction
	2) Provides widespread advertising coverage through thousands of digital publishers, cost-effectiveness, global reach, increased audience size, and multi-platform use.	2) Equipment malfunctions
	3) Users are more attached to the phone than to other devices.	3) High up-front costs
	4) Additional revenue can be made by moving beyond their traditional business through the automation of their business processes.	
	5) Equipped mobile devices and online SST with various features for customers, such as the capability to sign-up, check account status, check for updates, obtain technical assistance, perform billing-related transactions, and receive the latest promotions on a single platform	
	6) With the development of customer capability in terms of technology and mobile self-service, a customer has more opportunities to receive a personalized service experiences based on their specific knowledge and skills	
	7) Services are more targeted, focused, and personalized to maintain customer loyalty and increase an organization's revenue.	
	8) Ease of using SST facilities, saving time, low risk, and customization, allows service managers to focus on service performance drivers to provide the best customer service.	
	9) SST can reduce the live agent's burden and, at the same time, improve the customer's satisfaction and their commitment to solving the issues on their own.	
	10) Reduce the workforce through the automation of business processes to meet the demands of the future economy	

Existing Deployment of Mobile Apps in Malaysia

This section reviews four mobile apps used in the telecommunications industry in Malaysia. The selection was based on the app ratings given by customers. The four apps chosen were the Celcom Life App, MyUnifi App, Maxis App, and TIME App.

The Celcom Life App, by Celcom, was created exclusively to help customers to check and manage Postpaid and Celcom Xpax accounts. If using other Internet connections, customers can request a One Time PIN (OTP) for their mobile device. In addition, customers can download the latest app from the Google Play Store (Android), the Apple App Store (iOS) and the Huawei App Gallery. Celcom Life is free to download and use without incurring any internet charges, and it was specially designed for Celcom customers.

The MyUnifi app is an app for Unifi customers that was created to consolidate all the Unifi services. Customers can subscribe to Unifi services and manage their accounts in one app. In addition, the MyUnifi app is free for everyone and available for download via the Google Play Store and the Apple App Store. Customers can register to enjoy the features via a protected profile with personal information, giving a hassle-free experience and peace of mind. The app also allows customers to identify and link with current Unifi accounts. However, through the guest mode, existing or prospective customers can still browse TM's latest Unifi offerings.

The Maxis app is a secured self-help app for Maxis subscribers. The Maxis app is accessible with a Maxis ID. Customers can download the app on the Apple App Store or Google Play and sign up for a Maxis ID. However, if a customer uses Google maps retrieval from the rewards module or redirects to other websites or apps from the Maxis app, they will incur standard data charges.

The TIME Internet app allows the customer to manage a TIME account with the tap of a finger. In addition to viewing accounts, customers can pay bills and simplify their life with the TIME app. Table 3 summarises the key features of the existing mobile apps.

Table 3

Summary of the key features of the existing mobile apps, according to the top two highest ratings and bottom two lowest ratings

App Name	Features	Rating	Total Downloads	Total Reviewer
Celcom Life App	<ol style="list-style-type: none"> 1. Usage overview 2. Check statements & pay bills 3. Instant reload 4. Buy internet & add-ons 5. Roaming internet add-ons 6. Self-register Celcom Xpax 	4.4	1M+	176,720

Table 3 (Continue)

App Name	Features	Rating	Total Downloads	Total Reviewer
MyUnifi App	<ol style="list-style-type: none"> 1. Subscription of Unifi services 2. Account overview & activity 3. Product offering & subscription 4. Bill payment 5. Support (Unifi community, Live Chat, self-troubleshooting) 6. Rewards 	4.1	500K+	36,415
Maxis App	<ol style="list-style-type: none"> 1. Account overview & latest offers 2. Track data usage & purchase passes 3. Bill payment, download bill 4. Subscribe to e-Bill & manage direct debit 5. Maxis TV/ Maxis home fibre WiFi 6. Rewards 	3.5	1M+	36,542
TIME App	<ol style="list-style-type: none"> 1. Account summary & info 2. View and update billing information 3. View, download and pay bills 4. Register auto debit 5. Connection status 6. Self-care 	3.4	50K+	466

METHODOLOGY

This research implemented the Design Science Research Methodology (DSRM) to demonstrate the principles, practices, and procedures for a successful artefact (Peffer et al., 2007). The DSRM process consists of six steps, whereby each activity is carried out to answer the research questions. In Table 4, the research questions are mapped, along with the corresponding research objectives and the methodology applied to achieve the objectives.

Table 4

Research questions and methodology

Research Objective	Research Question	Methodology
i. To analyse customer preference and experience when using SST compared to the conventional method (traditional platform).	1.1 What current platforms were served to telecommunication customers in Malaysia?	<ol style="list-style-type: none"> i. Survey ii. Interview iii. Review the website of a service provider in Malaysia iv. Review the annual report of a service provider in Malaysia
	1.2 Does the mobile SST able to support customer needs and requirements?	<ol style="list-style-type: none"> i. Survey ii. Interview
	1.3 What are the customer preferences when contacting the service providers?	<ol style="list-style-type: none"> i. Survey ii. Interview

Table 4 (Continue)

Research Objective	Research Question	Methodology
ii. To formulate a mobile self-service framework based on existing processes and features.	2.1 What are the services that currently cannot be done online?	i. Survey ii. Interview iii. Review the website of a service provider in Malaysia iv. Review the annual report of a service provider in Malaysia
	2.2 How the other domains can support online registration and termination?	i. Literature review
	2.3 What are the customer preferences in digitizing these offline processes?	i. Survey ii. Interview
iii. To formulate a framework and through a mobile self-service solution based on existing features: 1. The interface 2. Billing 3. Account status 4. Ad-on to purchase 5. Rewards 6. Support received 7. Others features	3.1 What are the features that should have inside a self-service mobile application?	i. Survey ii. Interview iii. Review existing mobile applications of service providers in Malaysia iv. Review the website of a service provider in Malaysia
	3.2 What are the available self-service mobile applications in telecommunication industries?	i. Survey ii. Interview iii. Review existing mobile applications of service providers in Malaysia iv. Review the website of a service provider in Malaysia
	3.3 What are the available self-service web-based applications in telecommunication industries? Can these self-service practices be converted to mobile applications?	i. Survey ii. Interview iii. Review existing mobile applications of service providers iv. Review the website of a service provider in Malaysia
	3.4 What are the solutions for all available mobile SST in telecommunication?	i. Survey ii. Interview iii. Literature review iv. Review existing mobile applications of service providers in Malaysia v. Review the website of a service provider in Malaysia
	3.5 Are there are any researchers that focusing on compiling all available self-services along with the existing practices in the telecommunication industry? How about in other industries such as health and education?	i. Survey ii. Interview iii. Literature review
iv. To evaluate the formulated framework and process model based on existing features	4.1 How to evaluate the formulated framework?	i. Survey ii. Interview

This research reviewed journals and articles published within the period 2006 to 2020. The keywords are mobile applications, customer experience, human mobility restriction, digital contact centre, and customer loyalty. Through an in-depth study of existing research, this research identified many existing self-service mobile frameworks but found that most researchers have only focused on general factors and certain service areas. In many countries, telecommunications providers used to be state-owned monopolies that controlled the market and ensured that telecommunications were available to the public. Thus, the concepts and software depend on company strategies and market demand. Besides, few research studies focus on mobile self-services that drive companies to develop their customer service systems based only on user requirements studies. This research also reviewed the existing mobile apps and their features developed by telecommunications service providers in Malaysia. The reviews and selection were based on customer app ratings, which were the top two highest and the bottom two lowest. Thus, the apps reviewed were the Celcom Life App, MyUnifi App, Maxis App and TIME App. Based on their existing features, the apps were redesigned to deliver seamless and effortless user experiences with a refined look. As a result, a new framework for mobile self-service provision is proposed as guidelines for the telecommunications industry. Both customers and companies will benefit from the new framework, which helps to improve customer experience through the features provided. In addition, it should increase the number and loyalty of customers to service providers.

In this research, data was collected through primary and secondary sources. The primary data was collected using a survey and interviews. The secondary data was collected through a review of the journals, articles, web portals, existing mobile apps, and annual reports to gain the related information.

The survey questionnaire used Google Forms to gauge the information about the respondents' demographics, customer experience and recommendations for mobile app services. In addition, the questionnaire was distributed via email and social media such as Facebook and WhatsApp. Meanwhile, interview sessions with service providers were conducted through phone calls because of the implementation of social distancing due to the COVID-19 pandemic in Malaysia. Therefore, the opinions from selected service providers, including TM, Maxis, Celcom, and TIME, were needed to support the analysis for this research.

For the secondary data, information was collected from the respective service providers, e, so the sample size had to be greater than 75. Therefore, this research set 75 as the minimum sample size.

The design phase of the mobile app framework was based on the existing mobile applications developed by the service providers and the feedback from the survey results.

The design phase of the mobile application graphic user interface was based on the formulated framework. These graphic user interfaces (GUI) were displayed to customers and service providers during the evaluation.

The proposed mobile application framework and the GUI were evaluated based on the survey questions to get respondents' opinions and feedback. The questionnaire results were analysed using a Likert Scale to determine if the new requirements met the objective of this research. The existing evaluation frameworks designed by other researchers could be used as guidelines to formulate the new framework.

ANALYSIS AND RESULTS

According to Wang et al. (2019), five core features should be included while designing a mobile self-service. Based on a comparison of mobile apps in the literature review section: Celcom Life App, MyUnifi App, Maxis App, and TIME App, not all service providers include an Exclusive Deals and Rewards feature, while only the Celcom Life App has an Account Security (2FA) feature to verify the app account. Furthermore, Celcom Life App won the Best Use of Mobile award at the Customer Experience Asia Summit in 2018. Thus, the Exclusive Deals and Rewards and Account Security (2FA) features were included in the proposed framework. This section elicited customers' requirements as part of the initiative to resolve research problems which aim to provide comprehensive SST in telecommunication industries in Malaysia while adopting the existing mobile applications.

The newly proposed framework contains processes to enhance the services or features of mobile apps to improve customer experience. This research elicited customers' preferred transaction methods to integrate the processes.

According to Figure 1, 62% of customers chose transactions through Mobile Apps as the most desired. Meanwhile, 26% chose the web portal, and 7% chose walk-in (over the counter) as the least desired. Based on the survey, 92 customers gave the following reasons to contact their service provider's support team:

- Technical assistance (65 feedback items)
- Account info and bill-related (26 feedback items)
- Inquiry (22 feedback items)
- Product offering and Promotions (11 feedback items)
- Never (2 feedback items)

Operational Changes Due to the MCO Outlets/Kiosks

The newly proposed framework contains processes to accommodate the operational changes due to the MCO, which have impacted customer experience at physical counters. In addition, this research elicited the requirements of customers facing difficulties in performing transactions to formulate a mobile self-service framework.

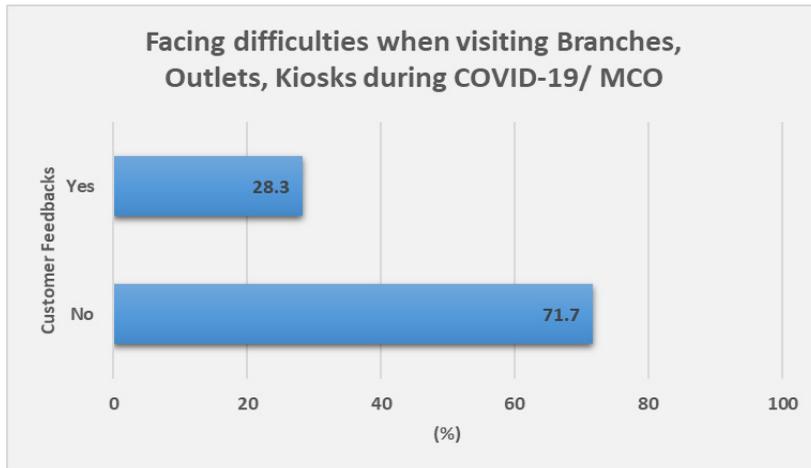


Figure 1. Analysis of customer preference when performing transactions

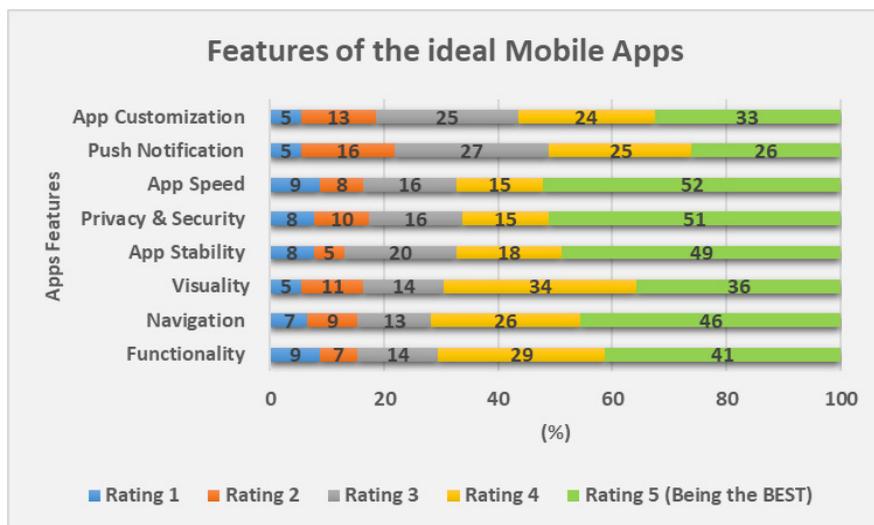


Figure 2. Analysis of operational changes due to the MCO

According to Figure 2, 28.3% of customers faced difficulties while performing a transaction at a physical counter during the MCO. Meanwhile, 71.7% of customers faced no difficulties. Twenty-four customers gave feedback on their difficulties during the MCO, which are:

- Long queues waiting time and do not exist in the form of SST, such as a request for account termination and transfer ownership (Feedback by 10 customers)
- Operating hours—closed and open at specific hours (Feedback by six customers)
- Following the Standard Operating Procedure (SOP) (Feedback by five customers)
- Others (feedback on services, people) (Feedback by three customers)

In addition, the newly proposed framework contains processes to analyse and formulate the framework through a mobile self-service solution based on existing features. This research elicited the ideal mobile app service practices to integrate the processes.

According to Figure 3, more than 41% of customers rated five for app speed, privacy & security, app stability, navigation, and functionality, which were regarded as the best practices. Meanwhile, 36% gave a rating of five for app visuality, 33% gave a rating of five for app customization, and 26% gave a rating of five for push notification. A mobile app's most important function or feature is to provide utility or value to customers. Thus, the prioritization and each requirement have been mapped out with the problem statement and previous study to provide a clear picture of critical features versus highly desired or nice to have features. Based on the survey, 11 customers gave the following suggestions for functionality or features:

- Outage or maintenance information (suggested by one customer)
- Connect with personnel support or agent (suggested by three customers)
- Log a technical report (suggested by one customer)
- Privacy & personalization (suggested by one customer)
- Security (suggested by two customers)
- Simple guidelines or manual for using the app (suggested by two customers)
- Bill reminder (suggested by one customer)

Summary of Analysis

Overall, the existing mobile applications supported the service providers' and helped customers perform the related transactions. The weaknesses to be resolved are listed below.

- i. None of the mobile apps consists of all services; thus, the SST in the telecommunication industry in Malaysia can be further improvised by expanding the technology for a customer to perform self-service.
- ii. There are too many reasons for contacting their service provider's support team resulting in congestion in the telephone line. Examples of reasons are requesting technical assistance or asking about account information and billing related.
- iii. Long waiting times and tedious processes are involved in getting an answer, update or solution at the physical branches, especially during MCO, where the working hour is shortened, and fewer customers can be attended to. The main features and services can be enhanced to benefit the customers. The existing services in the mobile apps can be enhanced by providing more guidelines and frequently asked questions (FAQ) to help customers use the application and the available features. These are customer requirements for improving mobile applications, which served as the basis for formulating the self-service framework.

THE FORMULATION OF A MOBILE SELF-SERVICE FRAMEWORK

It is important to enhance customers' engagement and experience through the quality of interactions and make the technical features of offline and online interactions easier to use to retain customer loyalty (Dhasan & Kowathanakul, 2021). Thus, this research aimed to formulate a Mobile Self-Service Framework for the Telecommunications Industry (MOFTI). The following section describes the proposed framework.

First, customers must verify their security using a two-factor authentication after logging in to their account successfully. Next, the customer can see the guidelines and FAQs under the dashboard application. The processes highlighted in orange are the five processes adopted from the existing mobile SST:

- Process 1 : Login
- Process 2 : Enter dashboard application
- Process 3 : Select the menu or features
- Process 4 : Perform the transaction or run content
- Process 5 : Received the transaction ID

The processes highlighted in brown are the new processes formulated by this research. In order to improve the customer experience by using a digital platform instead of the conventional method, this research formulates new processes:

- Process 6 : Display Guideline & FAQ

In order to improve the customer experience at the physical counter (Branches/ Outlets/Kiosks during MCO (research objective 2), this research formulates two new processes:

- e : Transfer Ownership

- Process 8 : Termination

In order to provide the mobile SST best practices for service providers, this research formulates two new processes:

- Process 9 : Display the list of actions that service providers can take. These actions are based on the requirements elicited.

- e : Service providers to enhance the features and services to improve customer experience based on literature review and survey results.

The other new features are described below:

There are three sub-processes under non-technical services: Transfer Ownership, Outages, and Termination. Transfer Ownership means the customer requests the transfer of an account or ownership to another person (a new owner), as shown in the sub-process below:

1. The owner applies to transfer ownership by filling in details such as their name and the new owner's Malaysian Identity Card (MyKad) number and then uploading a soft copy of the MyKad.

2. The new owner must upload a soft copy of their MyKad and accept the transfer process.
3. An agent from the telecommunications company requests more supporting documents and approves.

The term outages mean a process of handling downtime, whereby the system enables customers to view the planned maintenance or downtime in a specific area. The sub-process for outages includes the following points:

1. The customer selects the area code or service number.
2. The system displays the planned maintenance or downtime status.

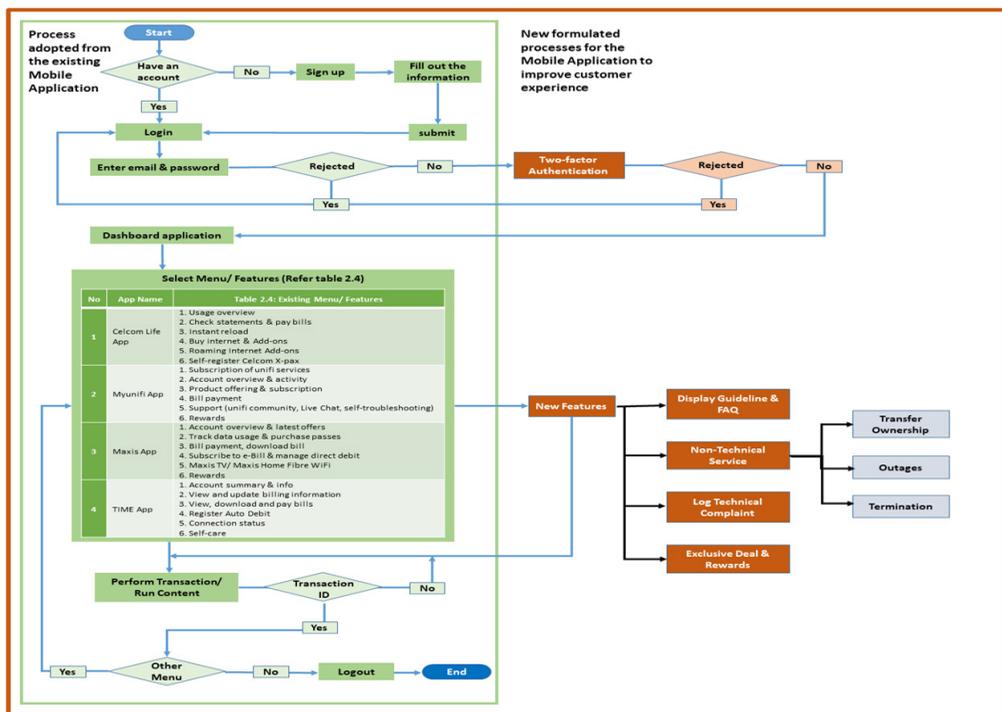


Figure 4. The proposed Mobile Self-Service Framework for the Telecommunications Industry (MOFTI)

The termination feature enables customers to terminate their accounts. The sub-processes for termination include:

1. The customer requests termination.
 2. The system executes persuasive features to persuade customers not to terminate the account. For example, the system offers exclusive deals and reward features.
 3. If the customer complies with the terms and conditions, the termination request will be approved.
- 3.1 A feature allows customers to request to switch their account to another service provider.

- 3.2 A notification is received from the new service provider.
- 3.3 The account is terminated.
4. If customers do not comply with the terms and conditions:
 - 4.1 A penalty is calculated.
 - 4.2 The customer pays the penalty.
 - 4.3 The process is escalated to step 3.

The exclusive deals and rewards feature aims to provide customers with special offers and points that they can use for any campaign, promotion, merchandise, or rewards redemption, to make new purchases using a combination of points, or to associate with other vendors for prizes and customer loyalty reward points. The sub-processes of the exclusive deals and rewards feature are:

1. The customer checks for exclusive deals and points.
2. Upon reaching a certain number of points, customers are allowed to redeem exclusive deals and rewards.
3. The agent is notified of the redemption.
4. The agent validates and processes the redemption.

The log technical complaint feature aims to replace chatting with live agents and long repetitive calls:

1. Self-troubleshooting information is displayed.
2. The customer creates a technical complaint.
3. The agent validates the complaint.
 - 3.1 Instructions are sent to the customer to rectify the problem.
 - 3.2 Requests are sent for technical personnel to resolve the issue.
 - 3.3 The technical personnel update the status.
4. If the problem is resolved, the customer closes the complaint.
5. If the complaint is unresolved company sends the customer. The process is escalated back to step 3.

EVALUATION

The evaluation of the analysis, survey results, and interview results aimed to validate that the newly proposed solution could enhance the customer experience through an SST that accommodated restrictions on people due to the MCO. This research developed a series of mobile app GUI based on the formulated framework to ease customers in understanding the framework while evaluating. The evaluation was done by collecting 2020 data from the telecommunications industry and analysing the data referring to the proposed mobile self-service framework, as illustrated in Figure 4, together with the respective GUI. The evaluation results show that more than 78% of customers would be satisfied if the following

features or services were available on their mobile apps. Furthermore, the questionnaire results show that more than 73% of customers agree that Mobile Apps are more convenient than contacting a support team via phone.

The Ideal Mobile Apps for the Proposed Mobile Self-Service Framework

The questionnaire results show that more than 41% of customers rated five (five being the best) for app functionality, followed by app navigation, app stability, app privacy & security and app speed. Meanwhile, less than 36% of customers rated the app visuality, app customization and push notifications based on the ideal Mobile App for the MOFTI.

Customer Preference for the Platform (Method) for the MOFTI

The questionnaire results show that more than 62% of customers choose the Mobile App as the most desired platform for the MOFTI. Tables 5 to 8 and Figures 5 to 7 describe the detailed findings of the MOFTI.

Table 5

Results of customer satisfaction for the MOFTI

Account overview, activity & other services related	Total of respondents (N=92)	Percentage %
Strongly agree	45	49.0
Agree	40	43.0
Neither agree nor disagree	7	8.0
Disagree	0	0
Strongly disagree	0	0
Bill statement & payment system	Total of respondents (N=92)	Percentage %
Strongly agree	48	52.0
Agree	38	41.0
Neither agree nor disagree	5	5.0
Disagree	1	1.0
Strongly disagree	0	0.0
Product offering, promotion & subscription	Total of respondents (N=92)	Percentage %
Strongly agree	30	33.0
Agree	44	48.0
Neither agree nor disagree	16	17.0
Disagree	1	1.0
Strongly disagree	1	1.0

Table 5 (Continue)

Support services/ self-care	Total of respondents (N=92)	Percentage %
Strongly agree	34	37.0
Agree	42	46.0
Neither agree or disagree	15	16.0
Disagree	1	1.0
Strongly disagree	0	0.0
Upgrade/ downgrade/ switching numbers	Total of respondents (N=92)	Percentage %
Strongly agree	40	43.0
Agree	35	38.0
Neither agree nor disagree	13	14.0
Disagree	4	4.0
Strongly disagree	0	0.0
Termination request/ transfer ownership	Total of respondents (N=92)	Percentage %
Strongly agree	36	39.0
Agree	36	39.0
Neither agree nor disagree	16	17.0
Disagree	4	4.0
Strongly disagree	0	0.0
Relocation request	Total of respondents (N=92)	Percentage %
Strongly agree	35	38.0
Agree	38	41.0
Neither agree nor disagree	14	15.0
Disagree	5	5.0
Strongly disagree	0	0.0
Store finder	Total of respondents (N=92)	Percentage %
Strongly agree	34	37.0
Agree	42	46.0
Neither agree nor disagree	14	15.0
Disagree	2	2.0
Strongly disagree	0	0.0
Rewards point & redemption	Total of respondents (N=92)	Percentage %
Strongly agree	42	46.0
Agree	40	43.0
Neither agree nor disagree	7	8.0
Disagree	3	3.0
Strongly disagree	0	0.0

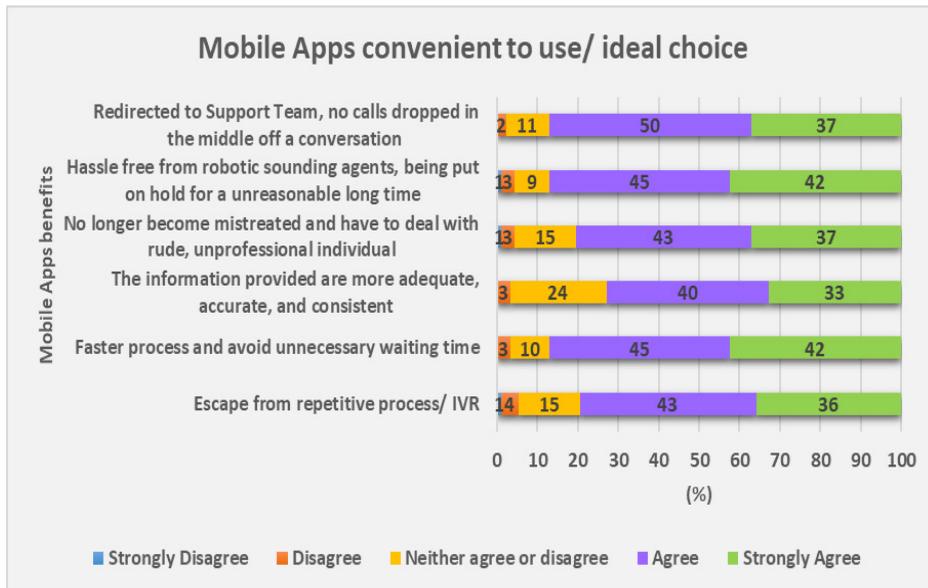


Figure 5. Results showing the mobile App as a convenient platform for the MOFTI

Table 6

Results showing the mobile App as a convenient platform for the MOFTI

Escape from repetitive process / IVR	Total of respondents (N=92)	Percentage %
Strongly agree	33	35.9
Agree	40	43.5
Neither agree nor disagree	14	15.2
Disagree	4	4.3
Strongly disagree	1	1.1
Faster process and avoid unnecessary waiting time for an answer, update, or solution	Total of respondents (N=92)	Percentage %
Strongly agree	39	42.4
Agree	41	44.6
Neither agree nor disagree	9	9.8
Disagree	3	3.3
Strongly disagree	0	0.0
The information provided is adequate, accurate, and consistent	Total of respondents (N=92)	Percentage %
Strongly agree	30	32.6
Agree	37	40.2
Neither agree nor disagree	22	23.9
Disagree	3	3.3
Strongly disagree	0	0.0

Table 6 (Continue)

No longer become mistreated and have to deal with rude, unprofessional individual	Total of respondents (N=92)	Percentage %
Strongly agree	34	37.0
Agree	40	43.5
Neither agree nor disagree	14	15.2
Disagree	3	3.3
Strongly disagree	1	1.1
Hassle-free from robotic-sounding agents, being put on hold for an unreasonable long time	Total of respondents (N=92)	Percentage %
Strongly agree	39	42.4
Agree	41	44.6
Neither agree nor disagree	8	8.7
Disagree	3	3.3
Strongly disagree	1	1.1
Redirected to support team for further assistance without having to contact the hotline no, facing disconnected or calls dropped in the middle of a conversation	Total of respondents (N=92)	Percentage %
Strongly agree	34	37.0
Agree	46	50.0
Neither agree nor disagree	10	10.9
Disagree	2	2.2
Strongly disagree	0	0.0

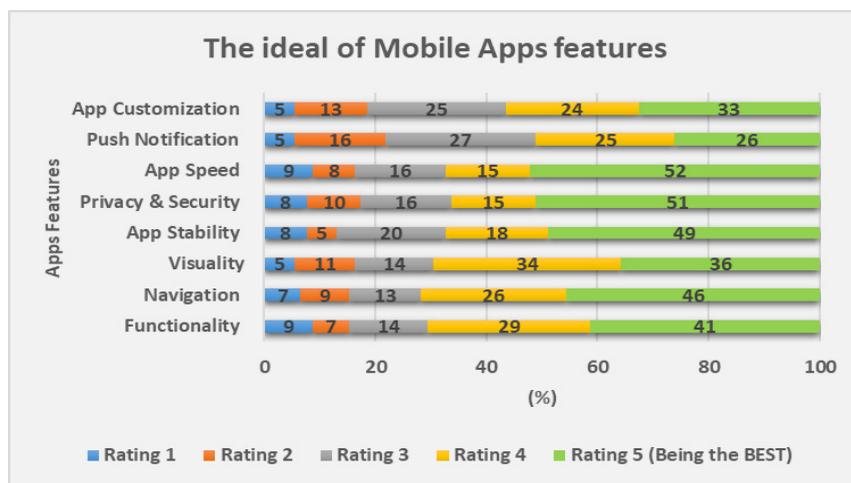


Figure 6. Results of the ideal mobile App for the MOFTI

Table 7

Results of the ideal mobile App for the MOFTI

Functionality (Relevant features, social integration, feedback system)	Total of respondents (N=92)	Percentage %
Rating 1	8	8.7
Rating 2	6	6.5
Rating 3	13	14.1
Rating 4	27	29.3
Rating 5 (Being the BEST)	38	41.3
Navigation (Able to access the information quickly and easily, simple menu selection)	Total of respondents (N=92)	Percentage %
Rating 1	6	6.5
Rating 2	8	8.7
Rating 3	12	13.0
Rating 4	24	26.1
Rating 5 (Being the BEST)	42	45.7
Visuality (Look and feel: interactive design, content layout, visually appealing)	Total of respondents (N=92)	Percentage %
Rating 1	5	5.4
Rating 2	10	10.9
Rating 3	13	14.1
Rating 4	31	33.7
Rating 5 (Being the BEST)	33	35.9
App Stability (App reliable & compatible, seldom crashed)	Total of respondents (N=92)	Percentage %
Rating 1	7	7.6
Rating 2	5	5.4
Rating 3	18	19.6
Rating 4	17	18.5
Rating 5 (Being the BEST)	45	48.9
Privacy & Security (Provide two-factor authentication: personal question, SMS confirmation code, biometric authentication: fingerprint and retina)	Total of respondents (N=92)	Percentage %
Rating 1	7	7.6
Rating 2	9	9.8
Rating 3	15	16.3
Rating 4	14	15.2
Rating 5 (Being the BEST)	47	51.1

Table 7 (Continue)

App Speed (Fast loading screen, high performance)	Total of respondents (N=92)	Percentage %
Rating 1	8	8.7
Rating 2	7	7.6
Rating 3	15	16.3
Rating 4	14	15.2
Rating 5 (Being the BEST)	48	52.2
Push Notification (Relevant and personalized push messaging option: manage alerts, secured messaging, receive timely updates, easy tracking of notification)	Total of respondents (N=92)	Percentage %
Rating 1	5	5.4
Rating 2	15	16.3
Rating 3	25	27.2
Rating 4	23	25.0
Rating 5 (Being the BEST)	24	26.1
App Customization (Personalization option: able to filter, control, sorts out important features, quick access)	Total of respondents (N=92)	Percentage %
Rating 1	5	5.4
Rating 2	12	13.0
Rating 3	23	25.0
Rating 4	22	23.9
Rating 5 (Being the BEST)	30	32.6

Service Provider Perspectives of the MOFTI

Based on the interviews, most service providers agreed that the MOFTI would benefit and improve the customer experience and services. However, risk factors such as security, fraud, and technology adaptation should be considered. Nevertheless, TIME has proven that customer transactions or services can be done online since they have never had a physical store. Table 9 describes the detailed results of the service providers’ perspectives of mobile apps during the interview sessions.

