IIUM ENGINEERING JOURNAL

Volume 25 Number 2

July 2024

IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

ISSN: 1511-788X E-ISSN: 2289-7860

http://journals.iium.edu.my/ejournal

IIUM ENGINEERING JOURNAL

Volume 25, Issue 2, July 2024

https:/doi.org/10.31436/iiumej.v25i2

Copyright Notice

Consent to publish: The Author(s) agree to publish their articles with IIUM Press.

Declaration: The Author(s) declare that the article has not been published before in any form, that it is not concurrently submitted to another publication, and that it does not infringe on anyone's copyright. The Author(s) holds the IIUM Press and Editors of the journal harmless against all copyright claims.

Transfer of copyright: The Author(s) hereby agree to transfer the article's copyright to IIUM Press, which shall have the non-exclusive and unlimited right to publish the article in any form, including in electronic media. For articles with more than one author, the corresponding author confirms that he/she is authorized by his/her co-author(s) to grant this copyright transfer.

The IIUM Engineering Journal follows the open access policy.

All articles published for open access will be immediately and permanently free for everyone to read, download, copy, and distribute for noncommercial purposes.

IIUM Engineering Journal at https://journals.iium.edu.my/ejournal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

IIUM ENGINEERING JOURNAL EDITORS

CHIEF EDITOR
Ahmad Faris Ismail, IIUM, Malaysia

TECHNICAL EDITOR
Khairul Azami Sidek, IIUM, Malaysia

EXECUTIVE EDITOR
Teddy Surya Gunawan, IIUM, Malaysia

ASSOCIATE EDITOR
Nor Farahidah Za'bah, IIUM, Malaysia

LANGUAGE EDITOR
Lynn Mason, Malaysia

COPY EDITOR
Aliza 'Aini Md Ralib, IIUM, Malaysia
Suriza Ahmad Zabidi, IIUM, Malaysia

MALAY TRANSLATOR
Nurul Arfah Che Mustapha, IIUM, Malaysia

EDITORIAL BOARD MEMBERS

Abdullah Al-Mamun, IIUM, Malaysia
Abdumalik Rakhimov, IIUM, Malaysia
Aishah Najiah Dahnel, IIUM, Malaysia
Alya Naili Rozhan, IIUM, Malaysia
Ani Liza Asnawi, IIUM, Malaysia
Hanafi Ani, IIUM, Malaysia
Hanafy Omar, Saudi Arabia
Huda Adibah Mohd. Ramli, IIUM, Malaysia
Konstantin Khanin, University of Toronto, Canada
Ma'an Al-Khatib, IIUM, Malaysia
Meftah Hrairi, IIUM, Malaysia
Mohamed B. Trabia, United States
Mohammad S. Alam, Texas A&M University-Kingsville, United States
Mohd Sultan Ibrahim Shaik Dawood, IIUM, Malaysia
Mustafizur Rahman, National University Singapore, Singapore
Noorasikin Samat, IIUM, Malaysia
Nor Fadhillah Mohamed Azmin, IIUM, Malaysia
Norsinnira Zainul Azlan, IIUM, Malaysia
Ossama Abdulkhalik, Michigan Technological University, United States
Waqar Asrar, IIUM, Malaysia

AIMS & SCOPE OF IIUMENGINEERING JOURNAL

The **IIUM Engineering Journal**, published biannually (January and July), is a carefully refereed international publication of the International Islamic University Malaysia (IIUM). Contributions of high technical merit within the span of engineering disciplines covering the main areas of engineering: Electrical and Computer Engineering; Mechanical and Manufacturing Engineering; Automation and Mechatronics Engineering; Material and Chemical Engineering; Environmental and Civil Engineering; Biotechnology and Bioengineering; Engineering Mathematics and Physics; and Computer Science and Information Technology are considered for publication in this journal. Contributions from other areas of engineering and applied science are also welcome. The IIUM Engineering Journal publishes contributions under *Regular papers and Invited review papers*. It also welcomes contributions that address solutions to the specific challenges of the developing world, as well as science and technology issues, from an Islamic and multidisciplinary perspective.

REFEREES' NETWORK

All papers submitted to the IIUM Engineering Journal will be subjected to a rigorous reviewing process through a worldwide network of specialized and competent referees. Each accepted paper should have at least two positive referees' assessments.

SUBMISSION OF A MANUSCRIPT

A manuscript should be submitted online to the IIUM-Engineering Journal website at https://journals.iium.