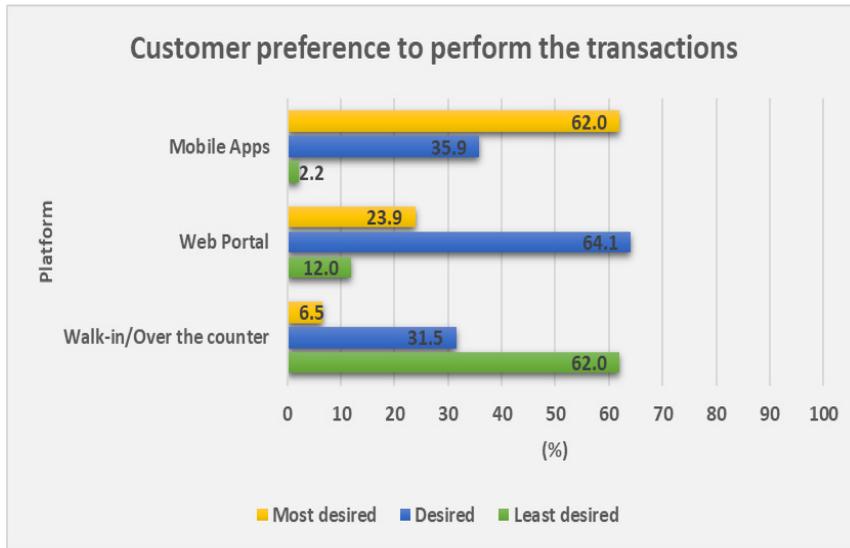


Figure 7. Results showing customer preference for the platform (method) for the MOFTI

Table 8

Results showing the customer preference for the platform (method) for the MOFTI

Mobile Apps	Total of respondents (N=92)	Percentage %
1-Most desired	57	62.0
2-Desired	33	35.9
3-Least desired	2	2.2
Web Portal	Total of respondents (N=92)	Percentage %
1-Most desired	22	23.9
2-Desired	59	64.1
3-Least desired	11	12.0
Walk-in/over the counter	Total of respondents (N=92)	Percentage %
1-Most desired	6	6.5
2-Desired	29	31.5
3-Least desired	57	62.0

Table 9

Results of the service providers' perspectives of the MOFTI

Mobile Apps improve the services/ bring benefit	Total of respondents (N=8)	Percentage %
Improve customer experience	8	100.0
Improve customer loyalty	8	100.0
Improve efficiency (function & support Services)	8	100.0
Reduce cost & increase revenue	8	100.0
The services that currently cannot be done online	Total of respondents (N=8)	Percentage %
Change ownership	3	37.5
Termination	6	75.0
Replacement sim card	3	37.5
Downgrade plan	1	12.5
All thru online platform (website/ Email/ call/ social media)	1	12.5

Summary of Evaluation

Table 10 summarises the scores and indicates that the newly proposed mobile self-service framework uplifts customer experience towards telecommunication service providers in Malaysia.

Table 10

Comparison between the existing framework and the MOFTI

No	Attributes	Score	
		Existing framework (%)	MOFTI (%)
Features/ Services			
1	The average results of the offering features/ services	76.10	78.00
Customer Satisfaction/ Experience			
2	The average results of customer satisfaction/ experience	44.60	73.00
Mobile Apps Factor			
3	The average results of the mobile apps factor	54.30	66.00
Platform/ Method			
4	Customer preference	19.40	62.0

The mean values for the MOFTI are generally higher than those for the existing framework. The customers agreed that the MOFTI would perform transaction-related tasks more satisfactorily than the existing framework. It is mainly because the MOFTI consists of ten newly formulated processes designed to:

- i. Enhance the services or features of a mobile app to improve the customer experience since there are flaws and limitations with conventional interaction (live agents) and IVR.
- ii. Improve the operational changes by making them digital due to the MCO, which has impacted customer experience at physical counters (e.g., at branches, outlets or kiosks).
- iii. Analyse and formulate the framework through a mobile self-service solution based on the existing features since there is a lack of studies on mobile app service practices.

Among the four aspects evaluated, most customers feel satisfied with the features or services offered by the MOFTI (78%), an increase of two percentage points compared to the existing frameworks (76%). The increment shows that the automation of physical counter services has improved customer satisfaction. Based on the customer experience results, there was a tremendous increment of about 28.4%, with customers finding it more convenient to use the MOFTI (73%) than the existing frameworks (44.6%).

CONCLUSION AND FUTURE WORK

The main aim of this research is to overcome the unpleasant conventional methods since there are flaws and limitations with conventional interaction (live agents) and IVR as well as with the operational changes due to the MCO, which has impacted customer experience at physical counters.

In realizing the importance of resolving these problems, this research achieves its first research objective: to analyse customer preference and experience when using the mobile self-service features provided by telecommunications service providers in Malaysia while managing the restrictions people face due to the MCO through analysis of the survey results. Then this research achieves the second research objective by formulating MOFTI to uplift customer experience based on an analysis of the literature review, the survey results, and the frameworks of the existing mobile SST. Next, this research achieves the third research objective by providing evidence that the proposed self-service framework can uplift customer experience towards telecommunication service providers evaluation results prove that the MOFTI could resolve the limitations of conventional interaction and is highly relevant to today's environment. Another strength of completing this research was customers' positive reception of the MOFTI. Therefore, this research recommends that MOFTI be referred by personnel from telecommunication industries in Malaysia to

enhance the improvements that made it possible to complete the primary task involving customer activities.

The novelties for this study are projected to benefit society and telecommunications service providers, considering that mobile apps have the highest potential to contribute to the customer experience through SST.

ACKNOWLEDGEMENT

This research was initiated by Telekom Malaysia (Headquarter) under the Strategic Planning Unit. We want to thank this unit, which provided insight and expertise that greatly assisted the research, the interpretations, and the conclusions of this paper.

REFERENCES

- Al-Mashraie, M., Chung, S. H., & Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, 144, Article 106476. <https://doi.org/10.1016/j.cie.2020.106476>
- Aziz, N. A., Othman, J., Lugova, H., Suleiman, A., behalf of the Economy, O., & Cluster, S. W. (2020). Malaysia's approach in handling COVID-19 onslaught: Report on the Movement Control Order (MCO) and targeted screening to reduce community infection rate and impact on public health and economy. *Journal of Infection and Public Health*, 13(12), 1823-1829. <https://doi.org/10.1016/j.jiph.2020.08.007>
- Barua, Z., Aimin, W., & Hongyi, X. (2018). A perceived reliability-based customer satisfaction model in self-service technology. *The Service Industries Journal*, 38(7-8), 446-466.
- Basole, R. C., & Karla, J. (2012). Value transformation in the mobile service ecosystem: A study of app store emergence and growth. *Service Science*, 4(1), 24-41. <https://doi.org/10.1287/serv.1120.0004>
- Beatson, A., Coote, L. V., & Rudd, J. M. (2006). Determining consumer satisfaction and commitment through self-service technology and personal service usage. *Journal of Marketing Management*, 22(7-8), 853-882. <https://doi.org/10.1362/026725706778612121>
- Bhale, U., & Bedi, H. S. (2021). A study on the customer experience of work from home mobile users during COVID 19 pandemic. In *Eighteenth AIMS International Conference on Management* (pp. 4-6). Elsevier.
- De Leon, M. V., Atienza, R. P., & Susilo, D. (2020). Influence of self-service technology (SST) service quality dimensions as a second-order factor on perceived value and customer satisfaction in a mobile banking application. *Cogent Business & Management*, 7(1), Article 1794241. <https://doi.org/10.1080/23311975.2020.1794241>
- Dhasan, D., & Kowathanakul, S. (2021). The impact of service quality, promotions and customer engagement in determining customer loyalty in the Thai mobile network industry. *ABAC Journal*, 41(1), 209-240.
- Fang, H., Wang, L., & Yang, Y. (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China. *Journal of Public Economics*, 191, Article 104272. <https://doi.org/10.1016/j.jpubeco.2020.104272>

- Lu, C., Geng, W., & Wang, I. (2015). The role of self-service mobile technologies in the creation of customer travel experiences. *Technology Innovation Management Review*, 5(2), Article 871.
- MCMC. (2020). *Industry performance report*. Malaysian Communications and Multimedia Commission. <https://www.mcmc.gov.my/en/resources/industry/industry-performance-report>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45-77. <https://doi.org/10.2753/MIS0742-1222240302>
- Rawson, A., Duncan, E., & Jones, C. (2013). The truth about customer experience. *Harvard Business Review*, 91(9), 90-98.
- Riahi, A., & Ristock, H. W. A. (2014). *U.S. Patent No. 8,767,948*. Washington, DC: U.S. Patent and Trademark Office.
- Salam, M. A., Saha, T., Rahman, M. H., & Mutsuddi, P. (2021). Challenges to mobile banking adaptation in COVID-19 pandemic. *Journal of Business*, 9(3), 101-113.
- Salomann, H., Kolbe, L., & Brenner, W. (2006). Self-services in customer relationships: Balancing high-tech and high-touch today and tomorrow. *E-Service*, 4(2), 65-84. <https://doi.org/10.2979/ESJ.2006.4.2.65>
- Stevens, J. P. (2020). *Applied multivariate statistics for the social sciences*. Routledge.
- Wang, J., Lai, J. Y., & Chang, C. H. (2016). Modeling and analysis for mobile application services: The perspective of mobile network operators. *Technological Forecasting and Social Change*, 111, 146-163. <https://doi.org/10.1016/j.techfore.2016.06.020>
- Wang, W. T., Ou, W. M., & Chen, W. Y. (2019). The impact of inertia and user satisfaction on the continuance intentions to use mobile communication applications: A mobile service quality perspective. *International Journal of Information Management*, 44, 178-193. <https://doi.org/10.1016/j.ijinfomgt.2018.10.011>
- Xu, C., Peak, D., & Prybutok, V. (2015). A customer value, satisfaction, and loyalty perspective of mobile application recommendations. *Decision Support Systems*, 79, 171-183. <https://doi.org/10.1016/j.dss.2015.08.008>
- Yang, H. D., Lee, J., Park, C., & Lee, K. (2014). The adoption of mobile self-service technologies: Effects of availability in alternative media and trust on the relative importance of perceived usefulness and ease of use. *International Journal of Smart Home*, 8(4), 165-178. <https://doi.org/10.14257/IJSH.2014.8.4.15>
- Zhou, Y. C., Xue, L., Liu, X. P., Wang, X. N., Liang, X. X., & Sun, C. H. (2010). Service storm: A self-service telecommunication service delivery platform with platform-as-A-service technology. In *2010 6th World Congress on Services* (pp. 8-15). IEEE Publishing. <https://doi.org/10.1109/SERVICES.2010.45>

Review Article

Dehydrated Food Waste for Composting: An Overview

Aziz Khalida¹, Veknesh Arumugam¹, Luqman Chuah Abdullah¹, Latifah Abd Manaf² and Muhammad Heikal Ismail^{1*}

¹Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Department of Environment, Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Food waste disposal has recently received much attention worldwide due to its major impact on environmental pollution and economic costs. Using high moisture content of food waste has the highest negative environmental impact due to increased greenhouse gas emissions, odor, and leachate. Drying technologies play an important role in reducing the moisture content of food waste, which is necessary for environmental sustainability and safety. The first part of this review highlights that sun-drying is the most cost-effective drying method. However, it has not been widely recommended for food waste management due to several limitations, including the inability to control sunray temperature and the inability to control end-product quality. Thermal drying eliminates moisture from food waste quickly, preventing hydrolysis and biodegradation. Thermal dryers, such as the GAIA GC-300 dryer, and cabinet dryer fitted with a standard tray, are the best alternative to sun drying. The second part of this review highlights that dehydrated food waste products are slightly acidic (4.7–5.1), have a high electrical conductivity (EC) value (4.83–7.64 mS cm⁻¹), with high nutrient content, due to low pH levels, dehydrated food waste is not suitable for direct use as a fertilizer for the plants. So, the dried food waste should be composted

before application to the plants because the composting process will dominate the limitation of phytotoxins, anoxia, salinity, and water repellence. Trench compost can be a good choice for decomposing dried organic waste because trench compost relies solely on soil decomposing microorganisms and insects.

Keywords: Dehydration, food waste, thermal drying, trench composting

ARTICLE INFO

Article history:

Received: 8 December 2021

Accepted: 9 March 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.33>

E-mail addresses:

khalidaaziz143@gmail.com (Aziz Khalida)

heikal@upm.edu.my (Muhammad Heikal Ismail)

viknesharumugam2@gmail.com (Veknesh Arumugam)

chuah@upm.edu.my (Luqman Chuah Abdullah)

latifahmanaf@upm.edu.my (Latifah Abd Manaf)

*Corresponding author

INTRODUCTION

The increasing global population growth of 7.6 billion has increased waste generation worldwide. It has enormously impacted the climate, animals, and human beings, recently attracting great global attention (Thani et al., 2019). In 2016, the Food Aid Foundation announced that Malaysia lost almost 15,000 tonnes of food waste, including 3,000 tons of food per day (Sulaiman & Ahmad, 2018) and around 15 million tonnes (234 kg/person/year or 50% of food) were lost annually in the United Kingdom (Salemdeeb et al., 2017). Food waste is one of the big problems in the world. Most developing countries lose billions of dollars per year due to food waste. This problem does not stop at the point where food is discarded. Around 95% of discarded food goes down in areas where anaerobic digestion processes nitrogen, carbon dioxide, or other gases, which has a catastrophic effect on global warming. Food waste issues appeared to be growing over the next 25 years because of intensive economic and population growth in most Asian countries, particularly Malaysia (Melikoglu et al., 2013).

Food waste exacerbates climate change. Because food waste produces methane, a gas that absorbs heat considerably more quickly than other greenhouse gases such as carbon dioxide, methane has a shorter lifetime than carbon dioxide, which results in the globe's rapidly warming (Zaki, 2019). In the last few years, food waste has increased in Malaysia, affecting the country's solid waste management system, including landfills and incineration. The government was currently constrained to alternative food waste disposal solutions such as compost, animal feed, and dehydration of food waste (Lim et al., 2016). Nowadays, significant factor considerations must be considered before implementing food waste disposal units as a wide-ranging solid waste management option, as solid waste is related to emissions of greenhouse gasses and other environmental damage. Continued research and development in food waste could better manage this choice to become a sustainable alternative to landfilling, incineration, and composting (Ismail et al., 2020a; Arumugam et al., 2021).

Utilizing a thermal GAIA GC-300 dryer is the best option for landfill and incineration. Food waste is called FORBI after it has been dried in this dryer; dried food waste that has high-calorie biomass with low moisture content, that product can be used to manufacture animal feed and raw material for compost (Papanikola et al., 2019). The dryer can handle 50 kg of food waste every eight hours. GAIA GC-300 dryer is becoming incredibly common due to its affordable price. It operates at 140 degrees and can eliminate any present microbes in the food waste by drying it out. Overall, this technique has reduced food waste in the long term (Vakalis et al., 2018). Dehydration has recently been investigated as a method of reusing food waste, especially in animal feed. According to an analysis of food waste conversion in a central animal feed manufacturing plant, substituting conventional animal feed can significantly reduce net greenhouse gas emissions (Hall, 2016). Dehydration

of food waste resulted in a 70% reduction in mass and a lower energy cost than regular food waste treatment. The low water content of the post-dehydration material delays decomposition and decreases odor, allowing for less frequent waste disposal. Dried food's physical and chemical properties make it a useful feedstock for value-added products like compost and animal feed (Schroeder et al., 2020).

Food waste and leftovers can be exploited as raw materials for manufacturing vital commercial compounds because it contains carbohydrates, proteins, lipids, and nutraceuticals. As well, the latest food waste treatment policy focuses on waste production reduction and less on recycling (Ravindran & Jaiswal, 2016). At the same time, the moisture content of food waste is over 80%. Using food waste with a high moisture content of 75 to 95% has the greatest negative impact on the environment due to increased greenhouse gas emissions and leachate released into the soil. The best solution to this problem is to dehydrate the food waste, which can then be used as animal feed or raw materials for fertilizer (Sotiropoulos et al., 2016).

A direct application of dehydrated food waste as a soil amendment appears implausible. Since food waste rehydrates and creates fungus. Thus, dry food contains significant nitrogen and carbon and should be composted before being placed in soil. Trench compost is recommended for dehydrated food waste because it does not require determining moisture levels, aeration, or sifting as with a compost pile (Compost Education Centre, 2010). Removing excess moisture from the source by drying reduces the waste content and volume in the kitchen. Continued study and development in food waste management might help dehydration become a more sustainable alternative to landfilling. This review has two parts: The first part of this review defined in general which methods can be utilized to dry food waste, based on published articles and research papers. The second part of the study focused on the physicochemical qualities of dehydrated food waste products to use as a biofertilizer after composting. This study was done to identify and select the best suitable food waste treatment for dehydration before composting for future research.

FOOD WASTE

Food waste is food collected from kitchen areas in restaurants and households which we cannot consume. A US Department of Agriculture (USDA) has found that 21% of fresh food at restaurants was not consumed. Usually, the restaurant service industry wastes 4–10% of the food before service to the customer (Sakaguchi et al., 2018). Food waste is a subset of food loss that consists of material meant for human consumption but not consumed. The border between food loss and food waste is not well defined. Food loss is common in the food value chain before it reaches the customer. It can occur during growing, harvesting, processing, or transportation. However, food waste happens due to mismanagement in the supply chain. Food waste is typically regarded as avoidable food loss. However, the

core reasons and motivations of the people involved in food waste are now the significant variables distinguishing food waste from food losses (Kibler et al., 2018).

Food is wasted at every step in the food supply chain (FSC); many points along the supply chain degrade or lose edible food mass (e.g., post-harvest handling, processing, distribution, and consumption). Figure 1 shows how food waste is generated in the FSC (Papargyropoulou et al., 2014). Food is wasted to a large extent in middle and high-income countries, both at the consumer and manufacturing levels. However, in underdeveloped nations, food is wasted at the beginning of the food supply chain (Chauhan et al., 2021).

Food waste per capita is the most prominent in Europe and North America (95–115 kg/year), while rock-bottom in Sub-Saharan Africa and South/Southeast Asia (6–11 kg/year). More than 40% of food losses occur in developing countries after harvesting and preparing the produce. However, in developed nations, over 40% at the retail and customer levels, most of the food waste is generated. Figure 2 shows that per person, wasted food varies around the globe (Blakeney, 2019). The amount of food wasted by consumers in developed countries (222 million tons) is nearly equal to the total food supply in sub-Saharan Africa (230 million tons). Figure 3 shows the volumes of food waste produced by commodity groups in various regions (Blakeney, 2019).

Environmental and social impacts on food surplus and waste were prioritized, according to Papargyropoulou et al. (2014). The three themes that were proposed for the framework were as follows:

- I. Food security and surplus
- II. Avoidable and unavoidable food waste
- III. Waste prevention and waste management

The food waste hierarchy is used to rank the options. The framework recommends recycling food waste into animal feed or composting after exhausting all preventative measures. If recycling becomes impossible, the next best option is the anaerobic digestion of food waste. Finally, when all other options have been exhausted, the only option left is disposed of in landfills. Figure 4 shows the framework for food waste (Papargyropoulou et al., 2014).

Classes of Food Waste

Disposed and abandoned food can be classified into two types, food waste and leftovers. Components of food waste can be separated from the original material. Most kitchen wastes was from food waste such as vegetables, fruit seeds, and peels (Al-Domi et al., 2011). In this grouping, inedible foods like coffee grounds were also included. Further, food waste consists of foodstuffs discarded during manufacturing, transport, and expired in stores, restaurants, and schools (Silvennoinen et al., 2015). Leftovers usually include the food residues left

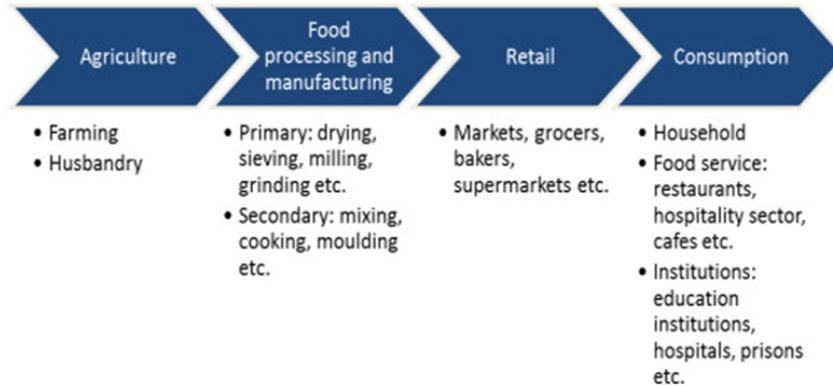


Figure 1. Activities that lead to food waste and losses in the food supply chain (Papargyropoulou et al., 2014)

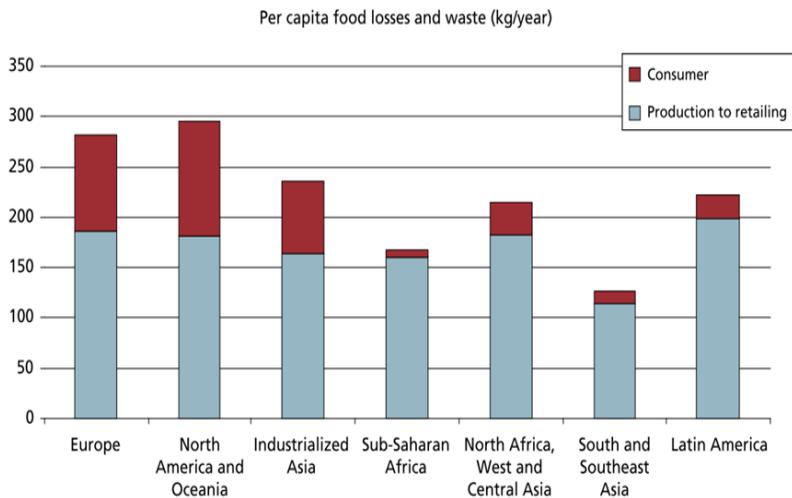


Figure 2. Food losses and waste per capita in various regions, at the consumption and pre-consumption stages (Blakeney, 2019)

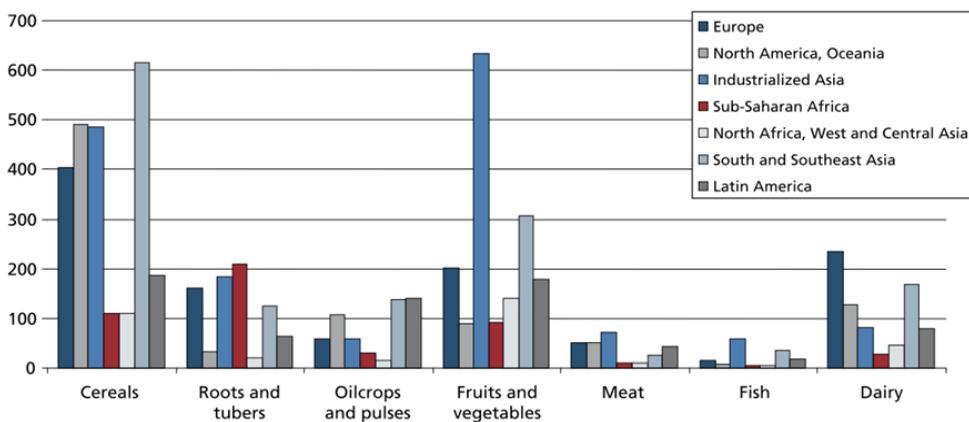


Figure 3. Commodity groups produce volumes of food waste in various regions (Blakeney, 2019)

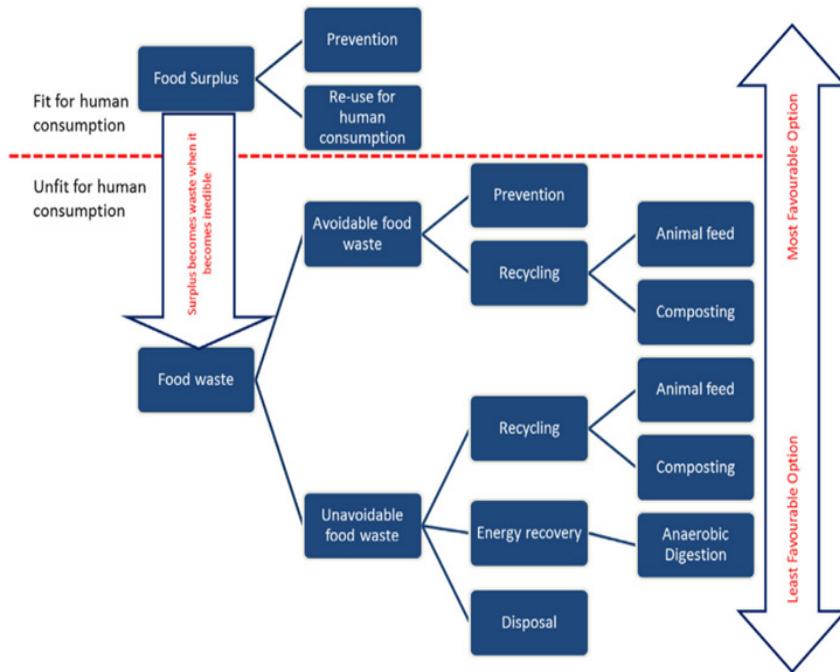


Figure 4. Framework for food surplus and waste (Papargyropoulou et al., 2014)

on plates that consumers dispose of. These wastes are also known as post-consumer food, and it usually includes domestic foods that have been discarded and are often co-mingled with yard debris upon disposal. This food waste can often be spotted in a large amount in dining restaurants, buffets, and hotels (McAdams et al., 2019). Leftovers were not consumed but were physically and economically unsustainable due to the environmental effects of manufacturing and processing raw materials into food. Discarded food scraps and leftovers were biodegradable wastes generated by many companies and homes, including the food industry, hospitality sectors, and households (Silvennoinen et al., 2015).

Cereals, roots and tubers, oil and pulses, fruits and vegetables, meats, fish and seafood, and dairy products are the seven subcategories of food waste. Functional compounds can be extracted from agricultural and food processing by-products. Therefore, it is necessary to understand analytical chemistry principles to extract and isolate these compounds from food waste (Galanakis, 2012). The applied methodologies are introduced with the following objectives:

- I. increasing the target compounds' yield
- II. meeting the needs of industrial manufacturing processes
- III. removal of impurities and toxic compounds from the high-value ingredients
- IV. processing and storing the product to ensure its integrity
- V. ensuring the food-grade nature of the final product

Some steps were eliminated or overlapped in Galanakis’s (2012) five recovery stages. Purification and encapsulation of target compounds occur after extraction of specific macromolecules Figure 5. This downstream scheme is used to recover two components or a macromolecule. Because proteins are macromolecules, this step can be skipped when the target compound is a protein (Galanakis, 2012). Before food waste can be used, it must undergo complex processing and extensive research. A thorough investigation of food waste type, volume, exploitation potential, and end-users are required to justify the investment. The environmental consequences of new manufacturing processes must also be considered. Finally, excessive food modification may pose health risks to consumers (Dhar, 2016).

Characteristics of Food Waste

Food waste characteristics generated worldwide were known as one of the top energy and substrata recovery solutions. Food waste values have varied, and 24% of these changes were due to their geographical origin, collecting source, and season (Fisgativa et al., 2016). The properties of the various types of solid waste were shown. Even as solid waste, all samples were disposed of certain samples (canned products, spent coffee, fruit and vegetable wastes, salad mix, and kitchen preparations waste have a solid content of < 30%). Substrates with

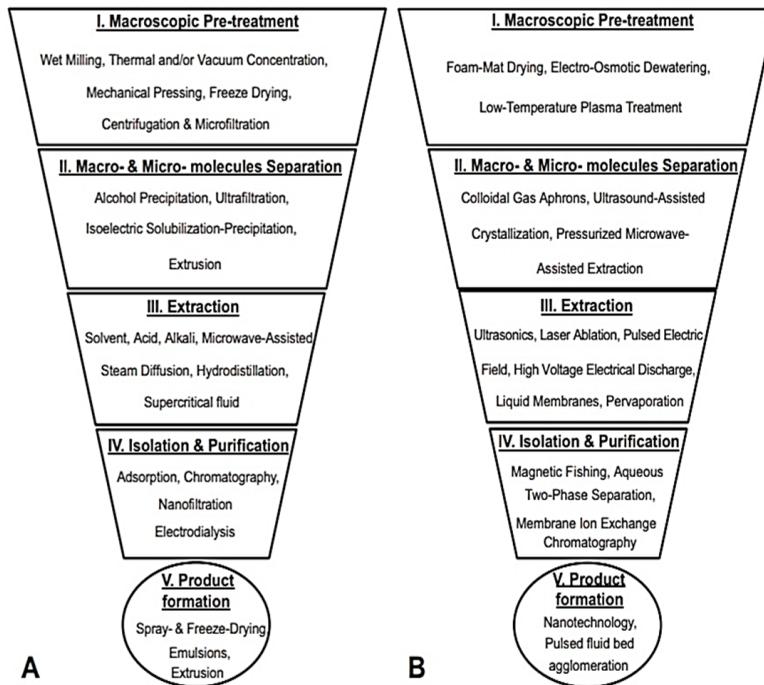


Figure 5. Food waste recovery stages: (A) Established (B) New Technologies (Galanakis, 2012)

high biodegradability have been highly rich in rapidly could be hydrolyzed carbohydrates and lipids. High-fat substrates led to increased synthesis of bio-methane compared with low-fat substrates (Ebner et al., 2016).

A typical kitchen waste contains 60–80% moisture, 3–5% ash, 40–60% carbohydrate, 18–30% volatiles, 10–30% protein, 15–40% fat, and 45–65% carbon. Fish and meat meals have three times the protein and moisture content of wheat meals, while wheat meals have a high composition of volatile matter and carbs (88–92%) (Palaniveloo et al., 2020). In 2012, researchers in the United States discovered the nutritional content of food waste. On average, food waste consumed 1,217 kcal of energy per day, 146.4 grams of carbohydrate, 32.8 grams of protein, 286.1 milligrams of calcium, 85.0 milligrams of magnesium, 450.3 milligrams phosphorus, 880.2 milligrams potassium, 264.2 milligrams sodium, and 3.9 milligrams zinc (Spiker et al., 2017). Various types of food waste were examined for potential value-added by-products. Many beneficial derivatives from fruits, vegetables, meat, and dairy products can be summarized in Figures 6A and 6B (Mirabella et al., 2014).

For the success of various valuation methodologies, various elements should be optimized. The physical/chemical features of the feedstock, like humidity, volatile materials and nutrient structure, pH and size, and configuration, determine, for instance, anaerobic digestion and biogas production (Ho & Chu, 2019). The addition of microorganisms plays a vital part in the anaerobic digestion of organic matter. The bacteria cause the effective breakup of complex organic compounds by a sequence of biochemical processes to create methane. By conversion of the organic waste part, biogas was created by the anaerobic digestion process. It contains numerous gases like methane, carbon dioxide, and others (60–65%). The organic component of the various waste (e.g., agricultural, kitchen, and local heavy waste) was used as a substratum for energy recovery in biogas. Besides process sustainability, the odor, sludge levels, and pathogens were minimized; digestate can even be used as a fertilizer. Finally, separating food waste by source and type can assist reduce the source and time-based composition variability, enhancing the consistency retrieval success of food waste products (Liaquat et al., 2017).

Moisture Content of Food Waste

Moisture content is one of the most significant performance variables affecting managing output and processing. There are two options to determine the moisture content, direct and indirect: In the first step, we can measure the moisture content after extracting the water. However, specific physical or chemical characteristics of the grain that influence moisture content are needed in the other portions, and then the humidity can be determined. The level of humidity is usually calculated either Based on a wet basis (% wb) or a dry basis (% db) (Obi et al., 2016). Moisture content (X_i) and drying rate (dX_i/dt) were determined using respective Equations 1 and 2 (Ismail et al., 2020b).

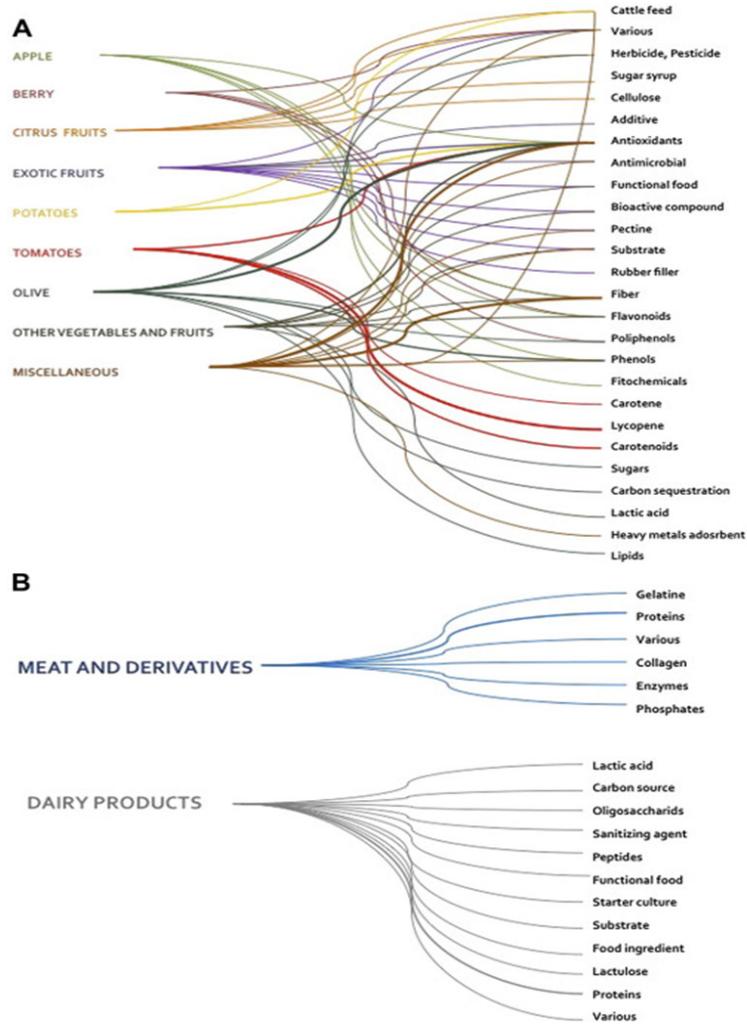


Figure 6. A. Valuable compounds derived from fruits and vegetables. B. Valuable compounds are derivable from meat and dairy (Mirabella et al., 2014).

$$\text{Moisture content (\%wb)}, X_i = \frac{M_i - M_{bd}}{M_{bd}} \quad (1)$$

$$\text{Moisture content (\%db)}, \frac{dX_i}{d_t} = \frac{M_i - M_{i+1}}{t_i - t_{i+1}} \quad (2)$$

Where: M presents the mass sample, t shows the drying time at (s), t_i is drying time in a time of (i), and bd presents the bone dry solid. The free humidity X vs time t graph looks like this.

One of the important factors is estimating the amount of moisture content in the organic material before drying (Ismail et al., 2020b). Because of the high moisture content, organic waste has a low calorific value, resulting in incomplete combustion and the release of dioxins and other toxic gases. Researchers discovered wet basis moisture content of food waste was 92.2%, and for leftovers, it was 64.4% (Liu et al., 2016). Furthermore, the moisture content of food waste gathered from various sources was typically over 65%, while the morning market food waste was 90% (Chua et al., 2019).

Moisture affects microbial activity and physical structure, affecting food waste biodegradation. In addition, moisture content impacts decomposition rate and is important in composting. Composting requires wet basis moisture concentrations of 25 to 80%, with 50 to 70% ideal (Makan et al., 2013). The high moisture content of food waste limits its ability to recover energy. However, there are currently various alternative strategies to reduce the moisture content of food waste, such as bio-drying as a zero-leachate generation technology. Also, dehydration near the source of food waste formation may benefit from reducing mass and transit costs for downstream upcycling processes like animal feed manufacturing, composting, or anaerobic digestion (Schroeder et al., 2020).

DRYING METHOD

Drying technologies have been implemented in developed and developing countries to enhance the efficiency of urban food waste to use solid wastes for sustainable sources of vitality, minimize fossil fuel dependencies, and ensure that waste is cleaner at sites. In developed and emerging countries, drying methods have been used to increase the performance of food waste to use solid wastes for environmental liveliness, make trash safer, and reduce the reliance on fossil fuels on-site (Tun & Juchelková, 2019). The most popular methods were freeze-drying, microwave, sun, and thermal drying.

Freeze-Drying

Freeze drying is commonly used to process dehydrated high-quality fruit and vegetables. The water solid-state, low temperatures, and the process of moisture sublimation during freeze-dry protection help protect the primary structure and form of the products, which also have low bulk density, high porosity, and stronger rehydration (Taylor et al., 2011).

Freeze-drying requires extracting a liquid formulation solvent, which is frozen and exposed to low pressure for the solvents' sublimation and then subject to a final desorption process for removing the unfrozen solvent (Nowak & Jakubczyk, 2020). Therefore, the method of drying can be divided into two stages: sublimation (primary drying) and desorption (secondary drying) because two similarly critical processes occur, i.e., freezing, in which almost all solvent has been transformed into a frozen solid, and drying process, in which the mixture eliminates almost all solvent (frozen or unfrozen) (Assegehegn et al., 2020).

Frozen, which consolidates food, is the first stage in freeze-drying. The freezing rate is crucial for the development and size of ice crystals—slow freezing is greater than or reciprocally than ice particles. In the food business, freeze-drying is limited to high-added-value products such as coffee, ingredients for ready-to-eat dishes (fruits and vegetables, meat and fish), and aromatic herbs (Wang et al., 2014). The MEADOW processor Freeze Drying Solid Waste has extensively treated solid waste items generated during extended human-crewed space flights. Freeze drying and vacuum drying were the principal drying methods explored. In either mode, a Peltier condenser gathers water vapor from the waste and converts it back into relatively clean water. The water activity level of the dried waste product is lower than that required for microorganisms to maintain metabolic activity. For the treated waste to remain stable, it must be stored and confined in a manner that prevents water from being reabsorbed (Wheeler et al., 2007). Many studies have demonstrated that freeze-drying is a good way to reduce food waste. However, they claimed that, due to its high cost, freeze-drying was an unsuitable method of waste disposal. Nevertheless, the special condition will help to reduce wastage.

Microwave Drying

Microwave drying is becoming increasingly popular as an effective drying process for different foodstuffs, such as fruits, vegetables, snack products, and dairy products (Wang et al., 2004). The drying of microwaves depends on the variation of electromagnetic energy into thermal energy by the polar molecules of the material. The motion between the bipolar molecules induces heating of the material; in this orientation method, the material generates ample heat to evaporate humidity from the mass. In addition, the bipolar molecules' rapid mobility provides a complete pressurizing gradient that facilitates the quick flow of liquid water and vapor to the material surface, which results in speedy drying without overheating the atmosphere (Khodifad & Dhamsaniya, 2020).

The microwave drying method is based on a particular volumetric heat mode facilitated by electromagnetic radiation of 915 MHz or 2450 MHz. The loss reaction of a dielectric component causes a fast power interaction with or dried up moisture and signal strength. A significant dropping time reduction usually involves the enhanced quality of the substance, making it a good drying technique (Feng et al., 2012). Microwave heating depends on the weight, weight, shape, size, simple thermal power, the composition of the sample, and dielectricity (Durance & Yaghmaee, 2011). Microwave drying is a revolutionary method; it is possible to dry food waste rapidly and efficiently using microwave drying techniques because there are several advantages to using microwaves over conventional drying methods, such as faster drying and lower energy use. In addition, the pore structure of raw materials improves from microwave drying (Liu et al., 2016). In general, using a microwave reactor to break down food waste has the potential to produce alternative fuels

for energy generation and transportation. In addition, microwave drying has aided in the thermal decomposition of trash and is a viable method for decreasing food waste (Anis et al., 2018).

Sun Drying

Sun drying is a traditional method of direct solar radiation and natural wind force. Most developing countries use open-air (sun drying) systems for drying agricultural products. It is a low-cost drying method for preserving and storing agricultural products (Dhumne, 2016). There are two types of direct and indirect solar dryers. Direct solar drying is called “sun drying” the product is heated directly by the sun’s rays, and the natural circulation of air removes moisture due to density differences. The hot air is transmitted over the drying material at indirect solar dryers and often passes through a solar collector, releasing moisture from the substance. The sun is the most powerful carbon-free energy source world has ever known. Sun-drying with free solar energy might also help save money on drying (Bennamoun & Li, 2018). Sun drying is a low-cost process, but this commodity is less consistent because of the possibility of bacterial runoff, pollen, birds, animals, and rainfall. The ability to exercise in ultraviolet radiation also adds to the lack and dryness of supplements, nude testing, and unacceptable color changes (Hegde et al., 2015).

Thermal Drying

One of the effective ways to improve materials quality is to use thermal drying. This mechanism can use for periodic drying. Therefore, products’ thermal and physical attributes for suitable dryer designs, such as water activity, moisture diffusion, heat transfer, and primary energy usage, become essential (Ismail et al., 2021). The thermal drying method can be classified into three types: atmospheric drying, low air drying, and hot air drying. Low-temperature processing offers lower corrosion risks and low capital and operating costs, but the process is sluggish and does not work well under some climatic conditions. Typically carried out at 15 to 501°C temperatures, the air is heated around 1 to 61°C (Al-kharabsheh & Goswami, 2004). On the other hand, low air drying is an option for drying unstable materials, mostly leafy vegetables, and it is the best way to preserve enough ascorbic acid (Venkatachalam et al., 2020).

Hot air drying has been the superlative widely used industrial drying process. Heat is transmitted through convection from hot air to the crop, and the liquid is evaporated by convection to air. Hot air drying is a viable means of fully exploiting the bioactive compounds’ possible physiological findings and the analysis of fruit waste. It is a complex process involving heat and mass transfer under the temporary structure, defined by several factors, including internal feeding and physicochemical properties and external properties, including temperature, flow, and drying humidity (Sozzi et al., 2021).

Hot air as a heat carrier is often used in a drying operation. The concept uses air heat and mass transfer to push into a closing structure. The famous dryer is designed to retain air uniformly over perforated racks where food waste spreads in the dryer chamber. Hot air drying is the least expensive dryer (Salim et al., 2017). Special low-cost thermal dryers such as the Schematic dryer and the cabinet drier with a traditional tray are used to dry food waste to minimize leachate and greenhouse gas. The main advantage of the cabinet dryer is that it can save energy as the temperature and airflow rose. According to the report, when food waste was subjected to dry at 70°C with 2m/s air velocity, it took just 120 minutes to dry, and the total amount of energy needed to dry food waste was found to be between 119.62 and 59.41 kWh (Branch & Borghei, 2021). The comparative findings from various drying processes recommended the thermal drying method since the thermal drying techniques for solid industrial items are accessible, efficient, and economical. Table 1 summarizes the drying processes mentioned in this section and their influence on the quality parameter.

Table 1

The characteristics of a few standard drying processes

Drying Method	Type of Feed	Mechanism	Advantage	Disadvantage	References
Freeze Drying	All types of food	Two steps: first, freezing water out of the source material; heating the frozen solid to induce the moisture sublimation mechanism	Reduce chemical compound changes; Minimum reduction and soluble solid transition; Helps avoid oxidation damage; Retention of volatile chemicals.	The cost of facilities is relatively high; the procedure is slow and costly.	Taylor et al (2011) Wang et al. (2014) Nowak and Jakubczyk (2020) Bhatta et al. (2020)
Microwave Drying	Different foodstuffs (fruit and vegetable)	The drying method is based on a particular volumetric heat mode facilitated by electromagnetic radiation of 915 MHz or 2450 MHz. The loss reaction of a dielectric component causes a fast power interaction with or dried up moisture and signal strength.	Quick heating in volume; Greater flow speed; short drying time; improved product quality; Shrinking energy uses a lower cost of operation.	High price for dryers at the start; Severe fragrance reduction and adverse perceptual alterations;	Wang et al. (2004) Changrue and Raghavan (2006) Feng et al. (2012)

Table 1 (Continue)

Drying Method	Type of Feed	Mechanism	Advantage	Disadvantage	References
Sun Drying	Fruits, vegetables	Two types: direct and indirect; the product is heated directly by the sun's rays. The hot air is transmitted over the drying material at indirect solar dryers and often passes through a solar collector, releasing moisture from the substance.	Low-cost drying method for preserving and storing agricultural products; The most significant carbon-free supply of energy worldwide	Extended drying time. The risk of bacterial rush in birds, animals, and plumage is less constant.	Dhumne et al. (2016) Maragkaki et al. (2016) Hegde et al. (2015) Bennamoun and Li (2018)
Thermal Drying	Fruits, vegetables	Two processes occur concurrently during thermal drying: Heat transfer to raise the temperature of the wet material and evaporate the moisture content.	Contribute to increased energy efficiency and decreased thermal drying time	Two main disadvantages are low costs and a long drying time in the lower rate cycle. In addition, the low thermal conductivity during this period limits the standard heat transmission to the inner parts of the material.	Cam et al. (2017) Rahman ad Perera (2007) Wang et al. (2004)

EFFECT OF DEHYDRATION ON QUALITY OF FOOD WASTE FOR COMPOSTING

In recent decades, excess supply has resulted in increased waste nature generations in urban and industrial installations. Implementing sustainable policies to conserve environmental sustainability is essential in this respect. That is a renewable technology for the disposal of organic pollution and the regulation by scale and volume removal of smelly greenhouse gas for various organic waste. It is a highly successful process in terms of current costs, compost generation, and decreased air and water pollution by treating process parameters more effectively than other waste management options (Dhamodharan et al., 2019).

Composting is one of the most viable options for converting the natural fracture into a beneficial organic fertilizer known as compost among the advanced techniques (Artola et al., 2009). The commodity from the composting process can be used as a bio-fertilizer containing plant-based and livestock-based nutrients such as potassium and nitrated

(sodium, calcium, magnesium, and chloride). In addition, Composting is used to recycle organic waste into a marketable product in advanced waste management techniques (Abdullah et al., 2018).

Composting is a simple process for producing a full-nutrient product from low-cost organic resources, which has a high-value commodity in agriculture and positively influences soil health (Pergola et al., 2020). Sustainable solid waste management is essential since it provides a means of handling biodegradable waste fractions. It has been used streamlined and promoted as a safe conversion of future organic waste to sustainably manageable solid waste (Ishola & Ishola, 2019). Some arguments say that composting reduces the overall greenhouse gas emissions, and others claim that it increases the amount of carbon dioxide (CO₂) and greenhouse gas (GHGs) emissions.

Composting is a convenient way to get rid of organic waste, but applying industrial drying methods before composting can preserve organic material quality and slash the time it takes to produce it (Ermolaev et al., 2019). Nowadays, significant factor considerations must be considered before implementing food waste disposal units as a wide-ranging waste management option, as are their related emissions of GHGs and other environmental damage. Continued research and development in food waste could better manage this choice to become a sustainable alternative to landfilling (Iacovidou et al., 2012). Dehydration of food waste is a technique that dewateres food waste to produce a low moisture product, greatly reducing the kitchen's waste content and waste volume (Sotiropoulos et al., 2015).

The product created from thermal dehydration of food waste may be more suited for soil application than compost and aerobic digestion because dehydrated food waste has a high concentration of Carbon, Nitrogen, and nutrition content that plant needs. In addition, using plant-growth-promoting microbes in dehydrated food waste, such as *Azospirillum*, can improve nutrient availability and soil health (Mahmood et al., 2019). Due to the characteristics of dehydrated food waste, it can be used on various crops. Organic waste that has been dehydrated provides several advantages, including being light, dry, odorless, and biologically inert. Organic biofertilizer or livestock feed is made from the high nutrient content of this product. Before putting dried food waste into the soil, it must be composted or buried. These methods will aid in eliminating phytotoxins, anoxia, salinity, and water repellence (O'Connor et al., 2022). The dehydrated organic raw material was also discovered to slow decomposition and decrease odors, allowing for less frequent trash collection. Furthermore, it was determined that the dried food waste's physical and chemical properties make it appropriate as a feedstock for value-added items such as compost and biofuels (Schroeder et al., 2020).

Compost made from dehydrated food waste requires much oxygen to function properly. When young, composted products are applied to soils, their high oxygen demand can stymie seed germination and plant growth. When water is added to dehydrated food waste, the decomposition process begins. This method can store food waste for an extended

period without losing nutritional value. To begin the composting process, add water to the dehydrated food waste (O'Connor et al., 2022). Food waste has a reduced moisture content, which can assist in controlling the water content of the organic substrate throughout the composting process. Also, purified dry food waste is pathogen-free, contains an appropriate amount of carbon and nitrogen, and is ideal for use as a compost feedstock. Dehydration not only retains the nutritional value of the raw material but also boosts productivity. Additionally, the remaining material may be archived for future reference. Otherwise, food waste will decay and become inert, eventually disappearing due to biodegradation (Loizidou, 2015).

According to the studies, due to some challenges dried food waste cannot be applied directly to the soil. Organic molecules made from dehydrated food waste have a high oxygen demand and are extremely phytotoxic, which can harm plant growth. Due to the increased oxygen demand, the nitrogen content of immature compost decreases when it is added to the soil. For a short period, food waste that has been dried can be stored and will not break down unless it is introduced to water. Adding water to the product completes the composting process since it sustains many microorganisms (O'Connor et al., 2021; Toundou et al., 2021).

Since dehydration of food waste has many advantages, it also has several drawbacks, one of which is the risk of financial loss. Therefore, using a laboratory drying system to treat food waste will have a significant economic impact. However, thermally GAIA Food Waste Dryer has recently overcome this difficulty by giving a non-biological, effective alternative to food waste dehydrators in terms of volume and bulk reduction (Dhar, 2016). Also, sun drying was the most cost-effective drying method, but it has not been widely suggested for food waste management due to several limitations, including the inability to control sun ray temperature and the inability to control product quality (Tony & Tayeb, 2011).

Physicochemical Properties of Dehydrated Food Waste for Compost

A proposed answer to the worldwide food waste disposal issue is to use food waste as biofertilizers. Dry food waste products had excellent electrical conductivity ($4.83\text{--}7.64\text{ mS cm}^{-1}$) and mild acidity ($4.7\text{--}5.1$) (O'Connor et al., 2022). pH is one of the critical characteristics of soil tests. It defines the soil considerably more than simply whether it is acidic or basic. The number of critical nutrients and toxicity of other elements can be anticipated due to their known pH relationship (Twain, 2020). Since Food waste is usually acidic in the early stages due to the creation of organic acids, it depends on the level of food waste and composition. Overall, the pH value in wet food waste ranged from $5\text{--}5.5$ to $6.9\text{--}7.3$. Due to microbial activity, the pH value will rise from 6 to 9% after composting (Kannah et al., 2020). In terms of acidity, the pH value of dried waste is very close to wet food waste. Thus, many experts believe that dry food waste can be used as raw material for composting.

Electrical conductivity stands for EC. It measures how well a substance drives electricity for the solution. The higher the conductivity, the more electrolytes. Furthermore, the greater the load, the higher the conductivity each particle carries. According to experts, the ideal EC values for agriculture compost fertilizer range from 2.0 to 3.5 mS/cm (Zaha et al., 2013). Compost with high EC rates might be very rich in nutrients since minerals have an outstanding performance. EC measures are supported by all ions charged into the soil solution. Some essential micronutrients, including NO_3 , NH_3 , and K, have reduced the requirement for additional fertilizers by their presence in compost. However, EC monitoring does not play the same function in soil feeding, and it is thus always suggested to follow the manufacturer's instructions (Crohn, 2016).

The effects of organic disposal on crop productivity depend on the material's different chemical compositions and sources. The main factors determine the composition and use of organic waste, such as C, N, K, P, and accessible biological germination experiments (Guerra-Rodríguez et al., 2001). As a soil-plant system, compost application is one of the most successful strategies. It might provide soil-plant systems with nutrients, stable organic material, and beneficial microorganisms (Haouas et al., 2021).

According to previous research findings, dried food waste can also be used as a raw material for composting. Because it is nutrient-rich, the amount of nutrient content in dehydrated food waste: Carbon 48.3%, nitrogen 3.26%, C/N 14.8%, and Sulphur 0.23% (Mahmood et al., 2019). One of the most important recent using dried food waste in compost to reduce total CH_4 emissions. Adjusting the moisture level during composting might help minimize greenhouse gas emissions and the environmental impact: For example, greenhouse gas emissions from composting food waste at two different humidity levels of 44 and 66% were examined over 20 days at 50°C and 16% oxygen. The total estimated Methane emission during the composting process was $35 \text{ g CH}_4\text{-C kg}^{-1}$ initial C at a moisture content of 66%. However, it was reduced to $0.04 \text{ g CH}_4\text{-C kg}^{-1}$ initial C at a moisture content of 44%. (Ermolaev et al., 2019).

Other parameters that need to be considered include the oxygen content and bulk density. Aeration is required to supply oxygen to microorganisms and promote the degradation of organic matter. Dehydrogenase activity is the easiest, fastest, and cheapest way to assess the stability and maturity of compost in ATP content and microbial bio-mass process (Tiquia, 2005). Composting alters the bioavailability of copper and phosphorus due to the decomposition of organic matter (microbial activity % and Humic acid/Fulvic acid ratio). C/N ratio of 22 allowed for successfully immobilizing copper, zinc, and phosphorus in compost (Wang et al., 2019). Aerobic composting is a decomposition process in which organic substrates break down in the presence of oxygen, producing carbon dioxide. Composting works best with a C/N ratio of 30–40% and a moisture of 50–65%, while new research suggests that C/N ratios as low as 20% may also be useful (Firdaus et al., 2018).

According to Schroeder et al. (2020), dried food waste with lower moisture content than ideal poses a significant issue; however, it is possible to compose this type of dry organic material by mixing it with other organic waste. For instance, combining yard waste high in nitrogen and moisture with raw food waste low in carbon can result in a healthful compost pile. Overall, Low humidity food waste, besides preserving nutrients, allows for the potential to produce a beneficial practical microbial photo-beneficial cum that may provide more advantages. Thus, managing dry organic waste is much easier than wet, and It can generate high-value products such as bio compost, biogas, and bioethanol (Loizidou, 2015). Table 2 Summarizes the optimum values of the various parameters of the composting processes.

Additionally, dehydrated food waste can be used as animal feed in nations where it is legal to feed animals with food waste. However, it must be processed very carefully before being fed. For example, the United States must heat food waste at 100°C for 30 minutes. Also, in Japan, food waste processed for use in animal feed is referred to as “ecofeed.” Preferably, raw materials containing uncooked meat should be heated at 70°C for 30 minutes or 80°C for 3 minutes before being used to make “ecofeed.” Because of the drying process, Ecofeed can be preserved for a longer time (Georganas et al., 2020). Compared to turning wet food waste into dry pig feed, dry feed poses fewer environmental and health risks. However, the higher fossil fuel inputs required to dehydrate municipal food wastes represent a large percentage of the difference between wet and dry diets shown in Figure 7. Dehydrating food wastes to produce dry feed involves gas and electricity due to their high-water content, which ranges from 65 to 80%. The finding indicated that food waste should always be fed as wet feed. Because these two approaches, wet and dry, are appropriate for various pig production systems (Salemdeeb et al., 2017b). Table 3 summarizes the components of dried food waste used as a feedstock for composting.

Table 2

Summarizes the optimal values of the various parameters to be carried out after composting process

Parameter	Optimum values	References
pH	Between 5.5–7.3	Kannah et al. (2020)
Electrical conductivity (EC)	2.89 ds/m	Khan and Ishaq (2011)
Total of C	46.8%	Jiménez and García (1992)
Total of N	0.4–3.5%	Harrison (2008)
Total of S	0.25–0.8%	Bahtiar et al. (2017)
Moisture content	50–65%	Firdaus et al. (2018)
C: N	25–30:1	Firdaus et al. (2018)

Table 3

Summarizes the components of dried food waste used as a feedstock for composting

Components	wt.%, dry basis
pH	4.7–5.1
electrical conductivity	4.83–7.64
Carbon	48.3
Nitrogen	3.26
C: N	14.8
Sulfur	0.23

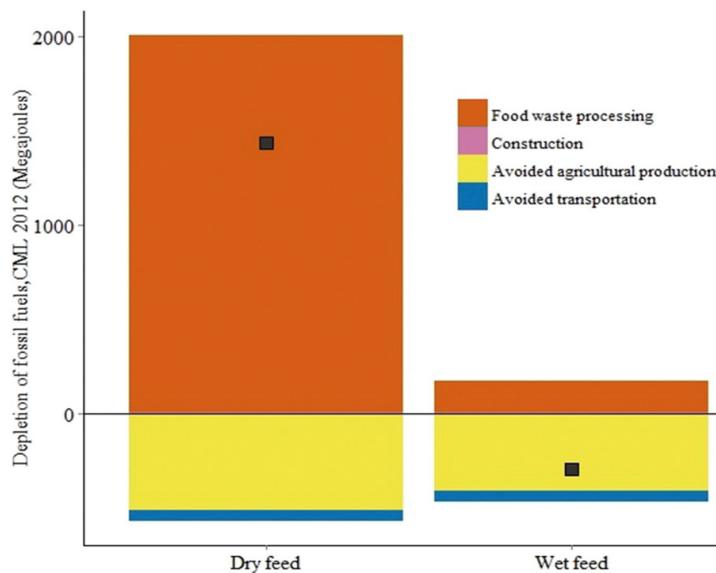


Figure 7. Manufacturing wet and dry feed using fossil fuel energy (MJ) (Salemdeeb et al., 2017)

Trench Composting of Dehydrated Food Waste

Trenching is a means to handle organic waste directly in the garden by burying it. It is a good way for plant roots to deposit nutrients in the soil. It also promotes the growth of deep root systems that provide water. Trenching for the plants generates nutrient-rich low-floor areas. The composting process is slower than in a managed backyard scenario, but more nitrogen is maintained throughout the process. For decomposing organic waste, trench compost relies only on the soil's decomposition of microorganisms and bugs. Since the substance is buried, it derives its water and oxygen from the earth underneath it. Additionally, the soil particles cling to decaying nitrogen, preventing it from escaping as methane and unpleasant odors (Compost Education Centre, 2010). Since the composting

process is anaerobic, the lack of oxygen makes it unsuitable for high-humidity organic waste, leading to pathogens in the final product. In this case, trench compost is a good option for dry waste. A trench system attracts a maximum of 54°C in 15 days, lasts for five days until day 21, and then begins settling. This procedure takes around 30 days. Composting of trenches is comparable to pit compost. Composting in a pit necessitates the excavation of a hole or trench. This technique is effective at concealing decomposing organic materials. It is particularly beneficial against termite attacks since most species dwell above ground level (Teresita et al., 2021).

The application of trench compost improves the soil quality and air retention, nutrients, and moisture, leading to healthy and flourishing plants (Brock et al., 2021). The disturbance and heterogeneity generated by human activities are substantially different in urban soil. By adding organic waste, significant increases in soil organic carbon have been identified, and soil physical and chemical characteristics improved and produced large amounts of organic carbon. In improving product quality and storage life after harvesting and eliminating product waste, the mechanical properties of garden produce play a significant role. The recent movement towards producing high-quality products and enhancing soil quality has been gradually using organic fertilizers (Jahanbakhshi & Kheiralipour, 2019). According to previous research, dehydrated food waste should be composted before being used as an organic fertilizer in the soil. Trench compost was recommended for dehydrated organic waste when comparing composting processes.

CONCLUSION

The use of food wastes as biofertilizers is a potential environmental solution to the current global food-waste disposal problem. The researchers concluded that dehydration of food waste is important and has a positive effect on reducing food waste during the management period. Sun-drying is the most cost-effective drying method, but it is not frequently recommended for food waste management due to some drawbacks, such as the inability to control sunray temperature and end-product quality. Thermal drying is an alternative method for sun-drying, which can dry food waste with high quality quickly. Thermal dryers, such as the GAIA GC-300 dryer and cabinet dryer fitted with a standard tray, provide a non-biological alternative to food waste dehydrators that effectively reduce the volume of waste generated. According to previous researchers, dry food waste can be used as a raw material for composting because it is high in nutrients and comparable to wet food waste. Dehydrated food waste contains 48.3% carbon, 3.26% nitrogen, 0.23% sulfur, 14.8% of carbon-nitrogen ratio, and it is slightly acidic (4.7–5.1) with high levels of electrical conductivity (EC) (4.83–7.64 mS cm⁻¹). Nevertheless, using dried food waste with a lower moisture content than the ideal moisture of 50–65 in the composting process has a significant problem; however, this type of dry organic material can be composted by

mixing it with other organic waste. For instance, combining yard waste high in nitrogen and low moisture with raw food waste low in carbon and high moisture can result in a healthful compost pile. Adjusting the moisture level before composting might help minimize greenhouse gas emissions and the environmental impact; the result of two different humidity levels of 44 and 66% throughout the composting process was shown. The total estimated methane output was 35 g CH₄-C kg⁻¹ initial C, which was reduced to 0.04 g CH₄-C kg⁻¹ initial C. As a result, the review recommended that food waste be dried to an adequate moisture level before composting to reduce greenhouse gas emissions, odor, and leachate.

ACKNOWLEDGEMENT

The study has been financially supported by Knowledge Transfer Grant Scheme (KTGS), vote number: 9437581 and Knowledge Transfer Programme (KTP), vote number: 6228161.

REFERENCE

- Abdullah, M., Rosmadi, H. A., Azman, N. Q. M. K., Sebera, Q. U., Puteh, M. H., Muhamad, A., & Zaiton, S. N. A. (2018). Effective drying method in the utilization of food waste into compost materials using effective microbe (EM). In *AIP Conference Proceedings* (Vol. 2030, No. 1, p. 020120). AIP Publishing LLC. <https://doi.org/10.1063/1.5066761>
- Al-Domi, H., Al-Rawajfeh, H., Aboyoucif, F., Yaghi, S., Mashal, R., & Fakhoury, J. (2011). Determining and addressing food plate waste in a group of students at the University of Jordan. *Pakistan Journal of Nutrition, 10*(9), 871-878. <https://doi.org/10.3923/pjn.2011.871.878>
- Al-kharabsheh, S., & Goswami, D. Y. (2004). *Solar Distillation and Drying*. University of Florida. <https://doi.org/10.1016/B0-12-176480-X/00319-3>
- Anis, S., Kurniawan, Y. A., Sumbodo, W., Alhakim, R., & Lestari, S. E. (2018). Thermal characteristics of microwave reactor for pyrolysis of food waste. *Journal of Physical Science, 29*, 1-13. <https://doi.org/10.21315/jps2018.29.s2.1>
- Artola, A., Barrera, R., Font, X., Gabriel, D., Gea, T., Mudhoo, A., & Sánchez, A. (2009). Composting from a sustainable point of view: Respirometric indices as key parameter. *Dynamic Soil, Dynamic Plant, 3*(1), 1-16.
- Arumugam, V., Abdullah, I., Yusoff, I. S., Abdullah, N. L., Tahir, R. M., Nasir, A. M., Omar, A. E., & Ismail, M. H. (2021). The impact of COVID-19 on solid waste generation in the perspectives of socioeconomic and people's behavior: A case study in Serdang, Malaysia. *Sustainability, 13*(23), Article 13045. <https://doi.org/10.3390/su132313045>
- Assegehegn, G., Brito-del la Fuente, E. J.M., & Gallegos, C. (2020). Freeze-drying: A relevant unit operation in the manufacture of foods, nutritional products, and pharmaceuticals. *Advances in Food and Nutrition Research, 93*, 1-58. <https://doi.org/10.1016/bs.afnr.2020.04.001>
- Bahtiar, S. A., Muayyad, A., Ulfaningtias, L., Anggara, J., Priscilla, C., & Miswar, M. (2017). Compost use banana weevil (*Musa Acuminata*) to boost growth and content of sugar sweet corn (*Zea mays* L. Saccharata). *Journal of Agricultural Science, 14*(1), 18-22. <https://doi.org/10.32528/agr.v14i1.405>

- Bennamoun, L., & Li, J. (2018). Drying process of food: Fundamental aspects and mathematical modeling. In *Natural and artificial flavoring agents and food dyes* (pp. 29-82). Academic Press. <https://doi.org/10.1016/B978-0-12-811518-3.00002-8>
- Bhatta, S., Janezic, T. S., & Ratti, C. (2020). Freeze-drying of plant-based foods. *Foods*, 9(1), 1-22. <https://doi.org/10.3390/foods9010087>
- Blakeney, M. (2019). *Food loss and food waste: Causes and solutions*. Edward Elgar Publishing. <https://doi.org/10.4337/9781788975391>
- Branch, E., & Borghei, A. M. (2021). *Evaluate the drying of food waste using cabinet dryer*. ResearchSquare. <https://doi.org/10.21203/rs.3.rs-874515/v1>
- Brock, C., Oltmanns, M., Matthes, C., Schmehe, B., Schaaf, H., Burghardt, D., Horst, H., & Spieß, H. (2021). Compost as an option for sustainable crop production at low stocking rates in organic farming. *Agronomy*, 11, 1-17. <https://doi.org/10.3390/agronomy11061078>
- Cam, I. B., Gulmez, H. B., Eroglu, E., & Topuz, A. (2017). Strawberry drying: Development of a closed-cycle modified atmosphere drying system for food products and the performance evaluation of a case study. *Drying Technology*, 36(12), 1460-1473. <https://doi.org/10.1080/07373937.2017.1409233>
- Changrue, V., & Raghavan, V. G. S. (2006). Microwave drying of fruits and vegetables. *Stewart Postharvest Review*, 2(6), 1-7. <https://doi.org/10.2212/spr.2006.6.4>
- Chauhan, C., Dhir, A., Akram, M. U., & Salo, J. (2021). Food loss and waste in food supply chains. A systematic literature review and framework development approach. *Journal of Cleaner Production*, 295, Article 126438. <https://doi.org/10.1016/j.jclepro.2021.126438>
- Chua, G. K., Tan, F. H. Y., Chew, F. N., & Mohd-Hairul, A. R. (2019). Nutrients content of food wastes from different sources and its pre-treatment. In *AIP Conference Proceedings* (Vol. 2124, No. 1, p. 020031). AIP Publishing LLC. <https://doi.org/10.1063/1.5117091>
- Compost Education Centre. (2010). *Trench composting* (Factsheet Series). <https://compost.bc.ca/wp-content/uploads/2021/03/5-Trenching.pdf>
- Crohn, D. M. (2016). *Assessing compost quality for agriculture*. California Digital Library. <https://doi.org/10.3733/ucanr.8514>
- Dhamodharan, K., Sudharsan, V., Veluchamy, C., Pugazhendhi, A., & Rajendran, K. (2019). Science of the total environment emission of volatile organic compounds from composting : A review on assessment, treatment and perspectives. *Science of the Total Environment*, 695, Article 133725. <https://doi.org/10.1016/j.scitotenv.2019.133725>
- Dhar, A. (2016). *Evaluation of food waste diversion potential and economics of using food waste dehydrators* (Doctoral dissertation). The University of Texas, USA.
- Dhumne, L. R., Bipte, V. H., & Jibhkate, Y. M. (2016). Solar dryers for drying agricultural products. *International Journal of Engineering Research*, 3(2), 80-84.
- Durance, T., & Yaghmaee, P. (2011). Microwave dehydration of food and food ingredients. *Comprehensive Biotechnology*, 2(1), 617-628. <https://doi.org/10.1016/B978-0-08-088504-9.00306-8>

- Ebner, J. H., Labatut, R. A., Lodge, J. S., Williamson, A. A., & Trabold, T. A. (2016). Anaerobic co-digestion of commercial food waste and dairy manure: Characterizing biochemical parameters and synergistic effects. *Waste Management*, *52*, 286-294. <https://doi.org/10.1016/j.wasman.2016.03.046>
- Ermolaev, E., Sundberg, C., Pell, M., Smårs, S., & Jönsson, H. (2019). Effects of moisture on emissions of methane, nitrous oxide and carbon dioxide from food and garden waste composting. *Journal of Cleaner Production*, *240*, Article 118165. <https://doi.org/10.1016/j.jclepro.2019.118165>
- Feng, H., Yin, Y., & Tang, J. (2012). Microwave drying of food and agricultural materials: Basics and heat and mass microwave drying of food and agricultural materials. *Food Engineering Reviews*, *4*(2), 89-106. <https://doi.org/10.1007/s12393-012-9048-x>
- Firdaus, A. R. M., Samah, M. A. A., & Hamid, K. B. A. (2018). CHNS analysis towards food waste in composting. *Journal CleanWAS*, *1*(1), 06-10. <https://doi.org/10.26480/jcleanwas.01.2018.06.10>
- Fisgativa, H., Tremier, A., & Dabert, P. (2016). Characterizing the variability of food waste quality : A need for efficient valorisation through anaerobic digestion. *Waste Management*, *50*, 264-274. <https://doi.org/10.1016/j.wasman.2016.01.041>
- Galanakis, C. M. (2012). Recovery of high added-value components from food wastes: Conventional, emerging technologies and commercialized applications. *Trends in Food Science and Technology*, *26*(2), 68-87. <https://doi.org/10.1016/j.tifs.2012.03.003>
- Georganas, A., Giamouri, E., Pappas, A. C., Papadomichelakis, G., Galliou, F., Manios, T., Tsiplakou, E., Fegeros, K., & Zervas, G. (2020). Bioactive compounds in food waste: A review on the transformation of food waste to animal feed. *Foods*, *9*(3), 1-18. <https://doi.org/10.3390/foods9030291>
- Guerra-Rodríguez, E., Vázquez, M., & Diaz-Raviña, M. (2001). Dynamics of physicochemical and biological parameters during the co-composting of chestnut burr/leaf litter with solid poultry manure. *Journal of the Science of Food and Agriculture*, *81*(7), 648-652. <https://doi.org/10.1002/jsfa.866>
- Hall, M. (2016). *Techno-environmental analysis of generating animal feed from wasted food products*. Rochester Institute of Technology.
- Haouas, A., El Modafar, C., Douira, A., Ibsouda-Koraichi, S., Filali-Maltouf, A., Moukhli, A., & Amir, S. (2021). Evaluation of the nutrients cycle, humification process, and agronomic efficiency of organic wastes composting enriched with phosphate sludge. *Journal of Cleaner Production*, *302*, Article 127051. <https://doi.org/10.1016/j.jclepro.2021.127051>
- Harrison, R. B. (2008). Composting and formation of humic substances. *Encyclopedia of Ecology*, *5*, 713-719. <https://doi.org/10.1016/B978-008045405-4.00262-7>
- Hegde, V. N., Hosur, V. S., Rathod, S. K., Harsoor, P. A., & Narayana, K. B. (2015). Design, fabrication and performance evaluation of solar dryer for banana. *Energy, Sustainability and Society*, *5*(1), 1-12. <https://doi.org/10.1186/s13705-015-0052-x>
- Ho, K. S., & Chu, L. M. (2019). Characterization of food waste from different sources in Hong Kong. *Journal of the Air and Waste Management Association*, *69*(3), 277-288. <https://doi.org/10.1080/10962247.2018.1526138>

- Iacovidou, E., Ohandja, D. G., Gronow, J., & Voulvoulis, N. (2012). The household use of food waste disposal units as a waste management option: A review. *Critical Reviews in Environmental Science and Technology*, 42(14), 1485-1508. <https://doi.org/10.1080/10643389.2011.556897>
- Ishola, T. M., & Ishola, E. T. (2019). Composting and sustainable development. In W. L. Filho (Ed.) *Encyclopedia of Sustainability in Higher Education* (pp. 1-8). Springer. https://doi.org/10.1007/978-3-319-63951-2_122-1
- Ismail, M. H., Ghazi, T. I. M., Hamzah, M. H., Manaf, L. A., Tahir, R. M., Nasir, A. M., & Omar, A. E. (2020a). Impact of movement control order (Mco) due to coronavirus disease (covid-19) on food waste generation: A case study in klang valley, malaysia. *Sustainability*, 12(21), 1-17. <https://doi.org/10.3390/su12218848>
- Ismail, M. H., Khan, K. A., Ngadisih, N., Irie, M., Ong, S. P., Hii, C. L., & Law, C. L. (2020b). Two-step falling rate in the drying kinetics of rice noodle subjected to pre-treatment and temperature. *Journal of Food Processing and Preservation*, 44(11), 1-11. <https://doi.org/10.1111/jfpp.14849>
- Ismail, M. H., Lik, H. C., Routray, W., & Woo, M. W. (2021). Determining the effect of pre-treatment in rice noodle quality subjected to dehydration through hierarchical scoring. *Processes*, 9(8), 1-12. <https://doi.org/10.3390/pr9081309>
- Jahanbakhshi, A., & Kheiralipour, K. (2019). Influence of vermicompost and sheep manure on mechanical properties of tomato fruit. *Food Science and Nutrition*, 7(4), 1172-1178. <https://doi.org/10.1002/fsn3.877>
- Jiménez, E. I., & García, V. P. (1992). Relationships between organic carbon and total organic matter in municipal solid wastes and city refuse composts. *Bioresource Technology*, 41(3), 265-272. [https://doi.org/10.1016/0960-8524\(92\)90012-M](https://doi.org/10.1016/0960-8524(92)90012-M)
- Kannah, R. Y., Merrylin, J., Devi, T. P., Kavitha, S., & Sivashanmugam, P. (2020). Bioresource technology reports food waste valorization: Biofuels and value added product recovery. *Bioresource Technology Reports*, 11, Article 100524. <https://doi.org/10.1016/j.biteb.2020.100524>
- Khan, A., & Ishaq, F. (2011). Chemical nutrient analysis of different composts (vermicompost and pitcompost) and their effect on the growth of a vegetative crop *Pisum sativum*. *Asian Journal of Plant Science and Research*, 1(1), 116-130.
- Khodifad, B. C., & Dhamsaniya, N. K. (2020). Drying of food materials by microwave energy - A review. *International Journal of Current Microbiology and Applied Sciences*, 9(5), 1950-1973. <https://doi.org/10.20546/ijemas.2020.905.223>
- Kibler, K. M., Reinhart, D., Hawkins, C., Motlagh, A. M., & Wright, J. (2018). Food waste and the food-energy-water nexus: A review of food waste management alternatives. *Waste Management*, 74, 52-62. <https://doi.org/10.1016/j.wasman.2018.01.014>
- Liaquat, R., Jamal, A., Tauseef, I., Qureshi, Z., Farooq, U., Imran, M., & Ali, M. I. (2017). Characterizing bacterial consortia from an anaerobic digester treating organic waste for biogas production. *Polish Journal of Environmental Studies*, 26(2), 709-716. <https://doi.org/10.15244/pjoes/59332>
- Lim, W. J., Chin, N. L., Yusof, A. Y., Yahya, A., & Tee, T. P. (2016). Food waste handling in Malaysia and comparison with other Asian countries. *International Food Research Journal*, 23, S1-S6.

- Liu, H., Jiaqiang, E., Ma, X., & Xie, C. (2016). Influence of microwave drying on the combustion characteristics of food waste. *Drying Technology*, 34(12), 1397-1405. <https://doi.org/10.1080/07373937.2015.1118121>
- Loizidou, A. S. D. M. M. (2015). Dehydration of domestic food waste at source as an alternative approach for food waste management. *Waste and Biomass Valorization*, 6(2), 167-176.
- Mahmood, A., Iguchi, R., & Kataoka, R. (2019). Multifunctional food waste fertilizer having the capability of Fusarium - Growth inhibition and phosphate solubility: A new horizon of food waste recycle using microorganisms. *Waste Management*, 94, 77-84. <https://doi.org/10.1016/j.wasman.2019.05.046>
- Makan, A., Assobhei, O., & Mountadar, M. (2013). Effect of initial moisture content on the in-vessel composting under air pressure of organic fraction of municipal solid waste in Morocco. *Journal of Environmental Health Science and Engineering*, 10(1), 1-9. <https://doi.org/10.1186/1735-2746-10-3>
- Maragkaki, A., Galliou, F., Markakis, N., Sabathianakis, G., Tsompanidis, C., Lolos, G., Mavrogiannis, G., Koukakis, G., Lasaridi, K., & Manios, T. (2016). Initial investigation of the solar drying method for the drying of olive oil by-products. *Waste and Biomass Valorization*, 7(4), 819-830. <https://doi.org/10.1007/s12649-016-9505-5>
- McAdams, B., von Massow, M., Gallant, M., & Hayhoe, M. A. (2019). A cross industry evaluation of food waste in restaurants. *Journal of Foodservice Business Research*, 22(5), 449-466. <https://doi.org/10.1080/015378020.2019.1637220>
- Melikoglu, M., Lin, C. S. K., & Webb, C. (2013). Analysing global food waste problem: Pinpointing the facts and estimating the energy content. *Central European Journal of Engineering*, 3(2), 157-164. <https://doi.org/10.2478/s13531-012-0058-5>
- Mirabella, N., Castellani, V., & Sala, S. (2014). Current options for the valorization of food manufacturing waste: A review. *Journal of Cleaner Production*, 65, 28-41. <https://doi.org/10.1016/j.jclepro.2013.10.051>
- Nowak, D., & Jakubczyk, E. (2020). The freeze-drying of foods - The characteristic of the process course and the effect of its parameters on the physical properties of food materials. *Foods*, 9(10), Article 1488. <https://doi.org/10.3390/foods9101488>
- O'Connor, J., Hoang, S. A., Bradney, L., Rinklebe, J., Kirkham, M. B., & Bolan, N. S. (2022). Value of dehydrated food waste fertiliser products in increasing soil health and crop productivity. *Environmental Research*, 204, Article 111927. <https://doi.org/10.1016/j.envres.2021.111927>
- O'Connor, J., Hoang, S. A., Bradney, L., Dutta, S., Xiong, X., Tsang, D. C. W., Ramadass, K., Vinu, A., Kirkham, M. B., & Bolan, N. S. (2021). A review on the valorisation of food waste as a nutrient source and soil amendment. *Environmental Pollution*, 272, Article 115985. <https://doi.org/10.1016/j.envpol.2020.115985>
- Obi, O. F., Ezeoha, S. L., & Egwu, C. O. (2016). Evaluation of air oven moisture content determination procedures for pearl millet (*Pennisetum glaucum* L.). *International Journal of Food Properties*, 19(2), 454-466. <https://doi.org/10.1080/10942912.2015.1038566>
- Palaniveloo, K., Amran, M. A., Norhashim, N. A., Mohamad-Fauzi, N., Peng-Hui, F., Hui-Wen, L., Kai-Lin, Y., Jiale, L., Chisn-Yee, M. G., Jing-Yi, L., Gunasekaran, B., Razak, S. A. (2020). Food waste composting and microbial community structure profiling. *Processes*, 8(6), 1-30. <https://doi.org/10.3390/pr8060723>

- Papanikola, K., Papadopoulou, K., Tsiliyannis, C., Fotinopoulou, I., Katsiampoulas, A., Chalarakis, E., Georgiopoulou, M., Rontogianni, V., Michalopoulos, I., Mathioudakis, D., Lytras, G. M., & Lyberatos, G. (2019). Food residue biomass product as an alternative fuel for the cement industry. *Environmental Science and Pollution Research*, 26(35), 35555-35564. <https://doi.org/10.1007/s11356-019-05318-4>
- Papargyropoulou, E., Lozano, R., K. Steinberger, J., Wright, N., & Ujang, Z. (2014). The food waste hierarchy as a framework for the management of food surplus and food waste. *Journal of Cleaner Production*, 76, 106-115. <https://doi.org/10.1016/j.jclepro.2014.04.020>
- Pergola, M., Persiani, A., Pastore, V., Palese, A. M., Adamo, C. D., De Falco, E., & Celano, G. (2020). Sustainability assessment of the green compost production chain from agricultural waste: A case study in Southern Italy. *Agronomy*, 10(2), Article 230. <https://doi.org/10.3390/agronomy10020230>
- Rahman, M. S., & Perera, C. O. (2007). Drying and food preservation. In *Handbook of food preservation* (pp. 421-450). CRC Press. <https://doi.org/10.1201/9781420017373-26>
- Ravindran, R., & Jaiswal, A. K. (2016). Exploitation of food industry waste for high-value products. *Trends in Biotechnology*, 34(1), 58-69. <https://doi.org/10.1016/j.tibtech.2015.10.008>
- Sakaguchi, L., Pak, N., & Potts, M. D. (2018). Tackling the issue of food waste in restaurants: Options for measurement method, reduction and behavioral change. *Journal of Cleaner Production*, 180, 430-436. <https://doi.org/10.1016/j.jclepro.2017.12.136>
- Sotiropoulos, A., Malamis, D., & Loizidou, M. (2015). Dehydration of domestic food waste at source as an alternative approach for food waste management. *Waste and Biomass Valorization*, 6(2), 167-176. <https://doi.org/10.1007/s12649-014-9343-2>
- Salemdeeb, R., zu Ermgassen, E. K. H. J., Kim, M. H., Balmford, A., & Al-Tabbaa, A. (2017). Environmental and health impacts of using food waste as animal feed: a comparative analysis of food waste management options. *Journal of Cleaner Production*, 140, 871-880. <https://doi.org/10.1016/j.jclepro.2016.05.049>
- Schroeder, J. T., Labuzetta, A. L., & Trabold, T. A. (2020). Assessment of dehydration as a commercial-scale food waste valorization strategy. *Sustainability*, 12(15), 1-13. <https://doi.org/10.3390/su12155959>
- Silvennoinen, K., Heikkilä, L., Katajajuuri, J., & Reinikainen, A. (2015). Food waste volume and origin : Case studies in the Finnish food service sector. *Waste Management*, 46, 140-145. <https://doi.org/10.1016/j.wasman.2015.09.010>
- Salim, N. S., Singh, A., & Raghavan, V. (2017). Potential utilization of fruit and vegetable wastes for food through drying or extraction techniques. *Novel Techniques in Nutrition and Food Science*, 1(2), 1-13. <https://doi.org/10.31031/NTNF.2017.01.000506>
- Sotiropoulos, A., Malamis, D., Michailidis, P., Krokida, M., & Loizidou, M. (2016). Research on the drying kinetics of household food waste for developing and optimizing domestic waste drying technique. *Environmental Technology*, 37(8), 929-939. <https://doi.org/10.1080/21622515.2015.1092588>
- Sozzi, A., Zambon, M., Mazza, G., & Salvatori, D. (2021). Fluidized bed drying of blackberry wastes: Drying kinetics, particle characterization and nutritional value of the obtained granular solids. *Powder Technology*, 385, 37-49. <https://doi.org/10.1016/j.powtec.2021.02.058>

- Spiker, M. L., Hiza, H. A. B., Siddiqi, S. M., & Neff, R. A. (2017). Wasted food, wasted nutrients: nutrient loss from wasted food in the united states and comparison to gaps in dietary intake. *Journal of the Academy of Nutrition and Dietetics*, 117(7), 1031-1040. <https://doi.org/10.1016/j.jand.2017.03.015>
- Sulaiman, N. F. A. R., & Ahmad, A. (2018). Save the food for a better future: A discussion on food wastage in Malaysia. *International Journal of Law, Government and Communication*, 3(10), 12-21.
- Taylor, P., Ong, S. P., Law, C. L., Ong, S. P., & Law, C. L. (2011). Drying kinetics and antioxidant phytochemicals retention of salak fruit under different drying and pretreatment conditions. *Drying Technology*, 29(4), 429-441. <https://doi.org/10.1080/07373937.2010.503332>
- Teresita, M., Villota, O., David, C., Casallas, P., Camilo, D., & Ayala, B. (2021). Composting of solid waste from the coffee milling process using trench composting and a bioreactor with the help of efficient microorganisms in the Libre University, Socorro headquarters. *Ingenieria Solidaria*, 17(1), 1-28. <https://doi.org/10.16925/2357-6014.2021.01.05>
- Thani, N. M., Kamal, S. M. M., Sulaiman, A., Taip, F. S., Omar, R., & Izhar, S. (2020). Sugar recovery from food waste via sub-critical water treatment. *Food Reviews International*, 36(3), 241-257. <https://doi.org/10.1080/87559129.2019.1636815>
- Tiquia, S. M. (2005). Microbiological parameters as indicators of compost maturity. *Journal of Applied Microbiology*, 99(4), 816-828. <https://doi.org/10.1111/j.1365-2672.2005.02673.x>
- Tony, M. A., & Tayeb, A. M. (2011, November 14-16). The use of solar energy in a low-cost drying system for solid waste management: Concept, design and performance analysis. In *Eurasia Waste Management Symposium* (pp. 14-16). Istanbul, Turkey.
- Toundou, O., Pallier, V., Feuillade-Cathalifaud, G., & Tozo, K. (2021). Impact of agronomic and organic characteristics of waste composts from togo on *Zea mays* L. nutrients contents under water stress. *Journal of Environmental Management*, 285, Article 112158. <https://doi.org/10.1016/j.jenvman.2021.112158>
- Tun, M. M., & Juchelková, D. (2019). Drying methods for municipal solid waste quality improvement in the developed and developing countries: A review. *Environmental Engineering Research*, 24(4), 529-542. <https://doi.org/10.4491/eer.2018.327>
- Twain, M. (2020). Soil pH and soil acidity. In *Roughing It* (pp. 107-115). University of California Press. <https://doi.org/10.1525/9780520948068-019>.
- Vakalis, S., Moustakas, K., Semitekolos, D., Novakovic, J., Malamis, D., Zoumpoulakis, L., & Loizidou, M. (2018). Introduction to the concept of particleboard production from mixtures of sawdust and dried food waste. *Waste and Biomass Valorization*, 9(12), 2373-2379. <https://doi.org/10.1007/s12649-018-0214-0>
- Venkatachalam, S. K., Vellingri, A. T., & Selvaraj, V. (2020). Low-temperature drying characteristics of mint leaves in a continuous-dehumidified air drying system. *Journal of Food Process Engineering*, 43(4), 1-15. <https://doi.org/10.1111/jfpe.13384>
- Wang, H., Zhang, M., & Adhikari, B. (2014). Food and bioproducts processing drying of shiitake mushroom by combining freeze-drying and mid-infrared radiation. *Food and Bioproducts Processing*, 94, 507-517. <https://doi.org/10.1016/j.fbp.2014.07.008>
- Wang, J., Xiong, Y. S., & Yu, Y. (2004). Microwave drying characteristics of potato and the effect of different microwave powers on the dried quality of potato. *European Food Research and Technology*, 219(5), 500-506. <https://doi.org/10.1007/s00217-004-0979-1>
- Wang, L., Li, Y., Prasher, S. O., Yan, B., Ou, Y., Cui, H., & Cui, Y. (2019). Organic matter, a critical factor to immobilize phosphorus, copper, and zinc during composting under various initial C/N ratios. *Bioresource Technology*, 289, Article 121745. <https://doi.org/10.1016/j.biortech.2019.121745>

- Wheeler Jr, R. R., Hadley, N. M., Dahl, R. W., Williams, T. W., Zavala Jr, D. B., Akse, J. R., & Fisher, J. W. (2007). Microwave enhanced freeze drying of solid waste. *Journal of Aerospace*, 116(1), 510-537. <https://doi.org/10.4271/2007-01-3266>
- Zaha, C., Dumitrescu, L., & Manciulea, I. (2013). Correlations between composting conditions and characteristics of compost as biofertilizer. *Bulletin of the Transilvania University of Brasov Engineering Sciences Series I*, 6(1), 51-58.
- Zaki, A. H. (2019, September 29). Waste not, want not - It's time we get serious about food waste. *New Straits Times*. <https://www.nst.com.my/lifestyle/sunday-vibes/2019/09/525506/waste-not-want-not-%E2%80%93-its-time-we-get-serious-about-food-waste>

Perceived Benefits, Perceived Barriers and Self-Efficacy Towards Weight Reduction Among Overweight and Obese Children in Kedah, Malaysia

Wan Mohd Nurussabah Abdul Karim, Hazizi Abu Saad*, Nurzalinda Zalbahar and Nurul Husna Mohd Sukri

Department of Nutrition, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

The objective of this study was to identify the psychological factors (perceived barriers to, perceived benefits of, and perceived self-efficacy toward weight reduction) associated with body mass index (BMI) among overweight (OW) and obese children. A cross-sectional survey was carried out among primary school children in Kedah, Malaysia. A multistage randomised cluster sampling technique was used in this study. The socio-demographic data were collected from the parents, and the researchers took anthropometric measurements of the children. A set of validated instruments were used to assess perceived benefits of weight reduction, perceived barriers to weight reduction, perceived self-efficacy in dietary practice and perceived self-efficacy in exercise among the participants. Data were analysed by using SPSS software for Windows Version 25.0. A total of 398 OW and obese children aged 10 and

11 years old (221 boys and 177 girls) were recruited. Based on the socio-demographic data, the father's educational level was found to be significantly associated with the child's body mass index (BMI) ($\chi^2 = 7.873, p = 0.049$). There were no significant differences in mean scores for perceived benefits, perceived barriers, and perceived self-efficacy in dietary practice and exercise between the OW and obese groups ($p > 0.05$). However, perceived self-efficacy in exercise was significantly associated with BMI

ARTICLE INFO

Article history:

Received: 11 December 2021

Accepted: 7 March 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.34>

E-mail addresses:

wannurussabah@moh.gov.my (Wan Mohd Nurussabah Abdul Karim)

hazizi@upm.edu.my (Hazizi Abu Saad)

nurzalinda@upm.edu.my (Nurzalinda Zalbahar)

n_husna@upm.edu.my (Nurul Husna Mohd Sukri)

*Corresponding author

among OW and obese children ($p < 0.05$). Health perceptions and beliefs variables are important and should be included in any future weight management programme among OW and obese children.

Keywords: Body mass index, body weight status, obesity, overweight, perceived barriers, perceived benefits, perceived self-efficacy

INTRODUCTION

Obesity is a multifactorial syndrome with genetic, sociocultural, environmental and psychological factors contributing to its aetiology. Currently, obesity-combating initiatives, especially through prevention programmes, are widely practised. Different subclasses of obesity are crucial for tailoring appropriate and efficient care (Buscemi et al., 2013). However, positive change must come from the person themselves and support within all parts of society, including governments, schools, businesses, non-profit organisations, neighbourhoods, communities and families, to make true progress. Hence, a detailed assessment of psychological and emotional risks and current eating and exercise habits are also important to help overweight (OW) or obese people maintain a healthier weight.

Children's obesity is one of the most important public health challenges of the 21st century. The number of OW and obese children and adolescents aged 5 to 19 has increased drastically worldwide, from 4% in 1975 to over 18% in 2016 (WHO, 2019). Findings from the National Health and Morbidity Survey Malaysia (NHMS) showed that the prevalence of OW and obese children and adolescents (10 to 17 years old) in the state of Kedah was 30.9% (IPH, 2017) and in Malaysia has increased from 30.4 to 32.5% (IPH, 2019). Socio-demographic characteristics such as children's locality have been suggested as important determinants of childhood obesity. Thus, from this study, the association between socio-demographic characteristics and childhood obesity can determine Kedah's nature and obesity situation.

Previous studies have shown that psychosocial factors, such as self-belief and self-efficacy, play major roles in the development of obesity (Lazzeretti, 2015; Faghri et al., 2016). Lazzeretti (2015) also added that a psychological asset could be strictly related to unhealthy behaviours, fickle compliance, and poor weight loss programme results. As they can lead to unhealthy lifestyles, they are known to be prominent factors in developing obesity. Thus, there has been increasing interest in understanding the beliefs or psychological functioning of obese children and the effect of psychological influences on the outcome of treatment. Psychological features or behavioural interventions play a key role in treating obesity. They have also been examined to minimise the lack of compliance and increase the effectiveness of dietary treatment. It indicates that higher success rates in treating obesity may correlate with good psychological functioning, and weight loss may be associated with beneficial effects on several care areas (Tseng et al., 2002).

There are several theories and models related to health and behaviour. The Health Belief Model (HBM) is commonly used in health promotion. Champion and Skinner (2008) stated that HBM is a psychological model that describes and predicts the health behaviour used to understand how a disease could be identified and avoided. The pillars of the HBM, namely perceived threat (perceived susceptibility and perceived seriousness or severity) and response effectiveness or modifying variables (perceived benefits, perceived barriers, and self-efficacy), were the basis of the HBM.

Studies on children's beliefs about or perceptions of weight reduction have been conducted in several countries, including South Korea (Park, 2011), Mexico (Rodríguez-Ventura et al., 2014), and seven Middle Eastern countries (Musaiger et al., 2013). Park (2011) reported that, regarding the effect of the health belief model in controlling obesity among elementary school children, health belief was a significant predictive variable affecting the weight-management activity and the desire to control obesity. In this study, only modifying variables were applied due to time constraints regarding the children's school. International studies have shown that childhood obesity can lead to depression in adulthood (Sánchez-Villegas et al., 2010), and obese children are far more likely to become obese adolescents and adults unless they adopt healthy exercise and eating habits. Thus, it is important to promote a better understanding and effective early intervention in younger children (Simmonds et al., 2016).

In Malaysia, studies on psychological or belief factors (barriers, benefits, self-efficacy) among OW and obese children are still lacking. To our knowledge, no similar study has reported the psychological factors for this particular group (OW and obese children) related to primary school children under 12 years old in Malaysia. Therefore, this study aimed to identify the psychological factors (perceived barriers, benefits, self-efficacy towards weight reduction) associated with body mass index (BMI) among OW and obese children.

METHODS

Participants and Recruitment

A cross-sectional study was carried out to investigate the associations of the independent variable (perceived barriers, benefits, and self-efficacy regarding weight reduction) with BMI among OW and obese children in Kedah. The sample size was estimated using the proportions formula created by Lemeshow et al. (1990) and data from the previous study in Kedah (IPH, 2017). The minimum sample size required for this study was 70 participants. In consideration of non-responses and missing data, the minimum sample size was increased by 30%, leading to a total of 91 participants. Participation in this study was selected based on the following eligibility criteria: boys and girls who were OW, $> + 1$ standard deviation (SD) - $\leq + 2SD$, or obese $> 2SD$ (WHO, 2020). Only participations who could provide

written consent and had parents who could read and understand the Malay language were included in this study. Those with physical or mental disabilities were excluded.

A multistage randomised cluster sampling technique was used in this study. A list of primary schools was obtained from the Jabatan Pendidikan Negeri Kedah to achieve the needed sampling size (2019). Participants were selected from rural and urban areas to include representative participants from the state of Kedah. All eleven districts in Kedah were involved in this study. One school was randomly selected from each district from March 2019 to September 2019 by visiting the selected schools. The respective school teachers provided the list of OW and obese children from grades 4 and 5, that is, ages 10 and 11, after being extracted from the National Physical Fitness Standard (SEGAK) test data. Overweight and obese school children aged 10–11 years from the selected schools were chosen based on the assumption that older children (aged 10 and above) were more likely than younger children to express their perceptions correctly, and six primary school children were excluded due to examination at a certification level before entering secondary school.

The written informed consent form was given to the parents of the participants (461), and 415 parents agreed to participate in this study. Of the 415 participants, 398 provided complete data (221 boys and 177 girls). The response rate of the study was 84.3%. Data were collected in three stages: in the first stage, a consent form and socio-demographic data were collected from the parents. The second stage covered anthropometric measurements of the children, while the third stage involved answering HBM questionnaires by children. The perceived benefits, perceived barriers, perceived self-efficacy in dietary practice, and perceived self-efficacy in exercise were among the constructs explored in this study. Respondents must complete the questionnaire (guided group interview) to obtain research information. The researcher explained the question and guided them question by question.

A second screening (using anthropometry measurements) was performed to ensure that only OW or obese children were involved in this study. Body weight was measured in kilograms using a Seca 803 digital scale, with a precision of 0.01 kg. Height was measured centimetres to the nearest 0.1 cm using a Seca 213 portable free-standing height measurement. Measurements were taken twice to obtain the averaged results to minimise measurement bias. This study was approved by the Research and Ethics Committee, Ministry of Health (NMRR-19-577-45778(IIR)). Written informed consent was obtained from all the participants and their parents.

Survey Instruments

The questionnaire included questions regarding demographic characteristics, such as family income, parents' education, and occupation, and questions related to the HBM (48 questions), which comprised four sections and were utilised for data collection. Parents were asked to answer the socio-demographic survey separately and were allowed to bring

it home with them. The instruments related to the health belief model (HBM) were adapted from Park (2011), who used a scale for HBM constructs in their study. The Cronbach's α of the instruments was ($\alpha = 0.859$) for the perceived benefits of weight reduction, ($\alpha = 0.805$) for perceived barriers to weight reduction, ($\alpha = 0.843$) for perceived self-efficacy in dietary practice, and ($\alpha = 0.831$) perceived self-efficacy in exercise. The response categories used a 4-point Likert scale (1—completely disagree, 2—disagree, 3—agree, 4—completely agree). For each item listed, a high score was associated with a more positive weight reduction perception. A total score was calculated by summing the marks obtained for perceived benefits, perceived barriers, and perceived self-efficacy. The total score was calculated for the perceived benefits of weight reduction by summing the marks obtained from items 1 to 8. The possible score for this section ranged from a minimum of 8 to 32. The total score was classified into three categories: the low score category ranged from 8 to 16, the moderate category ranged from 17 to 24, and the high category ranged from 25 to 32. Perceived barriers to weight reduction were categorised as low (score 12 to 24), moderate (25 to 36), or high (37 to 48). Perceived self-efficacy in dietary practice was categorised as low (16 to 32), moderate (33 to 48), or high (49 to 64). Perceived self-efficacy in exercise was also categorised as low (9 to 18), moderate (19 to 27), or high (28 to 36).

A guided group interview was used to obtain the information. The researcher explained the questions to every child and guided them question by question. The questionnaire was first translated into Bahasa Malaysia using the forward translation method (English to Bahasa Malaysia). It then was translated again using the backward translation method (Bahasa Malaysia to English) to ensure that the questionnaire's language, meaning, and content were correct and clear, according to two qualified English school teachers. The final version of the instrument was presented to the expert panel of this research group for approval before the data collection began.

Statistical Analysis

Data were analysed using the Statistical Package for Social Sciences (SPSS) software for Windows version 25.0 (SPSS Inc., 2010, Chicago, Illinois). Descriptive statistics were used to describe variables, such as BMI, perceived benefits, perceived barriers, and perceived self-efficacy. The chi-square test was used to determine associations between the variables. An independent sample t-test was used to compare the difference between the means of the OW and obese groups. Multiple logistic regression analysis was used to determine the relationship between HBM variables and children's BMI. The statistical significance was assigned for all statistical analyses at $p < 0.05$.

RESULTS

Respondent's Characteristics

A total of 398 children from 11 primary schools in Kedah were involved in this study; boys made up 55.5% of the sample, while girls comprised 44.5% (Table 1). The study sample was multi-ethnic in composition, while Malay children comprised the largest population, 97.7%. All respondents were between 10 and 11 years old, and most (61.6%) were from urban areas. The mean BMI of the children (25.77 ± 4.27 kg/m²) was higher among the boys than the girls (25.50 ± 3.73 kg/m²). However, no significant difference was found between BMI and sex.

Table 1

Characteristics of study respondents

Characteristics	Boys n (%)	Girls n (%)	Total n (%)	<i>p</i> -value
Sex	221 (55.5)	177 (44.5)	398 (100.0)	
Ethnic				
Malay	215 (97.3)	174 (43.7)	389 (97.7)	
Indian	1 (0.5)	0 (0.0)	1 (0.3)	
Others	5 (2.3)	3 (0.8)	8 (2.0)	
Standard (Age)				
4 (10 y/o)	102 (46.2)	71 (40.1)	173 (43.5)	
5 (11 y/o)	119 (53.8)	106 (59.9)	225 (56.5)	
School location				
Urban	147 (66.5)	98 (55.4)	245 (61.6)	
Rural	74 (33.5)	79 (44.6)	153 (38.4)	
Body weight status				
OW	40 (18.1)	56 (31.6)	96 (24.1)	
Obese	181 (81.9)	121 (68.4)	302 (75.9)	
BMI (kg/m ²)	25.77 ± 4.27	25.50 ± 3.73	25.65 ± 4.04	0.513

Independent t-test. **p* < 0.05

About 70% of their parents had attained a minimum secondary school education, and the father's educational level was found to be significantly associated with the child's BMI ($\chi^2 = 7.873$, *p* = 0.049); however, the mother's educational level was not associated with the child's BMI (Table 2). Most fathers for both categories (OW and obese children) were in elementary jobs, and most mothers were not in the labour force (99.4% homemakers). About 60% of the respondents were classified in the B40 monthly household income, with

a mean of RM1746 ± 861. The children were mostly from low- and bottom-income families based on Malaysia's B40, M40 and T20 household income distribution. In addition, 11.3% (B40 category) came from poor-income families (monthly income of less than RM950). The parents' occupational level and monthly household income were not significantly associated with the child's BMI.

Table 2

Difference in proportion of parental characteristics based on children BMI

Parental Characteristics	Father		Mother	
	OW n (%)	Obese n (%)	OW n (%)	Obese n (%)
Educational level ^a				
Not formal	1 (1.0)	0 (0)	3 (3.1)	1 (0.3)
Primary education	6 (6.3)	6 (2.0)	3 (3.1)	6 (2.0)
Secondary education	66 (68.8)	225 (74.5)	67 (69.8)	215 (71.2)
Tertiary education	23 (23.9)	71 (23.5)	23 (24.0)	80 (26.5)
Total	96 (100)	302 (100)	96 (100)	302 (100)
	$\chi^2 = 7.873, p = 0.049^*$		$\chi^2 = 6.276, p = 0.099$	
Occupational level ^b (MASCO, 2020)				
Managers	1 (1.0)	5 (1.7)	1 (1.0)	1 (0.3)
Professionals	13 (13.5)	35 (11.6)	13 (13.5)	45 (14.9)
Technicians and associate professionals	7 (7.3)	26 (8.6)	5 (5.2)	16 (5.3)
Clerical support workers	6 (6.3)	16 (5.3)	12 (12.5)	34 (11.3)
Service and sales workers	2 (2.1)	8 (2.6)	2 (2.1)	4 (1.3)
Skilled agricultural, forestry and fishery workers	10 (10.4)	34 (11.3)	3 (3.1)	8 (2.6)
Craft and related trade workers	3 (3.1)	21 (7.0)	1 (1.0)	11 (3.6)
Plant and machine operators/ assemblers	12 (12.5)	28 (9.3)	2 (2.1)	10 (3.3)
Elementary occupations	31 (32.3)	77 (25.5)	11 (11.5)	45 (14.9)
Armed forces occupations	7 (7.3)	25 (8.3)	0 (0)	2 (0.7)
Outside labour force (housewife/ pensioner)	4 (4.2)	27 (8.9)	46 (47.9)	126 (41.7)
Total	96 (100)	302 (100)	96 (100)	302 (100)
	$\chi^2 = 6.944, p = 0.804$		$\chi^2 = 5.548, p = 0.902$	

Table 2 (Continue)

Parental Characteristics	Father		Mother	
	OW n (%)	Obese n (%)	OW n (%)	Obese n (%)
Monthly Household Income ^c (RM)	Father & Mother			
	OW	Obese	Mean ± SD	
B40 (< RM3860)	61 (63.5)	196 (64.9)	RM1746 ± 861	
M40 (RM3860 – RM8,319)	26 (27.1)	85 (28.1)	RM5660 ± 1296	
T20 (> RM8319)	9 (9.4)	21 (7.0)	RM10820±2276	
Total	96 (100)	302 (100)	RM3521±2944	
	$\chi^2 = 0.617, p = 0.735$			

^a Educational level category based on the Department of Statistics Malaysia (2020)

^b Occupation classification-based Malaysia Standard Classification of Occupations (MASCO, 2020)

^c Monthly income categories referred to the Department of Statistic Malaysia (2017) distribution of households by income class in Malaysia. 1USD=RM4.2

Chi-square test - *p* indicates the level of significance in the difference in proportion (*p* < 0.05*)

Perceived Benefits of Weight Reduction

Children’s perceived benefits of weight reduction are presented in Table 3. The mean score of the total perceived benefits item for both groups was 3.18 ± 0.78 . Item 11, regarding the perception of “exercise better with friends”, had the highest mean (3.52 ± 0.61) and Item 6, regarding the perception of “getting friends”, received the lowest mean (2.99 ± 0.82). For all perceived benefits items, the perception of “attractiveness towards others” (Item 8) showed a significant difference (*p* = 0.046), whereby children in the obese group had a higher mean (3.13 ± 0.83) compared to the OW group (2.94 ± 0.84).

Table 3

Mean ± SD of perceived benefits of weight reduction

Perceived benefits	BMI status			<i>p</i> -value
	OW	Obese	All	
1. Health will be improved	3.08 ± 0.84	3.13 ± 0.82	3.12 ± 0.82	0.611
2. Chronic diseases will be prevented	3.27 ± 0.75	3.25 ± 0.79	3.26 ± 0.78	0.862
3. It is easy to act in daily life	3.35 ± 0.74	3.33 ± 0.71	3.34 ± 0.72	0.785
4. Confidence will be increased in everything	3.06 ± 0.78	3.15 ± 0.77	3.13 ± 0.77	0.322
5. Appearance will be better	3.32 ± 0.80	3.35 ± 0.65	3.34 ± 0.69	0.759

Table 3 (Continue)

Perceived benefits	BMI status			<i>p</i> -value
	OW	Obese	All	
6. I can get a boyfriend	2.92 ± 0.80	3.05 ± 0.84	2.99 ± 0.82	0.173
7. I will have a wide range of selection for clothes	3.20 ± 0.76	3.26 ± 0.83	3.25 ± 0.81	0.481
8. I will be attractive to others	2.94 ± 0.84	3.13 ± 0.83	3.09 ± 0.84	0.046*
9. People react favourably to me	2.89 ± 0.87	3.01 ± 0.84	2.98 ± 0.85	0.222
10. I will save money on food	3.07 ± 0.79	3.10 ± 0.76	3.09 ± 0.76	0.768
11. I will exercise and play the game better with friends	3.52 ± 0.63	3.52 ± 0.59	3.52 ± 0.61	0.989
TOTAL	3.15 ± 0.78	3.21 ± 0.77	3.18 ± 0.78	0.547

Response categories used a 4-point Likert scale (completely disagree = 1, disagree = 2, agree = 3, completely agree = 4). A score was given to each response from 1 to 4 to indicate higher scores strong feeling of benefits.

Perceived Barriers to Weight Reduction

As seen in Table 4, the mean of the total perceived barriers to weight reduction for all children was 2.51 ± 0.87 (range 1 - 4). For Item 8, “it is difficult to stand when delicious food is in front of me”, children in both groups had the highest mean (2.95 ± 0.85 for OW) and (2.90 ± 0.80 for obese). The lowest mean for the OW group was for “family often dines out: (Item 7), and for the obese group was for “family often eats food delivered” (Item 6). However, there was no significant difference ($p > 0.05$) between the OW and obese groups.

Table 4

Mean ± SD of perceived barriers to weight reduction

Perceived barriers	BMI status			<i>p</i> -value
	OW	Obese	All	
1. I don't know the right diet therapy for weight control.	2.72 ± 0.78	2.62 ± 0.85	2.64 ± 0.83	0.293
2. I don't know desirable snacks for weight control.	2.65 ± 0.83	2.51 ± 0.90	2.55 ± 0.89	0.203
3. I don't understand processed food's food and nutrition labels.	2.41 ± 0.89	2.57 ± 0.91	2.53 ± 0.91	0.132
4. I cannot make simple good snacks for weight control.	2.66 ± 0.74	2.55 ± 0.85	2.57 ± 0.82	0.254
5. I cannot refuse to eat some food when family and friends ask me to try it.	2.72 ± 0.90	2.68 ± 0.89	2.69 ± 0.89	0.725

Table 4 (Continue)

Perceived barriers	BMI status			
	OW	Obese	All	<i>p</i> -value
6. My family often eats food delivered.	2.16 ± 0.89	2.11 ± 0.87	2.12 ± 0.87	0.623
7. My family often dines out	2.10 ± 0.78	2.15 ± 0.85	2.14 ± 0.83	0.622
8. It is difficult to stand when delicious food is in front of me.	2.95 ± 0.85	2.90 ± 0.80	2.91 ± 0.81	0.596
9. I don't have time for dieting or exercise.	2.17 ± 0.93	2.30 ± 0.90	2.27 ± 0.90	0.204
10. I don't know the proper way to exercise for weight control.	2.45 ± 0.94	2.48 ± 0.91	2.47 ± 0.92	0.788
11. My family and friends seldom exercise.	2.51 ± 0.90	2.51 ± 0.86	2.51 ± 0.86	0.970
12. It is difficult to practice even though I make a plan.	2.63 ± 0.93	2.57 ± 0.89	2.59 ± 0.89	0.621
TOTAL	2.51 ± 0.86	2.50 ± 0.87	2.51 ± 0.70	0.503

Response categories used a 4-point Likert scale (completely disagree = 1, disagree = 2, agree = 3, completely agree = 4). A score was given to each response from 1 to 4 to indicate higher scores and fewer feelings of barriers. T-test (CI 95%), **p* < 0.05

Perceived Self-Efficacy in Dietary Practice

As seen in Table 5, the total mean of perceived self-efficacy in dietary practice for weight reduction for all children was 2.77 ± 0.93 (range 1–4). The highest mean was Item 3 regarding eating fresh food instead of processed food: (3.11 ± 0.89), and the lowest mean Item 13 regarding their: “refusal to eat even though offered by family members” (2.40 ± 0.88). There were no significant differences between the groups.

Table 5

Perceived self-efficacy in dietary practice

Perceived self-efficacy in dietary practice	BMI status			
	OW	Obese	All	<i>p</i> -value
1. I can eat three meals regularly.	2.81 ± 0.84	2.77 ± 0.90	2.78 ± 0.88	0.669
2. I can eat meals in moderate amounts.	2.80 ± 0.80	2.93 ± 0.76	2.90 ± 0.76	0.152
3. I can eat fresh food rather than processed one.	3.17 ± 0.84	3.05 ± 0.93	3.11 ± 0.89	0.287
4. I can refrain from eating sweets like candy and cookies.	2.75 ± 0.92	2.84 ± 0.84	2.82 ± 0.86	0.385
5. I can refrain from eating fatty food like fried food and fork belly.	2.58 ± 0.82	2.46 ± 0.80	2.49 ± 0.80	0.181
6. I can refrain from drinking carbonated drinks like cola and cider.	3.06 ± 0.99	3.06 ± 0.91	3.06 ± 0.93	0.997

Table 5 (Continue)

Perceived self-efficacy in dietary practice	BMI status			<i>p</i> -value
	OW	Obese	All	
7. I can eat various foods to avoid unbalance in my diet.	2.63 ± 1.01	2.70 ± 0.90	2.68 ± 0.93	0.478
8. I can refrain from eating just before going to bed.	2.80 ± 0.90	2.77 ± 0.91	2.78 ± 0.90	0.797
9. I can eat slowly, even when hungry	2.89 ± 0.96	2.85 ± 0.94	2.86 ± 0.94	0.756
10. I can stop to eat before filling my stomach, even if the food is delicious.	2.75 ± 0.62	2.60 ± 0.95	2.64 ± 0.95	0.176
11. I can get up early to eat breakfast.	2.77 ± 0.96	2.85 ± 0.97	2.83 ± 0.97	0.499
12. I can refrain from watching TV or reading a book when I eat.	2.50 ± 0.97	2.45 ± 1.01	2.46 ± 0.99	0.692
13. I can refuse to eat when my family or friends offer me food.	2.43 ± 0.88	2.37 ± 0.88	2.40 ± 0.88	0.607
14. I can refrain from eating when I am bored.	2.65 ± 0.90	2.63 ± 0.93	2.63 ± 0.92	0.853
15. I can refrain from eating when I am angry.	2.83 ± 0.98	2.85 ± 0.95	2.84 ± 0.96	0.898
16. I can refrain from eating when I am depressed.	2.82 ± 0.87	2.96 ± 1.96	2.93 ± 1.76	0.719
TOTAL	2.77 ± 0.89	2.76 ± 0.97	2.77 ± 0.93	0.572

Response categories used a 4-point Likert scale (completely disagree = 1, disagree = 2, agree = 3, completely agree = 4). A score was given to each response from 1 to 4 to indicate higher scores and less feeling of self-efficacy in dietary practice. T-test (CI 95%), **p* < 0.05

Perceived Self-Efficacy in Exercise

Perceived self-efficacy in exercise is presented in Table 6. The mean perceived self-efficacy in exercise for all children was 2.55 ± 0.91 (range 1–4). Item 1, “I can do light exercise every free moment,” had the highest mean (2.99 ± 0.82), and Item 5, “I can exercise in cold weather,” had the lowest mean. There were no significant differences between the two groups. Only Item 8, “I can exercise instead of watching TV in leisure time”, and Item 9, “I can go up and down for five floors,” showed significant differences (*p* = 0.006 and *p* = 0.05, respectively), with children in the OW group having a higher mean score than the obese group.

Table 6
Perceived self-efficacy in exercise

Perceived self-efficacy in exercise	BMI status			<i>p</i> -value
	OW	Obese	All	
1. I can do light exercise every free moment.	3.03 ± 0.85	2.96 ± 0.78	2.99 ± 0.82	0.448
2. I can exercise until getting short of breath.	2.98 ± 0.92	2.81 ± 0.94	2.85 ± 0.93	0.125
3. I can walk for a distance as far as 15 minutes walk.	2.95 ± 0.76	2.81 ± 0.87	2.84 ± 0.85	0.159
4. I can exercise for 30 minutes three times a week.	2.56 ± 1.05	2.36 ± 0.98	2.41 ± 0.99	0.085
5. I can exercise in cold weather.	1.86 ± 0.89	1.93 ± 0.85	1.91 ± 0.86	0.534
6. I can exercise in hot weather.	2.06 ± 0.97	2.04 ± 0.91	2.05 ± 0.92	0.833
7. I can exercise with friends after school.	2.70 ± 0.93	2.63 ± 0.86	2.65 ± 0.88	0.524
8. I can exercise instead of watching TV in my leisure time.	2.74 ± 1.03	2.43 ± 0.92	2.51 ± 0.95	0.006*
9. I can go up and down stairs to the fifth floor.	2.63 ± 0.94	2.41 ± 0.93	2.46 ± 0.94	0.050*
TOTAL	2.61 ± 0.93	2.49 ± 0.89	2.55 ± 0.91	0.307

Response categories used a 4-point Likert scale (completely disagree = 1, disagree = 2, agree = 3, completely agree = 4). A score was given to each response from 1 to 4 to indicate higher scores and less feeling of self-efficacy in exercise. T-test (CI 95%), **p* < 0.05

Table 7 shows the unadjusted and adjusted multiple logistic regression analysis of the psychological factors influencing obesity among children in Kedah. The odds of obesity were 2.43 times higher for the high score group of perceived self-efficacy in exercise compared to the low and moderate groups. Factors such as perceived benefits, perceived barriers and self-efficacy in dietary practice did not reveal any significant association with children obesity despite their significant relationship in univariate analysis.

Table 7
Psychological factors associated with obesity of OW and obese children in Kedah

Psychological factors score	Multiple logistic regression analysis			
	Crude OR (95% CI)	<i>p</i> -value	Adjusted OR (95% CI)	<i>p</i> -value
Perceived benefits				
Low & moderate	Reference	-	Reference	-
High	1.27 (0.14,11.55)	0.829	1.03 (0.11,9.62)	0.978

Table 7 (Continue)

Psychological factors Score	Multiple logistic regression analysis			
	Crude OR (95% CI)	<i>p</i> -value	Adjusted OR (95% CI)	<i>p</i> -value
Perceived barriers				
Low & moderate	Reference	-	Reference	-
High	1.29 (0.68,2.44)	0.430	1.31 (0.68,2.53)	0.416
Perceived self-efficacy in dietary practice				
Low & moderate	Reference	-	Reference	-
High	1.19 (0.71,2.00)	0.519	0.78 (0.43, 1.44)	0.430
Perceived self-efficacy in exercise				
Low & moderate	Reference	-	Reference	-
High	2.18 (1.27,3.76)	0.005*	2.43 (1.34,4.42)	0.004*

DISCUSSION

The distribution of the respondents by sex was almost equal, with 55.5% boys and 44.5% girls. Most children were Malays (97.7%), and the rest included one Indian and five others (Siamese). This proportion is because 85.3% of primary school children in Kedah are Malays, 9.8% are Chinese, and 4.9% are Indians and other races (Jabatan Pendidikan Negeri Kedah, 2019). To preserve the uniformity of the school environment, the targeted respondents were mostly children from national primary schools at the expense of the Chinese and Indian children who were mostly enrolled in the vernacular schools. Our study reports that urban children are heavier than rural children. This finding was in line with the National Health and Morbidity Survey (IPH, 2019), which found that urban children were more likely to be OW and obese than their rural counterparts. However, these findings were inconsistent with the IPH (2017) report for Kedah, showing that rural children had higher BMIs than urban children. The number of sampling locations may explain this difference. Two districts were involved in the study mentioned above, whereas eleven were involved in our study. This study also found that older children were more OW and obese than younger children, even though there was no significant difference in the mean between the two groups. In agreement with our findings, past studies have found that older children are more likely to be obese than younger children, but that body shape and weight concerns are increased (Mukherjee et al., 2016; da Cunha Feio Costa et al., 2016).

Gender differences were found to not be significant in mean BMI. It is in line with a local study which stated that there was no significant difference in the mean BMI between boys and girls (Ahmad et al., 2018). However, it contrasts with Anuar et al. (2005), who stated a significant difference in BMI, which was found to be higher among girls than boys.

This difference might be due to the study respondents' puberty age, when girls' bodies start to release hormones that cause physical changes, such as breasts and buttocks. Girls tend to have higher BMI because of rapid growth and physical changes associated with early puberty and sexual maturation (Ahmad et al., 2018).

About 70% of the parents in the current study had attained secondary education, and almost 25% had received tertiary education. A significant association was found between children's BMI and the father's educational level, but no association was found with the mother's educational level. It was consistent with a previous study that reported that the mother's education is associated with her children's physical development, though not as strong as the father's (Anuar et al., 2005). However, inconsistent with these results regarding the association between parents' educational level and obesity. Fitzgerald and Spaccarotella (2009) found no association between the level of education and obesity. Meanwhile, another finding stated that women with low education and socioeconomic status reported the cost of diet as a barrier, significantly more so with higher educated women (Sharifi et al., 2013).

For parents' occupational levels, this study found that three-quarters of fathers were in elementary jobs, while almost half of the mothers were homemakers. No association was found between the father's and mother's educational levels and children's BMI. Gnavi et al. (2016) mentioned that the prevalence of children's OW and obesity was inversely related to both the mother's and father's educational and occupational levels, with a higher prevalence observed in children born to parents with very low educational or occupational levels. Regarding household income, the present study shows that the mean household income for B40 (RM1,746) and T20 (RM10,820) is lower than the national finding for B40 (RM2,089) and T20 (RM11,469), but for the M40 category, this study found a higher amount (RM5,660) than the national finding (RM4,608) (Department of Statistics Malaysia, 2017). The majority (64.6%) of the respondents in this study came from B40 household families, with 27.9% from M40 and 7.5% from T20. Regarding the association of body weight status and socio-demographic factors among preschool children, Norimah et al. (2014) found no significant association between BMI and food intake, even though parents' educational level and household income were significantly associated with food intake.

This study found no significant difference between the means scores of the OW and obese groups for all the perceived benefits items except the item "I will be attractive to others" (Table 3). The respondents thought the greatest perceived benefit of weight reduction was that they would "exercise and play games better with friends", followed by "easy to act in daily life". A similar finding previously reported that being healthy and physically attractive is a prominent benefit of normal BMI among high school boys and that it can help them improve their physical appearance and attract girls (Tergerson & King, 2002). A previous finding by Park (2011) reported that the highest mean score of the perceived benefits was the item "health will be improved", followed by the item "easy to

act in daily life". Several studies have revealed that the obese group, rather than the OW group, perceived obesity to be a serious threat to their health and perceived that a threat or perceived benefit of a health condition could help predict changes in an individual's health behaviours (Moore et al., 2010; Okop et al., 2016). The result of this study also revealed that the children also knew that the items "preventing chronic diseases" and "better appearance" were benefits that they could get from reducing their weight, with mean scores of 3.26 ± 0.78 and 3.34 ± 0.69 , respectively.

The findings show no significant difference in the mean score for perceived barriers to weight reduction in children between the two groups (Table 4). The items "it is difficult to stand when delicious food is in front of me" in diet adherence and the "it is difficult to practise even though I make a plan" for exercise had the highest scores among the children. The lowest mean score for barriers to weight reduction in the OW group (2.10 ± 0.78) was "my family often dines out", and for the obese group (2.11 ± 0.87) was "my family often eats food delivered." In contrast, a previous study (Park, 2011) reported that the highest mean score of perceived barriers to weight reduction was the item "my family often eats food delivered", followed by the item "my family often dines out". This difference might be due to the respondents' family lifestyle and situational factors.

To reduce weight or achieve a normal BMI, acquiring support from family members and friends is very important to overcome perceived barriers, especially situational barriers. "Situational barriers" were the most important factor that kept people from sticking with their recommended diet and planned physical activity. Park (2011) explained the influences of family, peers, and mass media on weight control behaviours, which is related to this study. In relation to this, Serour et al. (2007) highlighted that situational barriers could affect dietary adherence among different groups within the population. He added that adhering to the diet was difficult when the everyday routine was interrupted, such as during weekends, when in a group, while travelling, and dining out. Other barriers to weight loss among OW and obese children were intrapersonal, interpersonal, community, institutional and public policy factors (Fitzgerald & Spaccarotella, 2009). Food is given as a reward, such as fast food as treats, parties, or good achievements. It is involuntary parental encouragement in raising emotional eaters, and the children learn to use food to help them cope with various emotions and feelings. Many children who become OW or obese might have fallen into these entrenched feeding patterns.

This study found that plan adherence scored the highest for physical activity or exercise barriers. Similarly, lack of time as a barrier to physical activity was one of the barriers in many previous studies (Andajani-Sutjahjo et al., 2004; Sit et al., 2008). Lack of interest and motivation to exercise for the item "my family and friends seldom exercise" ranked second highest for exercise barriers. It was consistent with the results stating that tiredness in exercising because of body weight was one of the internal barriers reported

by obese participants (Allison et al., 1999). Moreover, children and adolescents who were OW or obese were less likely to engage in physical activity or exercise because barriers, such as fear of strangers while playing outside, bad weather and too much homework, had a substantial effect on primary school children's levels of exercise (Miri et al., 2017). According to the present study, there was no significant difference between the mean score for perceived self-efficacy in the diet in both groups (Table 5), and this result is consistent with a previous study (Park, 2011). The highest average mean score was dedicated to the item "can eat fresh food rather than processed food," followed by the item "can refrain from drinking carbonated drinks like cola and cider." It is contrary to previous findings (Park, 2011) that the highest mean score of the perceived self-efficacy in the diet was the item "can stop to eat before filling stomach even if the food is delicious" and the item "can get up early to eat breakfast". This difference might be due to the individual respondent's family lifestyles and situational factors.

The lowest score or the most difficult to practise in perceived self-efficacy in dietary practice for both groups was "I can refuse to eat when my family members or friends offer food." It is inconsistent with Park (2011), who found that the lowest mean score in perceived self-efficacy in the diet was for "I can get up early to eat breakfast". According to Shin et al. (2011), individuals with greater overall self-efficacy to avoid eating when food is available may lose weight. That means these groups of children are not ready to reduce weight or control their appetite at this particular time. In relation to this, Kim et al. (2011) found that when obesity incidence was high, self-efficacy was low, and that increased self-efficacy was associated with greater weight reduction. The perceived self-efficacy item "I can refuse to eat when my family members or friends offer foods" portrays an individual's belief in their capabilities to exercise control over challenging demands and over their functioning. These include ambitious self-beliefs of the expectations that people set for themselves, what course of action they want to follow, how much time they spend on the given assignments and how long they persevere in the face of obstacles and setbacks. However, the study by Park (2011) found that the item "can refrain from eating when I am angry" was the most perceived self-efficacy in dietary practice.

Furthermore, there was no significant difference found between OW and obese children in perceived self-efficacy in exercise (Table 6) except for the items "I can exercise instead of watching TV in leisure time" and "I can go up and down stairs for five floors," with the OW group scoring higher than the obese group for both items. However, the greatest mean score for perceived self-efficacy in exercise in both groups was "can do light exercise every free moment," followed by "can exercise until getting short of breath." However, Park (2011) reported that the items "can walk for a distance as far as 15 minutes" and "can go up and down stairs for five floors" were the highest rated items that they could practise. The lowest mean score in both groups was for the item "can exercise in cold weather," while Park (2011) reported that the item "can exercise in hot weather" had the lowest mean score.

In Malaysia, the responses on understanding the weather or season might be different due to the “cold weather,” which would refer to rainy days or nights, and it should be noted that there it is not a norm for Malaysians when exercise in that situation, especially in rural areas. In Korea, however, this would refer to the winter season (Lim et al., 2015). For perceived self-efficacy in exercise, the item “I can exercise instead of watching TV in leisure time” shows a significant difference between the two groups, with the OW group having a higher mean score than the obese group. According to research conducted in Greece, the effect of television viewing time on childhood obesity is independent of physical activity levels and can be attributed to decreased overall energy intake while watching television (Manios et al., 2009).

Multiple logistic regression analysis findings showed that the high scoring group for perceived self-efficacy in exercise was associated with obesity. This result was consistent with prior research (Soliman et al., 2018). Other findings suggest that the obese Class I group was more significantly associated with behavioural intentions of weight reduction than the obese Class II and III groups. However, physical activity level remained an independent factor predicting obesity among middle-school students (Park, 2011). This finding is consistent with a Palestinian study which found that perceived barriers were not significant with behavioural intention of weight reduction (Soliman et al., 2018). The differences in findings might be due to other variables affecting children’s behaviour regarding weight reduction, such as friends, family, mass media and food taboos. Previous studies (Park, 2011; Saghafi- Asl et al., 2020; Mostafavi et al., 2014) have mentioned that perceived benefits seemed to be the most important factors of weight reduction, followed by perceived barriers.

Studies on psychological factors, health perceptions or beliefs among OW and obese children have been conducted widely in several countries (Buscemi et al., 2013; Faghri et al., 2016; Park, 2011; Gnavi et al., 2016). However, to our knowledge, there are no published studies on this topic in Malaysia. Therefore, this study focused on two categories of BMI that exceeded normal body weight status (OW and obese) in the relationship between HBM variables of weight reduction and obesity. Another advantage of this study is that it is considered a follow-up programme from the SEGAK test run by schools. Therefore, new and recent information on the perceptions or beliefs of OW and obese children in Malaysia, particularly regarding weight reduction, would be most useful for future intervention programme design.

CONCLUSION

In summary, no significant mean differences were found regarding BMI based on perceived benefits, perceived barriers, and perceived self-efficacy in dietary practice, except for some items regarding perceived self-efficacy in exercise. This study found that perceived

self-efficacy in exercise was significantly associated with obesity among OW and obese children in Kedah.

The study's outcomes showed that psychological factors and perceptions are also important for BMI status and, therefore, may be important to consider in weight management programs among children. Incorrect perceptions and beliefs about weight reduction among overweight and obese children should be corrected and prioritised in intervention programs for these groups of individuals. With a correct mindset, the children's body weight management program may become more effective and successful.

LIMITATIONS

There are several limitations of this study that could affect the study findings. First, no comparison can be made to normal children, as this study only involved OW and obese children. Second, despite guided group interviews, the data obtained solely relied on the ability of the children to grasp the actual meaning of each question in the instrument and their concentration on answering the question. Even though the validity of each instrument has been demonstrated, it is possible that it was over- or under-reported. Another limitation was that other potential factors, such as dietary and physical activity, were not measured to seek their association with BMI. Finally, when dealing with children in school, there was limited time available due to concern of interrupting the teaching and learning session. Thus, this study only focuses on selected variables and could not include others, as this age group of the population has some limitations, as mentioned. The data collected may be less accurate and less meaningful if too many questions or instruments are applied during data collection.

RECOMMENDATION FOR FUTURE RESEARCH

Based on the study's limitations, more OW and obese children should be included by choosing more schools, including vernacular schools. In addition, more research on psychological problems in obese children is required, as these issues can worsen over time. Since only a few factors are investigated in the current research, a follow-up study is needed to validate the results in Malaysian children who are OW or obese, as well as to identify a broader range of other HBM variables such as the perceived threat of obesity and cues to action for weight loss.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Based on understanding the three studied psychological perceptions of weight reduction among OW and obese children, a comprehensive and similar intervention programme could be designed for both groups because their perception of weight reduction shows no significant difference. Such programmes should emphasise the benefits of weight reduction

and other alternative activities that should be implemented, such as counselling and behaviour intervention to overcome the obstacles of weight reduction. Subsequently, this may raise awareness among family members or community members to start a healthier lifestyle. The primary school level can be an effective time to teach positive psychological beliefs about a healthy lifestyle and healthy body weight. Therefore, psychological factors should be considered in any future programme to improve children's BMI. The finding may have implications for designing a better health promotion, wellness educational programme, or intervention programme among OW and obese children.

ACKNOWLEDGMENTS

The authors thank all participants, parents, and teachers for their cooperation and willingness to be part of this study and all Nutritionists in the Kedah State Department of Health, Malaysia, for contributing to it.

REFERENCES

- Ahmad, A., Zulaily, N., Shahril, M. R., Fadzli, E., Syed, H., & Ahmed, A. (2018). Association between socioeconomic status and obesity among 12-year-old Malaysian adolescents. *PLoS ONE*, *13*(7), Article e0200577. <https://doi.org/10.1371/journal.pone.0200577>
- Allison, K. R., Dwyer, J. J. M., & Makin, S. (1999). Perceived barriers to physical activity among high school students. *Preventive Medicine*, *28*(6), 608-615. <https://doi.org/10.1006/pmed.1999.0489>
- Andajani-Sutjahjo, S., Ball, K., Warren, N., Inglis, V., & Crawford, D. (2004). Perceived personal, social and environmental barriers to weight maintenance among young women: A community survey. *International Journal of Behavioral Nutrition and Physical Activity*, *1*(1), 1-7. <https://doi.org/10.1186/1479-5868-1-15>
- Anuar, M. Z., Lim, C. T., Low, W. Y., & Harun, F. (2005). Factors affecting nutritional status of Malaysian primary school children. *Asia-Pacific Journal of Public Health*, *17*(2), 71-80. <https://doi.org/10.1177/101053950501700203>
- Buscemi, S., Castellini, G., Batsis, J. A., Ricca, V., Sprini, D., Galvano, F., Grosso, G., Rosafio, G., Caravello, M., & Rini, G. B. (2013). Psychological and behavioural factors associated with long-term weight maintenance after a multidisciplinary treatment of uncomplicated obesity. *Eating and Weight Disorders*, *18*.4, 351-358. <https://doi.org/10.1007/s40519-013-0059-2>
- Champion, V. L., & Skinner, C. S. (2008). The health belief model. In D. Kasprzyk (Ed.), *Health Behavior and Health Education* (pp. 45-65). John Wiley & Sons, Inc.
- da Cunha Feio Costa, L., Silva, D. A. S., dos Santos Alvarenga, M., & de Assis Guedes de Vasconcelos, F. (2016). Association between body image dissatisfaction and obesity among schoolchildren aged 7-10 years. *Physiology and Behavior*, *160*, 6-11. <https://doi.org/10.1016/j.physbeh.2016.03.022>
- Department of Statistics Malaysia. (2017). *Press release report of household income and basic amenities survey 2016*. Department of Statistics Malaysia. <https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=RUZ5REwveU1ra1hGL21JWVlPRmU2Zz09>

- Department of Statistics Malaysia. (2020). *Glossary A-Z*. Department of Statistics Malaysia. https://www.dosm.gov.my/v1/index.php?r=column/cglossary2&menu_id=eWd2VFdIZ2xpdzBmT2Y0a0pweDcwQT09&keyword=SjNETFJISStNdFk1d0szYmpYcVU4QT09&release=1#
- Faghri, P., Simon, J., Medina T, H., & Gorin, A. (2016). Effects of self-efficacy on health behavior and body weight. *Journal of Obesity & Weight Loss Therapy*, 06(06), Article 329. <https://doi.org/10.4172/2165-7904.1000329>
- Fitzgerald, N., & Spaccarotella, K. (2009). Barriers to a healthy lifestyle: From individuals to public policy. *An Ecological Perspective*, 47(1), 1-3.
- Gnavi, R., Spagnoli, T. D., Galotto, C., Pugliese, E., Carta, A., & Cesari, L. (2016). Socioeconomic status, overweight and obesity in prepuberal children: A study in an area of Northern Italy. *European Journal of Epidemiology*, 16(9), 797-803.
- IPH. (2017). *National health and morbidity survey (NHMS) 2017: Key findings from the adolescent health and nutrition surveys*. Institute for Public Health. <https://iku.moh.gov.my/images/IKU/Document/REPORT/NHMS2017/NHMS2017Infographic.pdf>
- IPH. (2019). *National health and morbidity survey 2019: NCDs – Non-communicable diseases: Risk factors and other health problems* (Technical Report – Volume 1). Institute of Public Health. https://iku.moh.gov.my/images/IKU/Document/REPORT/NHMS2019/Report_NHMS2019-NCD_v2.pdf
- Jabatan Pendidikan Negeri Kedah. (2019). *Maklumat asas pendidikan Jabatan Pendidikan Negeri Kedah* (Edisi Bilangan 3/2019) [Basic education information Kedah State Education Department (3/2019 Ed.)]. Jabatan Pendidikan Negeri Kedah.
- Kim, N. Y., Jeong, L. S., & Kim, J. S. (2011). A comparative study on the self-efficacy and health promoting behavior between obese and normal weight middle school students. *Journal of Korean Academy of Community Health Nursing*, 13(3), 828-837.
- Lazzeretti, L. (2015). Assessment of psychological predictors of weight loss: How and what for? *World Journal of Psychiatry*, 5(1), article 56. <https://doi.org/10.5498/wjp.v5.i1.56>
- Lemeshow, S., Hosmer, D. W., Klar, J., Lwanga, S. K., & World Health Organization. (1990). *Adequacy of Sample Size in Health Studies*. Chichester.
- Lim, Y. H., Park, M. S., Kim, Y., Kim, H., & Hong, Y. C. (2015). Effects of cold and hot temperature on dehydration: A mechanism of cardiovascular burden. *International Journal of Biometeorology*, 59(8), 1035-1043. <https://doi.org/10.1007/s00484-014-0917-2>
- MASCO. (2020). *Malaysia standard classification of occupations 2013*. Ministry of Human Resources Malaysia. https://www.dosm.gov.my/v1/uploads/files/4_Portal%20Content/3_Methods%20%26%20Classifications/2_List%20of%20References/updated%202019/MASCO_2013_EBOOK_BI.pdf
- Manios, Y., Kourlaba, G., Kondaki, K., Grammatikaki, E., Anastasiadou, A., & Roma-Giannikou, E. (2009). Obesity and television watching in preschoolers in greece: The genesis study. *Articles Epidemiology, Obesity*, 17(11), 2047-2053. <https://doi.org/10.1038/oby.2009.50>
- Miri, S. F., Javadi, M., Lin, C., Irandoost, K., & Pakpour, A. H. (2017). Health related quality of life and weight self-efficacy of life style among normal-weight, overweight and obese Iranian adolescents: A case control study. *International Journal of Pediatrics*, 5(47), 5975-5984. <https://doi.org/10.22038/ijp.2017.25554.2173>

- Moore, S. E., Harris, C., & Wimberly, Y. (2010). Perception of weight and threat to health. *Journal of the National Medical Association, 102*(2), 119-125. [https://doi.org/10.1016/s0027-9684\(15\)30499-5](https://doi.org/10.1016/s0027-9684(15)30499-5)
- Mostafavi, F., Salahshoori, A., Sharifirad, G., & Hassanzadeh, A. (2014). An assessment of the role of perceived benefits, barriers and self-efficacy in predicting dietary behavior in male and female high school students in the city of Izeh, Iran. *Journal of Education and Health Promotion, 3*(8). <https://doi.org/10.4103/2277-9531.127558>
- Mukherjee, S., Leong, H. F., & Wong, X. X. (2016). Waist circumference percentiles for Singaporean children and adolescents aged 6-17 years. *Obesity Research and Clinical Practice, 10*, S17-S25. <https://doi.org/10.1016/j.orcp.2016.04.004>
- Musaiger, A. O., Al-Mannai, M., Tayyem, R., Al-Lalla, O., Ali, E. Y. A., Kalam, F., Benhamed, M. M., Saghir, S., Halahleh, I., Djoudi, Z., & Chirane, M. (2013). Perceived barriers to healthy eating and physical activity among adolescents in seven arab countries: A cross-cultural study. *The Scientific World Journal, 2013*, Article 232164. <https://doi.org/10.1155/2013/232164>
- Norimah, A. K., Nasir, M. T. M., Hazizi, A. S., Suraya, P., Loh, S. H., & Nurliyana, A. R. (2014). Association of body weight status and socio-demographic factors with food habits among preschool children in Peninsular Malaysia. *Malaysian Journal of Nutrition, 20*(3), 303-315.
- Okop, K. J., Mukumbang, F. C., Mathole, T., Levitt, N., & Puoane, T. (2016). Perceptions of body size, obesity threat and the willingness to lose weight among black South African adults: A qualitative study. *BMC Public Health, 16*(1), 1-13. <https://doi.org/10.1186/s12889-016-3028-7>
- Park, D. Y. (2011). Utilizing the health belief model to predicting female middle school behavioral intention of weight reduction by weight status. *Nutrition Research Practice, 5*(4), 337-348. <https://doi.org/10.4162/nrp.2011.5.4.337>
- Rodríguez-Ventura, A. L., Pelaez-Ballestas, I., Sámano-Sámano, R., Jimenez-Gutierrez, C., & Aguilar-Salinas, C. (2014). Barriers to lose weight from the perspective of children with OW/obesity and their parents: A sociocultural approach. *Journal of Obesity, 2014*, Article 575184. <https://doi.org/10.1155/2014/575184>
- Saghafi-Asl, M., Aliasgharzadeh, S., & Asghari-Jafarabadi, M. (2020). Factors influencing weight management behavior among college students: An application of the health belief model. *PLoS ONE, 15*(2), 1-15. <https://doi.org/10.1371/journal.pone.0228058>
- Sánchez-Villegas, A., Pimenta, A. M., Beunza, J. J., Guillen-Grima, F., Toledo, E., & Martínez-González, M. A. (2010). Childhood and young adult OW/obesity and incidence of depression in the SUN project. *Obesity, 18*(7), 1443-1448. <https://doi.org/10.1038/oby.2009.375>
- Serour, M., Alqhenaei, H., Al-Saqabi, S., Mustafa, A. R., & Ben-Nakhi, A. (2007). Cultural factors and patients' adherence to lifestyle measures. *British Journal of General Practice, 57*(537), 291-295.
- Sharifi, N., Mahdavi, R., & Ebrahimi-Mameghani, M. (2013). Perceived barriers to weight loss programs for overweight or obese women. *Health Promotion Perspectives, 3*(1), 11-22. <https://doi.org/10.5681/hpp.2013.002>
- Shin, H., Shin, J., Liu, P., & Dutton, G. R. (2011). Self-efficacy improves weight loss in overweight / obese postmenopausal women during a 6-month weight loss intervention. *Nutrition Research, 31*(11), 822-828. <https://doi.org/10.1016/j.nutres.2011.09.022>

- Simmonds, M., Llewellyn, A., Owen, C. G., & Woolcott, N. (2016). Predicting adult obesity from childhood obesity: A systematic review and meta-analysis. *Obesity Reviews*, 17(2), 95-107. <https://doi.org/10.1111/obr.12334>
- Sit, C. H. P., Kerr, J. H., & Wong, I. T. (2008). Motives for and barriers to physical activity participation in middle-aged Chinese women. *Psychology of Sport and Exercise*, 9(3), 266-283. <https://doi.org/10.1016/j.psychsport.2007.04.006>
- Soliman, N. M., Elsayied, H. A. E., & Shouli, M. M. (2018). Application of HBM among youth at high risk for obesity in West Bank (Palestine). *American Journal of Nursing Science*, 7(3-1), 86-96. <https://doi.org/10.11648/j.ajns.s.2018070301.23>
- Tergerson, J. L., & King, K. A. (2002). Do perceived cues, benefits, and barriers to physical activity differ between male and female adolescents? *Journal of School Health*, 72(9), 374-380. <https://doi.org/10.1111/j.1746-1561.2002.tb03562.x>
- Tseng, M. C., Lee, Y. J., Chen, S. Y., Lee, M. B., Lin, K. H., Chen, P. R., & Lai, J. S. (2002). Psychobehavioral response and weight loss prediction in a hospital-Based weight reduction program. *Journal of the Formosan Medical Association*, 101(10), 705-711.
- WHO. (2019). *Growth reference data for 5-19 years*. World Health Organisation. <https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age>
- WHO. (2021). *Obesity and overweight: Fact sheet*. World Health Organisation. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

Evaluation of Acoustical Performance for Atrium Design with Respect to Skylight Geometry and Material in the Tropics

Nazli Che Din*, Tio Seng Zhi and Chia Keng Soon

The Centre for Building, Construction and Tropical Architecture (BuCTA), Faculty of Built Environment, Universiti Malaya, 50603 UM, Kuala Lumpur, Malaysia

ABSTRACT

Atriums, in most conditions, are considered the heart of buildings, the main spaces where different functions and activities are held. Nowadays, various atrium designs are incorporated in the building, varying from pyramidal to barrel vault. Therefore, it is of great importance to study the basic characteristics of sound fields in such spaces and their related materials. This research aims to determine the effect of various skylight geometry configurations on the atrium's acoustical performance. Next, an acoustical performance evaluation of the effect of skylight materials was performed by using computer simulation. Four selected atriums with different types of skylight geometries modeling and respective materials, i.e., glass, polycarbonate, and acrylic, have been used in this acoustical simulation. From the results, it can be concluded that an atrium with a pyramidal skylight provides a better reverberation time for music purposes. In contrast, all atriums' models' speech transmission index values were almost identical. However, no significant acoustical performance can be found in different used materials. The result also indicates that the shape of the atrium has a higher effect on the acoustical performance than the different materials used.

ARTICLE INFO

Article history:

Received: 16 December 2021

Accepted: 3 February 2022

Published: 19 September 2022

DOI: <https://doi.org/10.47836/pjst.30.4.35>

E-mail addresses:

nazlichedin@um.edu.my (Nazli Che Din)

z.tio@hotmail.com (Tio Seng Zhi)

soon_chia91@hotmail.com (Chia Keng Soon)

*Corresponding author

Keywords: Acoustical performance, atrium, geometry, ODEON, simulation, skylight

INTRODUCTION

A large open or covered space that continues through several levels of a building is an atrium in architecture. A courtyard house in

Ur, Mesopotamia, was identified as one of the traditional atrium forms in archaeological remains dating back to 3000 BC (Bednar, 1986). It was discovered as a central courtyard in ancient Roman and Greek architecture. The form of the atrium has gradually developed from a common traditional impression into high complexity methods to the difficulty in providing a sheltered courtyard (Saxon, 1983).

According to De Ruiter (1988), an atrium is a space with a large and high volume covered by glass-roofed. The atrium is also usually involved in large open spaces linking several floors, and it offers potential advantages as a building form over a typical building configuration (Gritch & Eason, 2016). Generally located at the center of the building, the atrium can be acted as a key circulation and transition space to proceed to other spaces (Wood & Salib, 2013; Pitts, 2013; Passe & Battaglia, 2015). It also gives an impression of large volume and has the potential to provide good air circulation and light penetration towards indoor comfort within the space (Jalil et al., 2014; Bai et al., 2015; Moosavi et al., 2014; Moosavi et al., 2015). Nowadays, the atrium is widely used in residential, commercial, and institutional buildings and not only acts as protection from the climate conditions but also creates a significant visual effect on the existing space.

Apart from serving as a transition zone between two spaces within the building, the atrium now hosts a variety of functions and activities, such as musical dance performances and exhibitions (Lee, 2019). Therefore, acoustic design or treatment should take into consideration its function. Normally, a larger volume and having too many reflecting materials installed in the space will result in higher reverberation time; thus, the reverberation time in the atrium should be lower and shorter if used for the reception or gathering space to allow acceptable speech intelligibility (Tio, 2016).

These days, the construction of public spaces with large atriums or high volumes, such as shopping malls and mixed-use developments, has increased. The acoustics in such a space is difficult to control and can contribute to poor acoustic comfort in the atrium (Nowicka, 2020). Also, the changes to meet the acoustic parameter may be highly extensive in cost afterward, so the early stage of acoustic design for the atrium should consider. Furthermore, acoustically reflecting construction materials and finishes, such as glass, concrete, and brick, are commonly utilized in the atrium and have lower sound absorption characteristics in overall frequency ranges, but it depends on the thickness of the materials (Fediuk et al., 2021). As a result, the acoustic performances in such places are poor since the reverberation time is longer. Subsequently, acoustic comfort in the space will become uncomfortable for the users. However, with recent rapid technologies and material studies, some solutions have become available to control the problem (Fuchs, 2001; Urban et al., 2016; Rychtarikova et al., 2017a; Rychtarikova et al., 2017b).

Based on Valtonen's (2014) observation, a shopping mall nowadays is not only a place to quickly do their routine shopping, but the time spent in the centers of the building or

atrium becomes favored by an individual or group of users. Long reverberation time in the atrium reduces the intelligibility of speech and human comfort (Šimek & Chmelík, 2021; Alnuman & Altaweel, 2020; Lavasani et al., 2021; Dökmeci, 2009), may even prevent people from hearing the announcement from the public address (PA) system. Until now, fewer or fewer acoustical studies have been conducted compared to the daylighting and thermal comfort studies for the atrium space.

Fundamental of Atrium Shape and Relationship with Sound

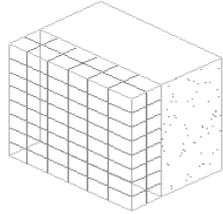
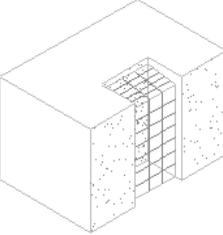
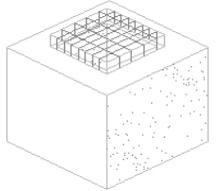
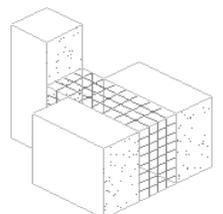
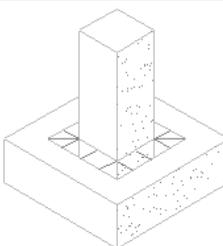
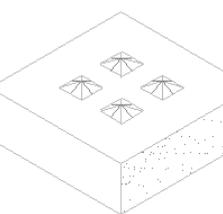
According to Gritch and Eason (2016), the shape and geometry of an atrium are largely determined by the product and rationale for the adjacent inhabited areas of the structure. Therefore, the arrangement or configuration of the atrium area has a significant impact on these places. As illustrated in Table 1, a few fundamental configurations of the atrium space are simple and complex. The simple types are Single Sided Atrium, Two Sided Atrium, Three Sided Atrium, Four Sided Atrium, and Linear Atrium. Furthermore, the complex types of the atrium can be divided into Bridging Atrium, Podium Atrium, Multiple Lateral Atrium, and Multiple Vertical Atrium. The minimalist atrium design is appropriate for modest single structures, while the complex atrium style is appropriate for massive complex buildings.

One of the key elements deciding the reverberation period in acoustics is the volume, shape, and size of space (Long, 2014). The reverberation time in a large volume of spaces will be longer, resulting in poor speech intelligibility (Jalil et al., 2016). When a sound source is applied in a room, sound waves can propagate from the source of sound until they hit the room's boundary. When a sound source is used in a room, sound waves scatter from the source until they reach the room's perimeter. Sound energy is reflected into the room space during propagation; some are absorbed and transmitted throughout the barrier. As a result, a room's volume, shape, and size changes may impact its acoustics (Maekawa et al., 2010). Sound rays from the sound source will focus the sound waves by reflecting the focal point in geometric shapes, such as a concave surface, as seen in Figure 1(a). Meanwhile, the surface in a convex shape will disperse the sound wave, as depicted in Figure 1(b). When compared to high-frequency sounds of short wavelength, the effect of sound ray diffraction is more obvious during long wavelength for low frequency (Ginn, 1978).

This study is divided into two objectives: i) to identify the acoustical condition of selected atrium design by using computer simulation, and ii) to evaluate the effectiveness of selected skylight materials on the acoustical performance of selected atrium space. In order to study and evaluate the acoustical performance in the atrium space, a series of acoustical simulations and analyses in different types of the atrium will be carried out using specialized room-acoustic software. The result of the acoustical performance of the selected atrium with different materials on the skylight will be generated and compared.

Table 1

Example of simple and complex basic configurations of atrium space (Gritch & Eason, 2016)

	<p>Simple Type Single-Sided Atrium: The atrium borders one side of the structure's occupied area</p>
	<p>Simple Type Three-Sided Atrium: The atrium borders three sides of the structure's occupied area</p>
	<p>Simple Type Four-sided Atrium: The atrium runs along four sides of the structure's occupied area</p>
	<p>Complex Type Bridging Atrium: The atrium connects several occupied areas of the structure.</p>
	<p>Complex Type Podium Atrium: The atrium sits at the bottom or below an occupied area of the structure</p>
	<p>Complex Type Multiple Lateral Atrium: Atrium areas spread throughout the layout on single or several stories.</p>

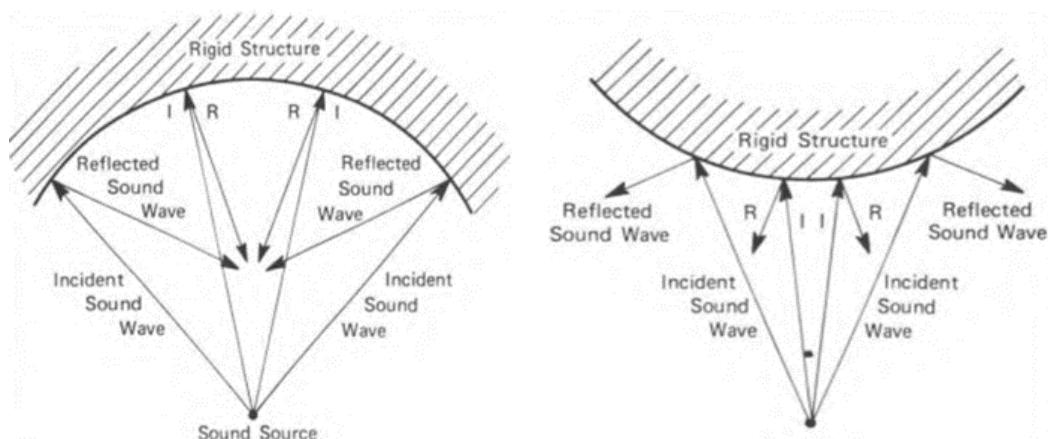


Figure 1. Reflections of sound rays on the curved surface; a) concave b) convex (Ginn, 1978)

Source. <https://www.bksv.com/media/doc/bn1329.pdf>

METHODS

Simulation Set-Up

Geometric Proportion of Atrium Space. The relationship between the length, width, and height of the atrium defines the differing degree of its light levels. The geometrical proportions of the atrium are similar to the well index ratio. Therefore, the shorter and broader atrium space will have improved daylight visibility because the height of the atrium defines the degree of the visible sky component. The Section Aspect Ratio (SAR), Plan Aspect Ratio (PAR), and Well Index Ratio (WI) can all be used to portray atrium geometry (Bednar, 1986; Saxon, 1983).

The height-to-width ratio determines the sectional proportion of the atrium. The lower SAR value suggests a lower atrium. For example, a shallow atrium has a SAR of less than 1.0, while a tall or narrow atrium has a SAR of larger than 2.0. (Atif, 1994). The PAR is the width-to-length ratio that defines whether the atrium is linear or square in plan proportion. A linear atrium is defined as one with a PAR of less than 0.4. (Bednar, 1986; Atif, 1994). Both the PAR and the SAR are included in the WI. It is also the best way to figure out how geometry affects the distribution of daylight illumination in the atrium. It is often the logarithmic scale index used to produce preliminary daylight predictions for various amounts of atrium volume (Atif, 1994; Boyer & Oh, 1988; Kim, 1987). The geometrical correlations are summarized as Equations 1, 2, and 3:

$$SAR = h / w \quad (1)$$

$$PAR = w / l \quad (2)$$

$$WI = h (l + w) / 2 (l) (w) \quad (3)$$

where, h = atrium height, w = atrium width, l = atrium length

Aside from the atrium's geometric proportions, the surface reflectivity of the well influences the quality and amount of sunshine inside the atrium, including the walls and floor. It is mostly determined by the type of materials utilized and the colors, textures, and finishes. When the surface is matte-finished with dispersed reflections, the polished and bright surface causes reflections. The ratio of the concrete wall to the glass surface determines the intensity of daylight created between the upper and lower floors of the atrium (Saxon, 1983; Navvab, 1990; Quek, 1989). Clear glass windows are less reflecting than solid walls. Because direct light is acquired mostly on the higher levels, a larger solid wall area is necessary to reflect the sun. The window area and ceiling height rise steadily towards the lower floors to obtain reflected light.

Skylight Glazing. Skylight glazing is a type of light-transmitting fenestration (elements that fill building envelope apertures) that forms all or a portion of a structure's roof for daylighting purposes. Several types of skylights are available in a wide range of sizes and shapes. It can improve the building's interior aesthetic by harmonizing the design form and characteristics of light penetration from the skylight. It is usually worked by properly selecting the skylights that enhance the ceiling grid and room proportions. A range of glass and plastic glazing materials, such as acrylics, polycarbonates, and fiberglass, are regularly available on the market for skylights. These materials are also available in various thicknesses and combination structure materials. Therefore, all these variables may affect the acoustical performance of the atrium.

Simulation 1. The major goal of this study was to determine the acoustical performance of various types of atriums. Therefore, choosing the sort of atrium was the first step. Generally, the atrium's basic model is based on the literature research "Design Principles of Atriums Buildings for the Tropic" (Ahmad & Rasdi, 2000), which identified the characteristics of a typical Malaysian atrium.

According to a review of the literature, the four-sided rectangular atrium with top-lit is the most prevalent atrium design in Malaysia; Plan-Aspect-Ratio (PAR) of 1:3; Section-Aspect-Ratio (SAR) of 1:1; The atrium void height should be four levels, and the floor to floor height should be 12 feet (about 4 m); The corridor floor to ceiling height should be 9 feet (about 3 meters), and the corridor width should be 12 feet (roughly 4 meters), with 1.2 meters of transparent glass balustrade or railing. Table 2 shows the details of the description of each room model created in Simulation 1 and has a similar size with approximately 643 m² of floor area in all models.

Four types of the atrium were selected in this research. The atrium is a model with different skylight geometry with the same materials used, as shown in Table 3. The four different types of atrium skylight are the Single Pitched (F1), Double Pitched (F2), Barrel

Vault (F3), and Pyramid (F4). For the computational study of computer modeling, the architectural characteristics of these atrium models have been streamlined. This simulation purposely identifies the effectiveness of the skylight geometry so that the materials assigned in this study are not identical with any existing atrium and are to be used similarly for all models.

Table 2

Description of each room model in Simulation 1 & 2 (sound source in red dots and receiver in blue dots)

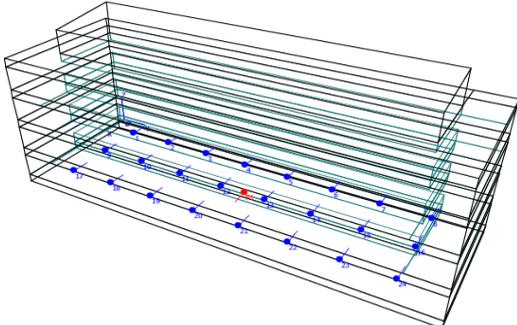
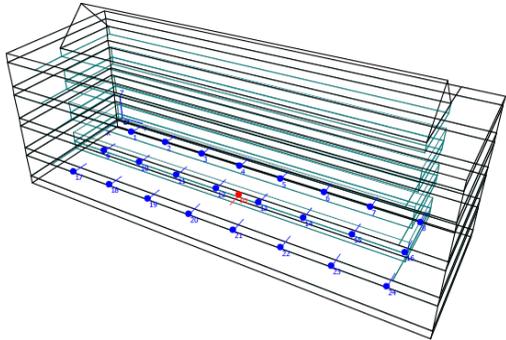
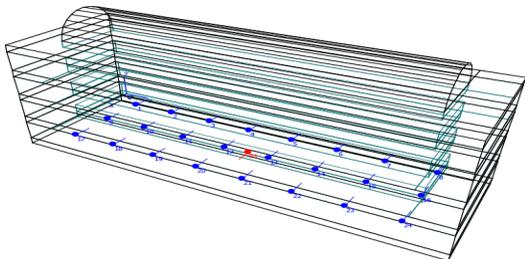
Model	Volume, m ³	Max Height, m
 <p>F1: Single Pitch Shape</p>	10566.3	18.22
 <p>F2: Double Pitch Shape</p>	10331.8	20.74
 <p>F3: Barrel Vault Shape</p>	10470.2	19.87

Table 2 (Continue)

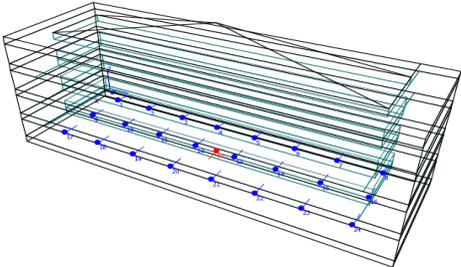
Model	Volume, m ³	Max Height, m
	10651.5	22.29
F4: Pyramid Shape		

Table 3

Absorption coefficients of materials applied for Simulation 1 (Ref. from ODEON material library)

Materials	Frequency (Hz)			
	250	500	1000	2000
Floor: Ceramic tiles. Perforation = 12%. Mineral wool in a cavity (Stroem, 1979)	0.44	0.68	0.79	0.56
Walls: Smooth brickwork, 10mm dep pointing, pit sand mortar (Kristensen, 1984)	0.09	0.12	0.16	0.22
Ceiling: Perf. 27 mm gypsumboard (16%), d= 4.5mm 300mm from ceiling (Dalenback, CATT)	0.55	0.60	0.90	0.86
Skylight Glazing: 6mm Single pane glass (Ref. Multiconsult, Norway)	0.06	0.04	0.03	0.02
Railing: Solid Glass block (Ref. Multiconsult, Norway)	0.02	0.02	0.02	0.02

Table 4

Absorption coefficients of materials applied for Simulation 2 (Ref. from ODEON material library)

Materials	Frequency (Hz)			
	250	500	1000	2000
Glass: 6 mm Single pane glass (Ref. Multiconsult, Norway)	0.06	0.04	0.03	0.02
Polycarbonate: 6 mm polycarbonate sheet (Ref. Panelite, 2016)	0.20	0.16	0.15	0.10
Acrylic: 6 mm acrylic sheet (Ref. Deamp, Norway)	0.25	0.20	0.18	0.10

Table 5

Sound power of an omnidirectional speaker was used in both simulations (Omni.SO8)

Frequency, Hz	250	500	1000	2000
Sound power, dB	69.6	74.8	71.8	63.8

Simulation 2. Using the inadequate acoustical performance of skylight geometry derived from Simulation 1, the next goal of this study was to assess the effectiveness of selected skylight materials on the acoustical performance of a specified atrium space.

A selected atrium model with similar materials is used except for the skylight glazing. Furthermore, two additional materials have been assigned (Table 4), i.e., polycarbonate and acrylic, to compare the impact of the acoustical performance in the atrium spaces. The motivation for materials used in this simulation is based on the observation that common materials used for atrium skylights in Malaysia include glass, polycarbonate, and acrylic.

Simulation Procedure. In the beginning, the four types of the selected atriums with different skylight geometries were modeled using Google Sketchup®, and then models were exported into ODEON Room Acoustic Software 13.0 Industrial (Cristensen & Koutsouris, 2015). A new room's validity was verified when assigned in ODEON to ascertain the simulation's accuracy. The verification process involves checking whether data is coherent and in the proper format. In addition, it involves a water tightness test of the room through tracing rays in the 3D Investigate Rays or 3D Billiard window, as depicted in Figure 2, to ensure that the room model is entirely enclosed. All architectural details such as an ornament, cornice, and framing were not included in the room modeling because such details do not produce any strong early reflections to the receiver (Jalil et al., 2019).

The sound source and receiver were defined systematically. First, a single point of natural raised sound (Table 5) was used as a sound source. The sound source was located in the center of the atrium and lifted 1.7 m from the floor level (stage height) to provide an even sound distribution in the atrium space. The receiver positions were then spread consistently across the entire floor space. Table 2 depicts the distribution of sound sources (red dots) and receiver sites (blue dots) for each room model. All simulation results from Simulations 1 and 2 were compared in two key parameters: reverberation time (RT) and speech transmission index (STI).

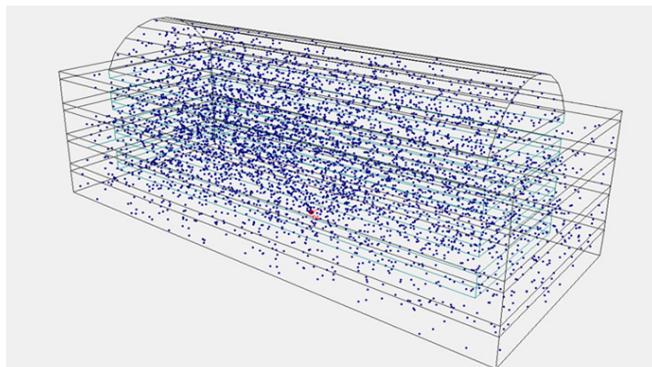


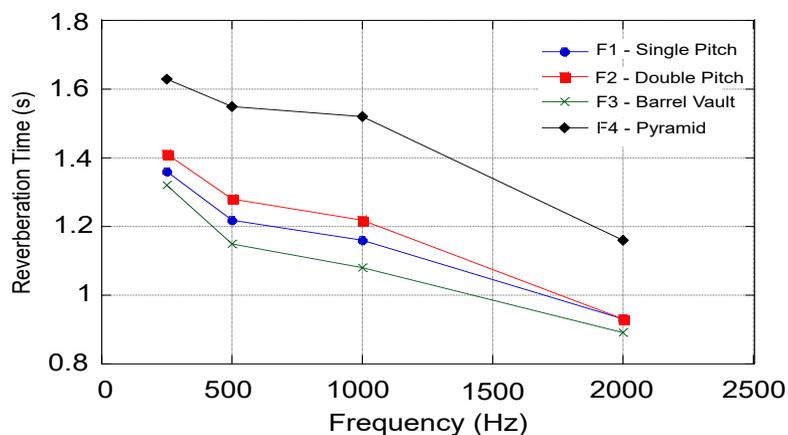
Figure 2. Example of the validation process using water tightness test by 3D billiard in the model room of F3 rectangular atrium with barrel vault-shaped skylight

RESULTS & DISCUSSIONS

Simulation 1: Acoustical Performance vs Skylight Geometry

Figure 3 shows the comparison of RT between all simulated models. A similar tendency of RT can be found in all models from 250 Hz until 2000 Hz. F4 exhibits higher RT values than the other atriums, followed by F2, F1, and F3. The Pyramid Skylight F4 may be better suited for a musical event due to its longer reverberation period, producing a warmer sound for musical activity. An atrium space considered good for a musical event has an RT value between 1.5 s to 1.8 s or a ‘fair’ value of 1.3 s to 1.5 s and 1.8 s to 1.9 s at 500 Hz and 1000 Hz, respectively (Egan, 2007). Based on the data obtained in Figure 3, all RT values decreased as the frequency increased. It was probably attributed to the fact that most materials do not absorb low frequencies well, so room reverberation is shorter at higher frequencies and longer at lower frequencies. This phenomenon is also supported by Ginn (1978) that the effect is more noticeable at low frequencies region.

Table 6 shows that all room models achieved a good STI and have an almost identical average STI, around 0.7. In general, all room models have attained outstanding agreement and provided speech intelligibility at the surrounding middle location of the atrium. However, a lower STI value can be observed where the distance between the sound source and the receiver exceeds around 5 m. There is a clear correlation between the STI mean value and the distance between the sound source and the receiver point, where the STI value decreases with the increased distance from the sound source. Generally, it can be inferred that F4 has a better STI value for a strong subjective perception of speech intelligibility.



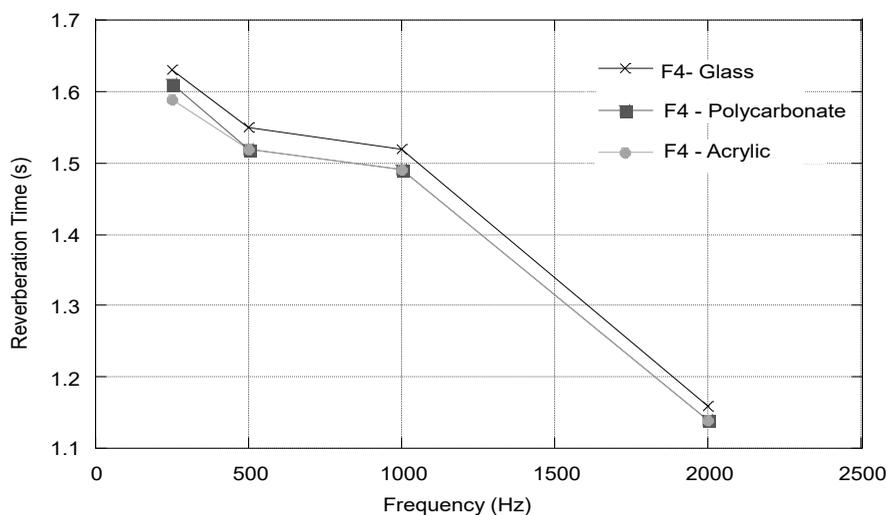
Atrium Type	T30 (s)			
	250 Hz	500 Hz	1000 Hz	2000 Hz
Single Pitch - F1	1.36	1.22	1.16	0.93
Double Pitch - F2	1.41	1.28	1.22	0.93
Barrel Vault - F3	1.32	1.15	1.08	0.89
Pyramid - F4	1.63	1.55	1.52	1.16

Figure 3. Comparison of reverberation time, T30 for each room model in Simulation 1

Table 6

Comparison of sound transmission index, STI for each model in Simulation 1

Parameter	Atrium Type			
	F1	F2	F3	F4
STI Max	0.86	0.87	0.87	0.85
STI Min	0.63	0.63	0.63	0.64
STI Ave	0.69	0.70	0.70	0.70
Distance (m) @STI-0.75	4.68	5.87	5.87	5.96



Materials	T30 (s)			
	250 Hz	500 Hz	1000 Hz	2000 Hz
Glass	1.63	1.55	1.52	1.16
Polycarbonate	1.61	1.52	1.49	1.14
Acrylic	1.59	1.52	1.49	1.14

Figure 4. Comparison of reverberation time, T30 for each room model in Simulation 2

Table 7

Comparison of sound transmission index, STI for each model in Simulation 2

Atrium pyramid skylight with material type			
Parameter	F4-Glass	F4-Polycarbonate	F4-Arylic
STI Max	0.85	0.86	0.86
STI Min	0.64	0.65	0.65
STI Ave	0.70	0.71	0.71
Distance (m) @STI-0.75	5.96	6.62	6.62

Simulation 2: Acoustical Performance vs Skylight Material

For Simulation 2, the F4 model was chosen and compared the two additional different types of materials used in the atrium skylight. A similar methodology was applied in this simulation, except the materials were assigned differently to evaluate their effectiveness in acoustical performance.

In this simulation, an atrium with similar skylight geometry and volume was used; the difference in RT value is caused by the difference in the absorption coefficient of materials used at the atrium skylight. However, there is no distinct agreement for all simulated RT observed in Figure 4, whereby the maximum dispersion is only 0.04 s.

Furthermore, no substantial difference in STI value can be found in Table 7 except for the distance of STI at 0.75 when a comparison of different materials has been performed. The distance of STI at 0.75 with polycarbonate and acrylic was slightly increased compared to the glass type of material. However, it can be concluded that the additional two types of materials gave a less significant performance on STI even though higher absorptive materials were used compared to the material glass.

CONCLUSION

This study is a series of investigations on different type of atrium skylights and their impact of skylight materials on acoustical performance using a computer simulation. From the simulation result, it can be concluded that the atrium with a pyramid skylight provides a better and longer reverberation time for music function. However, a less significant acoustical performance can be observed in simulated results when different materials for atrium skylights are used. It can also be concluded that the shape of the atrium has a higher impact on the acoustical performance than the materials used in this study. This study contributes to further investigation, such as on-site psychological and physical measurements, for evaluating human acoustic comfort in different atriums.

ACKNOWLEDGEMENT

The authors declared no conflicts of interest concerning this article's research, authorship, and/or publication. This work was conducted as part of completing the Master of Architecture degree requirements and supported by the Department of Architecture, Faculty of Built Environment, Universiti Malaya.

REFERENCES

- Ahmad, M. H., & Rasdi, M. T. H. M. (2000). *Design principles of atrium buildings for the tropics*. Penerbit UTM.
- Alnuman, N., & Altaweel, M. Z. (2020). Investigation of the acoustical environment in a shopping mall and its correlation to the acoustic comfort of the workers. *Applied Sciences*, 10(3), Article 1170. <https://doi.org/10.3390/app10031170>
- Atif, M. R. (1994). *Daylighting and cooling of atrium buildings in warm climates: Impact of the top-fenestration and wall mass area*. Texas A&M University.
- Bai, G., Gong, G., Yu, C. W., & Zhen, O. (2015). A combined, large, multi-faceted bulbous façade glazed curtain with open atrium as a natural ventilation solution for an energy efficient sustainable office building in Southern China. *Indoor and Built Environment*, 24(6), 813-832. <https://doi.org/10.1177/1420326X15602048>
- Bednar, M. J. (1986). *The New Atrium*. McGraw-hill book company.
- Boyer, L. L., & Oh, M. S. (1988). Computer prediction and measurement comparison of daylighting performance in selected atrium buildings using SERI algorithms. *ASHRAE Transactions*, 94(1), 799-811.
- Cristensen, C. L., & Koutsouris, G. (2015). *Odeon room acoustics software version 13 (User's Manual)*. Odeon A/S.
- De Ruiter, E. P. J. (1988). Atria in shopping centres, office buildings and hospitals. In *Proceedings of the Institute of Acoustics* (Vol. 10, No. Part 8, pp. 299-307). Institute of Acoustics.
- Dökmeci, P. N. (2009). *Acoustical comfort evaluation in enclosed public spaces with a central atrium: a case study in food court of cepa shopping center, Ankara* (Doctoral dissertation). Bilkent Universitesi, Turkey.
- Egan, M. D. (2007). *Architectural acoustics*. J. Ross Publishing.
- Fediuk, R., Amran, M., Vatin, N., Vasilev, Y., Lesovik, V., & Ozbakkaloglu, T. (2021). Acoustic Properties of Innovative Concretes: A Review. *Materials*, 14(2), Article 398. <https://doi.org/10.3390/ma14020398>
- Fuchs, H. V. (2001). Alternative fibreless absorbers - New tools and materials for noise control and acoustic comfort. *Acta Acustica United with Acustica*, 87(3), 414-422.
- Ginn, K. B. (1978). *Architectural acoustics*. Brüel & Kjaer. <https://www.bksv.com/media/doc/bn1329.pdf>.
- Gritch, T., & Eason, B. (2016). *Building envelope design guide - Atria systems*. Whole Building Design Guide. http://www.wbdg.org/design/env_atria.php.

- Jalil, N. A. A., Din, N. C., & Daud, N. I. M. K. (2014). A literature analysis on acoustical environment in green building design strategies. In *Applied Mechanics and Materials* (Vol. 471, pp. 138-142). Trans Tech Publications Ltd. <https://doi.org/10.4028/www.scientific.net/AMM.471.138>.
- Jalil, N. A. A., Din, N. C., & Keumala, N. (2016). Assessment on acoustical performance of green office buildings in Malaysia. *Indoor and Built Environment*, 25(4), 589-602. <https://doi.org/10.1177/1420326X14559855>
- Jalil, N. A. A., Din, N. C., Keumala, N., & Razak, A. S. (2019). Effect of model simplification through manual reduction in number of surfaces on room acoustics simulation. *Journal of Design and Built Environment*, 19(3), 31-41. <https://doi.org/10.22452/jdbe.vol19no3.4>.
- Kim, K. S. (1987). *Development of daylighting prediction algorithms for atrium design*. Texas A&M University.
- Lavasani, M., Sluyts, Y., Urbán, D., & Rychtarikova, M. (2021, October 25-27). Acoustic comfort evaluation based on architectural aspects in atria. In *Proceedings of Euronoise 2021* (pp. 1-9). Madera, Portugal. <http://ftp.sea-acustica.es/fileadmin/Madeira21/ID269.pdf>
- Lee, J. H. (2019). Identifying spatial meanings of atria in built environment and how they work. *Journal of Asian Architecture and Building Engineering*, 18(3), 247-261. <https://doi.org/10.1080/13467581.2019.1627216>
- Long, M. (2014). *Architectural Acoustics*. Elsevier.
- Maekawa, Z., Rindel, J., & Lord, P. (2010). *Environmental and architectural acoustics*. CRC Press. <https://doi.org/10.4324/9780203931356>
- Moosavi, L., Mahyuddin, N., & Ghafar, N. A. (2015). Atrium cooling performance in a low energy office building in the Tropics, a field study. *Building and Environment*, 94, 384-394. <https://doi.org/10.1016/j.buildenv.2015.06.020>
- Moosavi, L., Mahyuddin, N., Ghafar, N. A., & Ismail, M. A. (2014). Thermal performance of atria: An overview of natural ventilation effective designs. *Renewable and Sustainable Energy Reviews*, 34, 654-670. <https://doi.org/10.1016/j.rser.2014.02.035>
- Navvab, M. (1990). Outdoors indoors. *Daylighting within atrium spaces, LD+ A*, 20(5), 6-31.
- Nowicka, E. (2020). The acoustical assessment of the commercial spaces and buildings. *Applied Acoustics*, 169, Article 107491. <https://doi.org/10.1016/j.apacoust.2020.107491>
- Passe, U., & Battaglia, F. (2015). *Designing Spaces for Natural Ventilation: An architect's guide*. Routledge.
- Pitts, A. (2013). Thermal comfort in transition spaces. *Buildings*, 3(1), 122-142. <https://doi.org/10.3390/buildings3010122>
- Quek, C. K. (1989). *Design of atrium. Building in the warm humid tropics* (Unpublished M. Phil. dissertation). Darwin College Cambridge, France.
- Rychtarikova, M., Urban, D., Kassakova, M., Maywald, C., & Glorieux, C. (2017b). Perception of acoustic comfort in large halls covered by transparent structural skins. In *Proceedings of Meetings on Acoustics 173EAA* (Vol. 30, No. 1, p. 015005). Acoustical Society of America.
- Rychtarikova, M., Urban, D., Maywald, C., Zelem, L., & Kassakova, M. (2017a). Advantages of ETFE in terms of acoustic comfort in atria and large halls. In *Proceedings of Advanced Building Skins* (pp. 646-654). Advanced Building Skins GmbH.

- Saxon, R. (1983). *Atrium buildings: Development and design*. Architectural Press.
- Šimek, R., & Chmelík, V. (2021, October 25-27). Investigation of acoustical phenomenon in atria covered by structural glass roof. In *Proceedings of Euronoise 2021* (pp. 1-6). Madera, Portugal. <http://ftp.sea-acustica.es/fileadmin/Madeira21/ID232.pdf>
- Tio, S. Z. (2016). *Evaluation of acoustical performance for atrium design by using computer simulation model* (Master dissertation). Universiti Malaya, Malaysia.
- Urban, D., Zrnková, J., Zát'ko, P., Maywald, C., & Rychtáriková, M. (2016). Acoustic comfort in atria covered by novel structural skins. *Procedia Engineering*, 155, 361-368.
- Valtonen, M. (2014). *Acoustic design of a public space using perforated panel resonators* (Master Thesis). Aalto University, Finland.
- Wood, A., & Salib, R. (2013). *Guide to natural ventilation in high rise office buildings*. Routledge.



REFEREES FOR THE PERTANIKA JOURNAL OF SCIENCE & TECHNOLOGY

Vol. 30 (4) Oct. 2022

The Editorial Board of the Pertanika Journal of Science and Technology wishes to thank the following:

Abam Fidelis <i>(MOUUAU, Nigeria)</i>	Dipankar Pal <i>(IIPe, India)</i>	Kek Sie Long <i>(UTHM, Malaysia)</i>
Adilah Abdul Gapor <i>(UM, Malaysia)</i>	Dipesh Shikchand Patle <i>(MNNIT)</i>	Khairil Amir Sayuti <i>(USM, Malaysia)</i>
Adriana Irawati Nur Ibrahim <i>(UM, Malaysia)</i>	Elya Nabila Abd Bahri <i>(UM, Malaysia)</i>	Khanh Le Nguyen Quoc <i>(TMU, Taiwan)</i>
Ahmad Ilyas Rushdan <i>(UTM, Malaysia)</i>	Emienour Muzalina Mustafa <i>(UMT, Malaysia)</i>	Lim Gin Keat <i>(USM, Malaysia)</i>
Ahmad Khusairi Azemi <i>(USM, Malaysia)</i>	Erzam Marlisah <i>(UPM, Malaysia)</i>	Ma Nyuk Ling <i>(UMT, Malaysia)</i>
Ahmad Taufek Abdul Rahman <i>(UiTM, Malaysia)</i>	Fahrul Zaman Huyop <i>(UTM, Malaysia)</i>	Maryati Md Dasor <i>(UiTM, Malaysia)</i>
Ahmad Zhafran Ahmad Mazlan <i>(USM, Malaysia)</i>	Farshid Keynia <i>(GUAT, Iran)</i>	Masrah Azrifah Azmi Murad <i>(UPM, Malaysia)</i>
Ajune Wanis Ismail <i>(UTM, Malaysia)</i>	Farzaneh Mohamadpour <i>(USB, Iran)</i>	Maz Jamilah Masnan <i>(UNIMAP, Malaysia)</i>
Aman Mohd Ihsan Mamat <i>(UiTM, Malaysia)</i>	Gan Keng Hoon <i>(USM, Malaysia)</i>	Melody Kimi <i>(UNIMAS, Malaysia)</i>
Amit Grover <i>(SBSSTC, India)</i>	Gerald Franz <i>(UPJV, France)</i>	Mohamad Rezi Abdul Hamid <i>(UPM, Malaysia)</i>
Ayad Hameed Mousa <i>(University of Karbala, Iraq)</i>	Goh Eg Su <i>(UTM, Malaysia)</i>	Mohammad H Yarmohammadian <i>(MUI, Iran)</i>
Azizul Hakim Lahuri <i>(UPM, Malaysia)</i>	Hafizah Hanim Mohd Zaki <i>(IIUM, Malaysia)</i>	Mohd Faizal Fauzan <i>(UKM, Malaysia)</i>
Azwan Iskandar Azmi <i>(UNIMAP, Malaysia)</i>	Haghifam Mahmoud-Reza <i>(Tarbiat Modares University, Iran)</i>	Mohd Hanafi Ani <i>(IIUM, Malaysia)</i>
Bimo Ario Tejo <i>(UPM, Malaysia)</i>	Hani Syahida Zulkafli <i>(UPM, Malaysia)</i>	Mohd Najib Yacob <i>(UMP, Malaysia)</i>
Chandran Somasundram <i>(UM, Malaysia)</i>	Hauwau Ibrahim Mustapha <i>(MAAUN, Niger)</i>	Mohd Nizam Ahmad <i>(UniKL, Malaysia)</i>
Christopher Heng Yii Sern <i>(USM, Malaysia)</i>	Iskandar Ishak <i>(UPM, Malaysia)</i>	Mohd Tahir Ismail <i>(USM, Malaysia)</i>
Chua Han Bing <i>(Curtin University, Malaysia)</i>	Jaffar Syed Mohamed Ali <i>(IIUM, Malaysia)</i>	

Muhammad Abbas Ahmad Zaini (UTM, Malaysia)	Nur Syakina Jamali (UPM, Malaysia)	Sayed Sayeed Ahmad (AGU, United Arab Emirates)
Muhammad Hanif Ramlee (UTM, Malaysia)	Patil Harshal (ADYPU, India)	Siti Rohaiza Ahmad (UBD, Brunei Darussalam)
Mus`ab Abd. Razak (UPM, Malaysia)	Peng Rui (BJUT, China)	Sreetheran Maruthaveeran (UPM, Malaysia)
Muzamir Isa (UNIMAP, Malaysia)	Rafidah Husen (UiTM, Malaysia)	Suhaili Abu Bakar @ Jamaludin (UPM, Malaysia)
Nani Draman (USM, Malaysia)	Raja Izamshah Raja Abdullah (UTeM, Malaysia)	Wan Nurul Huda Wan Zainal (UMP, Malaysia)
Noor Akhmazillah Mohd Fauzi (UTHM, Malaysia)	Rajyalakshmi K (KLU, India)	Wan Zakiyatussariroh Wan Husin (UiTM, Malaysia)
Nooranida Arifin (UM, Malaysia)	Razlina Abdul Rahman (USM, Malaysia)	Zakariya Yahya Algamal (University of Mosul, Iraq)
Nor Hanuni Ramli (UMP, Malaysia)	Rohayah Che Amat (UTM, Malaysia)	Zarani Mat Taher (UTM, Malaysia)
Norharyati Harum (UTeM, Malaysia)	Rusli Abdullah (UPM, Malaysia)	Zarirah Mohamed Zulperi (UPM, Malaysia)
Nur Fadhilah Ibrahim (UMT, Malaysia)	Samuel Olusegun David (FUPRE, Nigeria)	
Nur Hartini Mohd Taib (USM, Malaysia)	Sayed Inayatullah Shah (IIUM, Malaysia)	

ADYPU – Ajeenkya DY Patil University
AGU – Al Ghurair University
BJUT – Beijing University of Technology
FUPRE – Federal University of Petroleum Resources
GUAT – Graduate University of Advanced Technology
IIFE – Indian Institute of Petroleum & Energy
IIUM – International Islamic University Malaysia
KLU – Koneru Lakshmaiah Education
MAAUN – Maryam Abacha American University of Niger
MNNIT – Motilal Nehru National Institute of Technology
MOUAU – Michael Okpara University of Agriculture
MUI – Medical University of Isfahan
SBSSTC – Shaheed Bhagat Singh State Technical Campus
TMU – Taipei Medical University
UBD – Universiti Brunei Darussalam

UiTM – Universiti Teknologi MARA
UKM – Universiti Kebangsaan Malaysia
UM – Universiti Malaya
UMP – Universiti Malaysia Pahang
UMT – Universiti Malaysia Terengganu
UniKL – University Kuala Lumpur
UnimAP – Universiti Malaysia Perlis
UNIMAS – Universiti Malaysia Sarawak
UPIV – University of Picardie Jules Verne
UPM – Universiti Putra Malaysia
USB – University of Sistan & Baluchestan
USM – Universiti Sains Malaysia
UTeM – Universiti Teknikal Malaysia Melaka
UTHM – Universiti Tun Hussein Onn Malaysia
UTM – Universiti Teknologi Malaysia

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.

Pertanika Journal of Science & Technology

Our goal is to bring high-quality research to the widest possible audience

INSTRUCTIONS TO AUTHORS

(REGULAR ISSUE)

(Manuscript Preparation & Submission Guide)

Revised: November 2020

Please read the *Pertanika* guidelines and follow these instructions carefully. The Chief Executive Editor reserves the right to return manuscripts that are not prepared in accordance with these guidelines.

MANUSCRIPT PREPARATION Manuscript Types

Pertanika accepts submission of mainly 4 types of manuscripts

- that have not been published elsewhere (including proceedings)
- that are not currently being submitted to other journals

1. Regular article

Regular article is a full-length original empirical investigation, consisting of introduction, methods, results, and discussion. Original research work should present new and significant findings that contribute to the advancement of the research area. *Analysis and Discussion* must be supported with relevant references.

Size: Generally, each manuscript is **not to exceed 6000 words** (excluding the abstract, references, tables, and/or figures), a maximum of **80 references**, and **an abstract of less than 250 words**.

2. Review article

A review article reports a critical evaluation of materials about current research that has already been published by organising, integrating, and evaluating previously published materials. It summarises the status of knowledge and outlines future directions of research within the journal scope. A review article should aim to provide systemic overviews, evaluations, and interpretations of research in a given field. Re-analyses as meta-analysis and systemic reviews are encouraged.

Size: Generally, it is expected **not to exceed 6000 words** (excluding the abstract, references, tables, and/or figures), a maximum of **80 references**, and **an abstract of less than 250 words**.

3. Short communications

Each article should be timely and brief. It is suitable for the publication of significant technical advances and maybe used to:

- (a) reports new developments, significant advances and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope;
- (b) reports/discuss on significant matters of policy and perspective related to the science of the journal, including 'personal' commentary;
- (c) disseminates information and data on topical events of significant scientific and/or social interest within the scope of the journal.

Size: It is limited to **3000 words** and have a maximum of **3 figures and/or tables, from 8 to 20 references, and an abstract length not exceeding 100 words**. The information must be in short but complete form and it is not intended to publish preliminary results or to be a reduced version of a regular paper.

4. Others

Brief reports, case studies, comments, concept papers, letters to the editor, and replies on previously published articles may be considered.

Language Accuracy

Pertanika emphasises on the linguistic accuracy of every manuscript published. Articles can be written in **English** or **Bahasa Malaysia** and they must be competently written and presented in clear and concise grammatical English/Bahasa Malaysia. Contributors are strongly advised to have the manuscript checked by a colleague with ample experience in writing English manuscripts or a competent English language editor. For articles in Bahasa Malaysia, the title, abstract and keywords should be written in both English and Bahasa Malaysia.

Author(s) **may be required to provide a certificate** confirming that their manuscripts have been adequately edited. **All editing costs must be borne by the authors.**

Linguistically hopeless manuscripts will be rejected straightaway (e.g., when the language is so poor that one cannot be sure of what the authors are really trying to say). This process, taken by authors before submission, will greatly facilitate reviewing, and thus, publication.

MANUSCRIPT FORMAT

The paper should be submitted in **one-column format** with 1.5 line spacing throughout. Authors are advised to use Times New Roman 12-point font and *MS Word* format.

1. Manuscript Structure

The manuscripts, in general, should be organised in the following order:

Page 1: Running title

This page should **only** contain the running title of your paper. The running title is an abbreviated title used as the running head on every page of the manuscript. The running title **should not exceed 60 characters, counting letters and spaces.**

Page 2: Author(s) and Corresponding author's information

General information: This page should contain the **full title** of your paper **not exceeding 25 words**, with the name of all the authors, institutions and corresponding author's name, institution and full address (Street address, telephone number (including extension), handphone number, and e-mail address) for editorial correspondence. **The corresponding author must be clearly indicated with a superscripted asterisk symbol (*).**

Authors' name: The names of the authors should be named **in full without academic titles**. For Asian (Chinese, Korean, Japanese, Vietnamese), please write first name and middle name before surname (family name). The last name in the sequence is considered the surname.

Authors' addresses: Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers.

Tables/figures list: A list of the number of **black and white/colour figures and tables** should also be indicated on this page. See "**5. Figures & Photographs**" for details.

Example (page 2):

Fast and Robust Diagnostic Technique for the Detection of High Leverage Points

Habshah Midi^{1,2*}, Hasan Talib Hendi¹, Jayanthi Arasan² and Hassan Uraibi³

¹*Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

²*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

³*Department of Statistics, University of Al-Qadisiyah, 88 -Al-Qadisiyah -Al-Diwaniyah, Iraq*

E-mail addresses

habshah@upm.edu.my (Habshah Midi)

h.applied.t88@gmail.com (Hasan Talib Hendi)

jayanthi@upm.edu.my (Jayanthi Arasan)

hssn.sami1@gmail.com (Hassan Uraibi)

*Corresponding author

List of Table/Figure: Table 1.

Figure 1.

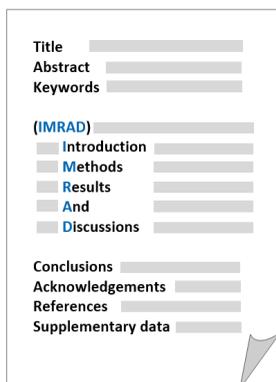
Page 3: Abstract

This page should **repeat** the **full title** of your paper with only the **Abstract**, usually in one paragraph and **Keywords**.

Keywords: *Not more than 8 keywords in alphabetical order must be provided to describe the content of the manuscript.*

Page 4: Text

A regular paper should be prepared with the headings *Introduction, Materials and Methods, Results and Discussions, Conclusions, Acknowledgements, References, and Supplementary data* (if any) in this order. The literature review may be part of or separated from the *Introduction*.



MAKE YOUR ARTICLES AS CONCISE AS POSSIBLE

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words Introduction, Materials and Methods, Results, And, Discussion. It indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: Title, Authors, Keywords, Abstract, Conclusions, and References. Additionally, some papers include Acknowledgments and Appendices.

The Introduction explains the scope and objective of the study in the light of current knowledge on the subject; the Materials and Methods describes how the study was conducted; the Results section reports what was found in the study; and the Discussion section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the Journal's instructions to authors.

2. Levels of Heading

Level of heading	Format
1 st	LEFT, BOLD, UPPERCASE
2 nd	Flush left, Bold, Capitalise each word
3 rd	Bold, Capitalise each word, ending with .
4 th	Bold italic, Capitalise each word, ending with .

3. Equations and Formulae

These must be set up clearly and should be typed double-spaced. Numbers identifying equations should be in square brackets and placed on the right margin of the text.

4. Tables

- All tables should be prepared in a form consistent with recent issues of *Pertanika* and should be numbered consecutively with Roman numerals (Table 1, Table 2).
- A brief title should be provided, which should be shown at the top of each table (APA format):

Example:

Table 1

PVY infected Nicotiana tabacum plants optical density in ELISA

- Explanatory material should be given in the table legends and footnotes.
- Each table should be prepared on a new page, embedded in the manuscript.
- Authors are advised to keep backup files of all tables.

**** Please submit all tables in Microsoft word format only, because tables submitted as image data cannot be edited for publication and are usually in low-resolution.**

5. Figures & Photographs

- Submit an original figure or photograph.
- Line drawings must be clear, with a high black and white contrast.
- Each figure or photograph should be prepared on a new page, embedded in the manuscript for reviewing to keep the file of the manuscript under 5 MB.
- These should be numbered consecutively with Roman numerals (Figure 1, Figure 2).
- Provide a brief title, which should be shown at the bottom of each table (**APA format**):

Example: Figure 1. PVY-infected in vitro callus of Nicotiana tabacum

- If a figure has been previously published, acknowledge the original source, and submit written permission from the copyright holder to reproduce the material.
- Authors are advised to keep backup files of all figures.

**** Figures or photographs must also be submitted separately as TIFF or JPEG, because figures or photographs submitted in low-resolution embedded in the manuscript cannot be accepted for publication. For electronic figures, create your figures using applications that are capable of preparing high-resolution TIFF files.**

6. Acknowledgement

Any individuals and entities who have contributed to the research should be acknowledged appropriately.

7. References

References begin on their own page and are listed in alphabetical order by the first author's last name. Only references cited within the text should be included. All references should be in 12-point font and double-spaced. If a Digital Object Identifier (DOI) is listed on a print or electronic source, it is required to include the DOI in the reference list. Use Crossref to find a DOI using author and title information.

NOTE: When formatting your references, please follow the **APA-reference style** (7th edition) (refer to the examples). Ensure that the references are strictly in the journal's prescribed style, failing which your article will **not be accepted for peer-review**. You may refer to the *Publication Manual of the American Psychological Association* (<https://apastyle.apa.org/>) for further details.

Examples of reference style are given below:

Books		
	Insertion in text	In reference list
Book/E-Book with 1-2 authors	<p>Information prominent' (the author's name is within parentheses):</p> <p>... (Staron, 2020)</p> <p>... (Darus & Rasdi, 2019)</p> <p>... Or</p> <p>'Author prominent' (the author's name is outside the parentheses):</p> <p>(Starron, 2020)...</p> <p>Darus and Rasdi (2019) ...</p>	<p>Staron, M. (2020). <i>Action research in software engineering</i>. Springer International Publishing. https://doi.org/10.1007/978-3-030-32610-4</p> <p>Darus, A., & Rasdi, I. (2019). <i>Introduction to occupational health a workbook</i>. UPM Press.</p>
Book/E-Book with 3 or more authors	<p><i>For all in-text references, list only the first author's family name and followed by 'et al.'</i></p> <p>Information prominent' (the author's name is within parentheses):</p> <p>... (Yusof et al., 2020)</p> <p>... Or</p> <p>'Author prominent' (the author's name is outside the parentheses):</p> <p>Yusof et al. (2020) ...</p>	<p>Yusof, N. A., Azmi, U. Z. M., Ariffin, N., & Rahman, S. F. A. (2020). <i>Biosensors and chemical sensors: A practical approach</i>. UPM Press.</p>
Book/E-Book with more than 20 authors		<p>For books with more than 20 authors, please follow the guidelines for journal articles with more than 20 authors.</p>
Chapter in an edited Book/E-Book	<p>Inform ation pr ominent' (the author 's name is within parentheses):</p> <p>... (Mainzer, 2020) ...</p> <p>... (Tang et al., 2020) ...</p> <p>Or</p> <p>'Author prominent' (the author's name is outside the parentheses):</p> <p>Mainzer (2020) ...</p> <p>Tang et al. (2020) ...</p>	<p>Mainzer, K. (2020). Logical thinking becomes automatic. In K. Mainzer (Ed.), <i>Artificial intelligence-When do machines take over?</i> (pp. 15-45). Springer. https://doi.org/10.1007/978-3-662-59717-0_3</p> <p>Tang, W., Khavarian, M., Yousefi, A., & Cui, H. (2020). Properties of self-compacting concrete with recycled concrete aggregates. In R. Siddique (Ed.), <i>Self-Compacting Concrete: Materials, Properties, and Applications</i> (pp. 219-248). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-817369-5.00009-X</p>

	Insertion in text	In reference list
Editor	<p>Information prominent' (the author's name is within parentheses): ... (Kesharwani, 2020) (Lanza et al., 2020) ... Or 'Author prominent' (the author's name is outside the parentheses): Kesharwani (2020) ... Lanza et al. (2020) ...</p>	<p>Kesharwani, P. (Ed.). (2020). <i>Nanotechnology based approaches for tuberculosis treatment</i>. Academic Press.</p> <p>Lanza, R., Langer, R., Vacanti, J. P., & Atala, A. (Eds.). (2020). <i>Principles of tissue engineering</i>. Academic press. https://doi.org/10.1016/C2018-0-03818-9</p>
Several works by the same author in the same year	<p>Information prominent' (the author's name is within parentheses): ... (Aggarwal & Aggarwal, 2020a, 2020b) ... Or 'Author prominent' (the author's name is outside the parentheses): Aggarwal & Aggarwal (2020a, 2020b) ...</p>	<p>Aggarwal, P., & Aggarwal, Y. (2020a). Strength properties of SCC. In R. Siddique (Ed.), <i>Self-Compacting Concrete: Materials, Properties, and Applications</i> (p. 83-115). Woodhead Publishing. doi: https://doi.org/10.1016/B978-0-12-817369-5.00004-0</p> <p>Aggarwal, P., & Aggarwal, Y. (2020b). Carbonation and corrosion of SCC. In R. Siddique (Ed.), <i>Self-Compacting Concrete: Materials, Properties, and Applications</i> (p. 147-193). Woodhead Publishing. doi: https://doi.org/10.1016/B978-0-12-817369-5.00007-6</p>
Journals		
Journal article with 1-2 authors	<p>Information prominent' (the author's name is within parentheses): ... (Laan & Fox, 2019) ... Or 'Author prominent' (the author's name is outside the parentheses): Laan and Fox (2019) ...</p>	<p>Laan, E., & Fox, J. W. (2019). An experimental test of the effects of dispersal and the paradox of enrichment on metapopulation persistence. <i>Oikos</i>, 129(1), 49-58. https://doi.org/10.1111/oik.06552</p>
Journal article with 3 or more authors	<p><i>For all in-text references, list only the first author's family name and followed by 'et al.'</i> Information prominent' (the author's name is within parentheses): ... (Midi et al., 2020) (Shagufta et al., 2017) ... Or 'Author prominent' (the author's name is outside the parentheses): Midi et al. (2020) ... Shagufta et al. (2017) ...</p>	<p>Midi, H., Hendi, H. T., Arasan, J., & Uraibi, H. (2020). Fast and Robust Diagnostic Technique for the Detection of High Leverage Points. <i>Pertanika Journal of Science & Technology</i>, 28(4), 1203-1220.</p> <p>Shagufta, B., Sivakumar, M., Kumar, S., Agarwal, R. K., Bhilegaonkar, K. N., Kumar, A., & Dubal, Z. B. (2017). Antimicrobial resistance and typing of Salmonella isolated from street vended foods and associated environment. <i>Journal of Food Science and Technology</i>, 54(8), 2532-2539. doi: https://doi.org/10.1007/s13197-017-2698-1</p>
Journal article with more than 20	<p>Information prominent' (the author's name is within parentheses): ... (Wiskunde et al., 2019) ... Or 'Author prominent' (the author's name is outside the parentheses): Wiskunde et al. (2019) ...</p>	<p>Wiskunde, B., Arslan, M., Fischer, P., Nowak, L., Van den Berg, O., Coetzee, L., Juárez, U., Riyaziyyat, E., Wang, C., Zhang, I., Li, P., Yang, R., Kumar, B., Xu, A., Martinez, R., McIntosh, V., Ibáñez, L. M., Mäkinen, G., Virtanen, E., ... Kovács, A. (2019). Indie pop rocks mathematics: Twenty One Pilots, Nicolas Bourbaki, and the empty set. <i>Journal of Improbable Mathematics</i>, 27(1), 1935-1968. https://doi.org/10.0000/3mp7y-537</p>
Journal article with an article number	<p>Information prominent' (the author's name is within parentheses): ... (Roe et al., 2020) ... Or 'Author prominent' (the author's name is outside the parentheses): Roe et al. (2020) ...</p>	<p>Roe, E. T., Bies, A. J., Montgomery, R. D., Watterson, W. J., Parris, B., Boydston, C. R., Sereno, M. E., & Taylor, R. P. (2020). Fractal solar panels: Optimizing aesthetic and electrical performances. <i>Plos One</i>, 15(3), Article e0229945. https://doi.org/10.1371/journal.pone.0229945</p>
Journal article with missing information	<p>Information prominent' (the author's name is within parentheses): ... (Alfirevic et al., 2017) (Hayat et al., 2020) (Fan et al., 2020) ...</p>	<p>Missing volume number Alfirevic, Z., Stampalija, T., & Dowswell, T. (2017). Fetal and umbilical Doppler ultrasound in high-risk pregnancies (review). <i>Cochrane Database of Systematic Reviews</i>, (6), 1-163. https://doi.org/10.1002/14651858.CD007529.pub4. Copyright</p>

	Insertion in text	In reference list
Journal article with missing information	Or 'Author prominent' (the author's name is outside the parentheses): Alfirevic et al. (2017) ... Hayat et al. (2020) ... Fan et al. (2020) ...	Missing issue number Hayat, A., Shaishta, N., Mane, S. K. B., Hayat, A., Khan, J., Rehman, A. U., & Li, T. (2020). Molecular engineering of polymeric carbon nitride based Donor-Acceptor conjugated copolymers for enhanced photocatalytic full water splitting. <i>Journal of colloid and interface science</i> , 560, 743-754. https://doi.org/10.1016/j.jcis.2019.10.088 Missing page or article number Fan, R. G., Wang, Y. B., Luo, M., Zhang, Y. Q., & Zhu, C. P. (2020). SEIR-Based COVID-19 Transmission Model and Inflection Point Prediction Analysis. <i>Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China</i> , 49(3). https://doi.org/10.12178/1001-0548.9_2020029
Several works by the same author in the same year	Information prominent' (the author's name is within parentheses): ... (Chee et al., 2019a, 2019b) ... Or 'Author prominent' (the author's name is outside the parentheses): Chee et al. (2019a, 2019b) ...	Chee, S. S., Jawaid, M., Sultan, M. T. H., Alothman, O. Y., & Abdullah, L. C. (2019a). Accelerated weathering and soil burial effects on colour, biodegradability and thermal properties of bamboo/kenaf/epoxy hybrid composites. <i>Polymer Testing</i> , 79, Article 106054. https://doi.org/10.1016/j.polymeresting.2019.106054 Chee, S. S., Jawaid, M., Sultan, M. T. H., Alothman, O. Y., & Abdullah, L. C. (2019b). Evaluation of the hybridization effect on the thermal and thermo-oxidative stability of bamboo/kenaf/epoxy hybrid composites. <i>Journal of Thermal Analysis and Calorimetry</i> , 137(1), 55-63. https://doi.org/10.1007/s10973-018-7918-z
Newspaper		
Newspaper article – with an author	... (Shamshuddin, 2019) ... Or ... Shamshuddin (2019) ...	Shamshuddin, J. (2019, September 23). Lynas plant waste residue can be used to boost oil palm growth? <i>New Straits Times</i> . https://www.nst.com.my/opinion/letters/2019/09/523930/lynas-plant-waste-residue-can-be-used-boost-oil-palm-growth
Newspaper article – without an author	("Zoonotic viruses," 2017). OR "Zoonotic viruses" (2017) ... Use a shortened title (or full title if it is short) in Headline Case enclosed in double quotation marks.	Zoonotic viruses like swine flu are ticking time bombs, say experts. (2020, July 4). <i>New Straits Times</i> , 3.
Dissertation/Thesis		
Published Dissertation or Thesis References	... (Rivera, 2016) ... Or ... Rivera (2016) ...	Rivera, C. (2016). <i>Disaster risk management and climate change adaptation in urban contexts: Integration and challenges</i> [Doctoral dissertation, Lund University]. Lund University Publications. https://lup.lub.lu.se/search/ws/files/5471705/8570923.pdf
Unpublished Dissertation or Thesis References	... (Brooks, 2014) ... Or ... Brooks (2014) ...	Brooks, J. D. (2015). <i>Bamboo as a strengthening agent in concrete beams for medium height structures</i> [Unpublished Doctoral dissertation]. The University of Washington.
Conference/Seminar Papers		
Conference proceedings published in a journal	... (Duckworth et al., 2019) ... Or Duckworth et al. (2019) ...	Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., & Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. <i>Proceedings of the National Academy of Sciences, USA</i> , 116(47), 23499-23504. https://doi.org/10.1073/pnas.1910510116
Conference proceedings published as a book chapter	... (Bedenel et al., 2019) ... Or Bedenel et al. (2019) ...	Bedenel, A. L., Jourdan, L., & Biernacki, C. (2019). Probability estimation by an adapted genetic algorithm in web insurance. In R. Battiti, M. Brunato, I. Kotsireas, & P. Pardalos (Eds.), <i>Lecture notes in computer science: Vol. 11353. Learning and intelligent optimization</i> (pp. 225-240). Springer. https://doi.org/10.1007/978-3-030-05348-2_21

	Insertion in text	In reference list
Online	... (Gu et al., 2018) ... Or Gu et al. (2018) ...	Gu, X., Yu, J., Han, Y., Han, M., & Wei, L. (2019, July 12-14). <i>Vehicle lane change decision model based on random forest</i> . [Paper presentation]. 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China. https://doi.org/10.1109/ICPICS47731.2019.8942520
Government Publications		
Government as author	First in-text reference: Spell out the full name with the abbreviation of the body. ... National Cancer Institute (2019) ... Or ... (National Cancer Institute, 2019) ... Subsequent in-text reference: ... NCI (2019) ... Or ... (NCI, 2019) ...	National Cancer Institute. (2019). <i>Taking time: Support for people with cancer</i> (NIH Publication No. 18-2059). U.S. Department of Health and Human Services, National Institutes of Health. https://www.cancer.gov/publications/patient-education/takingtime.pdf

8. General Guidelines

Abbreviations: Define alphabetically, other than abbreviations that can be used without definition. Words or phrases that are abbreviated in the *Introduction* and following text should be written out in full the first time that they appear in the text, with each abbreviated form in parenthesis. Include the common name or scientific name, or both, of animal and plant materials.

Authors' Affiliation: The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved to another institution, the current address may also be stated in the footer.

Co-Authors: The commonly accepted guideline for authorship is that one must have substantially contributed to the development of the paper and share accountability for the results. Researchers should decide who will be an author and what order they will be listed depending upon their order of importance to the study. Other contributions should be cited in the manuscript's *Acknowledgements*.

Similarity Index: All articles received must undergo the initial screening for originality before being sent for peer review. *Pertanika* does not accept any article with a similarity index exceeding **20%**.

Copyright Permissions: Authors should seek necessary permissions for quotations, artwork, boxes or tables taken from other publications or other freely available sources on the Internet before submission to *Pertanika*. The *Acknowledgement* must be given to the original source in the illustration legend, in a table footnote, or at the end of the quotation.

Footnotes: Current addresses of authors if different from heading may be inserted here.

Page Numbering: Every page of the manuscript, including the title page, references, and tables should be numbered.

Spelling: The journal uses American or British spelling and authors may follow the latest edition of the Oxford Advanced Learner's Dictionary for British spellings. Each manuscript should follow one type of spelling only.

SUBMISSION OF MANUSCRIPTS

All submissions must be made electronically using the **ScholarOne™ online submission system**, a web-based portal by Clarivate Analytics. For more information, go to our web page and click "**Online Submission (ScholarOne™)**".

Submission Checklist

1. MANUSCRIPT:

Ensure your manuscript has followed the *Pertanika* style particularly the first-4-pages as explained earlier. The article should be written in a good academic style and provide an accurate and succinct description of the contents ensuring that grammar and spelling errors have been corrected before submission. It should also not exceed the suggested length.

2. DECLARATION FORM:

Author has to sign a declaration form. In signing the form, authors declare that the work submitted for publication is original, previously unpublished, and not under consideration for any publication elsewhere.

Author has to agree to pay the publishing fee once the paper is accepted for publication in Pertanika.

3. COVER LETTER:

In Step 6 of the ScholarOne system, author is asked to upload a cover letter in *Pertanika* format. Please ignore this instruction and replace the cover letter with the **Declaration Form**.

Note:

COPYRIGHT FORM: Author will be asked to sign a copyright form when the paper is accepted. In signing the form, it is assumed that authors have obtained permission to use any copyrighted or previously published material. All authors must read and agree to the conditions outlined in the form and must sign the form or agree that the corresponding author can sign on their behalf. Articles cannot be published until a signed form (original pen-to-paper signature) has been received.

Visit our Journal's website for more details at <http://www.pertanika.upm.edu.my/>.

ACCESS TO PUBLISHED MATERIALS

Under the journal's open access initiative, authors can choose to download free material (via PDF link) from any of the journal issues from *Pertanika*'s website. Under "**Browse Journals**" you will see a link, "*Regular Issue*", "*Special Issue*" or "*Archives*". Here you will get access to all current and back-issues from 1978 onwards. No hard copy of journals or offprints are printed.

Visit our Journal's website at:

http://www.pertanika.upm.edu.my/regular_issues.php for "Regular Issue"

http://www.pertanika.upm.edu.my/cspecial_issues.php for "Special Issue"

http://www.pertanika.upm.edu.my/journal_archives.php for "Archives"

PUBLICATION CHARGE

Upon acceptance of a manuscript, a processing fee of RM 750 / USD 250 will be imposed on authors; RM 750 for any corresponding author affiliated to an institution in Malaysia; USD 250 for any corresponding author affiliated to an institution outside Malaysia. Payment must be made online at <https://paygate.upm.edu.my/action.do?do=>

Any queries may be directed to the **Chief Executive Editor's** office via email to executive_editor.pertanika@upm.edu.my

<i>Review Article</i>	
A Review on the Effect of Zein in Scaffold for Bone Tissue Engineering <i>Istikamah Subuki, Khairun Nor Ashikin Nasir and Nur Azrini Ramlee</i>	2805
Robust Hybrid Classification Methods and Applications <i>Friday Zinzendoff Okwonu, Nor Aishah Ahad, Innocent Ejiro Okoloko, Joshua Sarduana Apanapudor, Saadi Ahmad Kamaruddin and Festus Irimisose Arunaye</i>	2831
Significant Factors in Agile Software Development of Effort Estimation <i>Pantjawati Sudarmaningtyas and Rozlina Mohamed</i>	2851
Transformer Population Failure Rate State Distribution, Maintenance Cost and Preventive Frequency Study Based on Markov Model <i>Nor Shafiqin Shariffuddin, Norhafiz Azis, Jasronita Jasni, Mohd Zainal Abidin Ab Kadir, Muhammad Sharil Yahaya and Mohd Aizam Talib</i>	2879
MOFTI: Mobile Self-Service Framework to Uplift Customer Experience for the Telecommunications Industry in Malaysia <i>Zainab Abu Bakar, Hazrina Sofian and Nazean Jomhari</i>	2903
<i>Review Article</i>	
Dehydrated Food Waste for Composting: An Overview <i>Aziz Khalida, Veknesh Arumugam, Luqman Chuah Abdullah, Latifah Abd Manaf and Muhammad Heikal Ismail</i>	2933
Perceived Benefits, Perceived Barriers and Self-Efficacy Towards Weight Reduction Among Overweight and Obese Children in Kedah, Malaysia <i>Wan Mohd Nurussabah Abdul Karim, Hazizi Abu Saad, Nurzalinda Zalbahar and Nurul Husna Mohd Sukri</i>	2961
Evaluation of Acoustical Performance for Atrium Design with Respect to Skylight Geometry and Material in the Tropics <i>Nazli Che Din, Tio Seng Zhi and Chia Keng Soon</i>	2983

ZnO Multilayer Thin Films as the Seed Layer for ZnO Nanorods: Morphology, Structural and Optical properties <i>Rohanieza Abdul Rahman, Muhammad AlHadi Zulkefle, Sukreen Hana Herman and Rosalena Irma Alip</i>	2613
<i>Review Article</i>	
A Review on 3D Augmented Reality Design Technique and Inward Leakage Testing on Protective Face Mask <i>Nur Amirah Kamaluddin, Murizah Kassim and Shahrani Shahbudin</i>	2639
Suppression of Coffee-Ring Effect on Nitrocellulose Membrane: Effect of Polyethylene Glycol <i>Sarah Sorfeena Shahrudin, Norhidayah Ideris and Nur Atikah Kamarulzaman</i>	2659
Multi-Objective Optimal Control of Autocatalytic Esterification Process Using Control Vector Parameterization (CVP) and Hybrid Strategy (HS) <i>Fakhrony Sholahudin Rohman, Dinie Muhammad, Iylia Idris, Muhamad Nazri Murat and Ashraf Azmi</i>	2669
Automated Cryptocurrency Trading Bot Implementing DRL <i>Aisha Peng, Sau Loong Ang and Chia Yean Lim</i>	2683
RNA Isolation from Environmental Samples of a Harmful Algal Bloom for Metatranscriptome Next-Generation Sequencing <i>Diana Lorons, Kenneth Francis Rodrigues, Madihah Jafar Sidik and Grace Joy Wei Lie Chin</i>	2707
A Comparative Study of RNA-Seq Aligners Reveals Novoalign's Default Setting as an Optimal Setting for the Alignment of HeLa RNA-Seq Reads <i>Kristine Sandra Pey Adum and Hasni Arsad</i>	2727
The Kinetic Evaluation and DFT Study of Cis-[Pt(Asc)(NH ₃) ₂] Complex as an Inhibitor to Type 2 Diabetic Human Amylase <i>Khalid Farouk Al-Rawi, Khaldoon Taher Maher, Othman Ibrahim Alajrawy and Firas Taher Maher</i>	2747
Comparative Computational Study of Double Rotating Cylinder Embedded on Selig S1223 Aerofoil and Flat Plate for High Altitude Platform <i>Hidayatullah Mohammad Ali, Azmin Shakrine Mohd Rafie, Mohd Faisal Abdul Hamid and Syaril Azrad Md. Ali</i>	2767
Risk Assessment on Robotic Surgery Using Bayesian Network <i>Teh Raihana Nazirah Roslan and Chee Keong Ch'ng</i>	2789

Evaluation of Phytochemical and Mineral Composition of Malaysia's Purple-Flesh Sweet Potato <i>Rosnah Shamsudin, Nurfarhana Shaari, Mohd Zuhair Mohd Noor, Nazatul Shima Azmi and Norhashila Hashim</i>	2463
Cultivation of <i>Rhodotorula Toruloides</i> Using Palm Oil Mill Effluent: Effect on the Growth, Lipid Production, and Waste Removal <i>Ignatia Justine, Grace Joy Wei Lie Chin, Wilson Thau Lym Yong and Mailin Misson</i>	2477
The Relationship Between Bangladesh's Financial Development, Exchange Rates, and Stock Market Capitalization: An Empirical Study Using the NARDL Model and Machine Learning <i>Rehana Parvin</i>	2493
Organ Dose and Radiation Exposure Risk: A Study Comparing Radiation Dose Using Two Software Packages <i>Abdullah Ali M Asiri</i>	2509
Effective Emergency Management: Scrutinizing the Malaysia Lead Responding Agency Planning and Information Management Approach During Disaster Exercise <i>Khairilmizal Samsudin, Fatin Najihah Ghazali, Nur Hannani Abdul Ghani, Mohamad Fahmi Hussin, Ainul Husna Kamarudin and Haikal Kasri</i>	2521
A Topic Modeling and Sentiment Analysis Model for Detection and Visualization of Themes in Literary Texts <i>Kah Em Chu, Pantea Keikhosrokiani and Moussa Pourya Asl</i>	2535
Data Safety Prediction Using Bird's Eye View and Social Distancing Monitoring for Penang Roads <i>Lek Ming Lim, Majid Khan Majahar Ali, Mohd. Tahir Ismail and Ahmad Sufiril Azlan Mohamed</i>	2563
Effect of Process Conditions on Catalytic Hydrothermal Oxidation of <i>p</i> -Xylene to Terephthalic Acid <i>Mohamad Zarqani Yeop, Kamariah Noor Ismail and Ahmad Rafizan Mohamad Daud</i>	2589
Hole Quality Assessment of Drilled Carbon Fiber Reinforced Polymer (CFRP) Panel Using Various Custom Twist Drill Geometries <i>Muhammad Faris Shauqi Kamaruzaman, Muhammad Hafiz Hassan and Muhammad Fauzinizam Razali</i>	2603

Pertanika Journal of Science & Technology
Vol. 30 (4) Oct. 2022

Content

Foreword <i>Chief Executive Editor</i>	i
Application of Groupwise Principal Sensitivity Components on Unbalanced Panel Data Regression Model for Gross Regional Domestic Product in Kalimantan <i>Desi Yuniarti, Dedi Rosadi and Abdurakhman</i>	2315
Evaluation of Subcritical Organic Rankine Cycle by Pure and Zeotropic of Binary and Ternary Refrigerants <i>Omid Rowshanaie, Mohd Zahirasri Mohd Tohir, Faizal Mustapha, Mohammad Effendy Ya'acob and Hooman Rowshanaie</i>	2333
Optimal Material Selection for Manufacturing Prosthetic Foot <i>Fahad Mohanad Kadhim, Samah Falah Hasan and Sadiq Emad Sadiq</i>	2363
Development of Regression Models for the Prediction of NiTi Archwire Forces in an Orthodontic Bracket System Using Response Surface Methodology <i>Sareh Aiman Hilmi Abu Seman, Ching Wei Ng, Muhammad Fauzinizam Razali and Abdus Samad Mahmud</i>	2377
Robust Linear Discriminant Rule Using Double Trimming Location Estimator with Robust Mahalanobis Squared Distance <i>Yik Siong Pang, Nor Aishah Ahad and Sharipah Soaad Syed Yahaya</i>	2393
Design Guidelines for Vegetative Approach Application to Mitigate Riverbank Landslides in Mangkang Wetan and Mangunharjo Villages, Semarang Coastal Area, Indonesia <i>Santy Paulla Dewi and Rina Kurniati</i>	2407
Quality Evaluation of Color and Texture of the Dabai Fruit (<i>Canarium odontophyllum</i> Miq.) at Different Temperatures and Times of Blanching <i>Rosnah Shamsudin, Siti Hajar Ariffin, Wan Nor Zanariah Zainol @ Abdullah, Nazatul Shima Azmi and Arinah Adila Abdul Halim</i>	2427
<i>Review Article</i>	
Computed Tomography and Other Imaging Modalities in Pediatric Congenital Heart Disease <i>Hongying Chen, Norafida Bahari, Suraini Mohamad Saini and Noramaliza Mohd Noor</i>	2439



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@upm.edu.my
Tel. No.: +603- 9769 1622

PENERBIT
UPM
UNIVERSITI PUTRA MALAYSIA
PRESS

<http://www.penerbit.upm.edu.my>
Email: penerbit@upm.edu.my
Tel. No.: +603- 9769 8851

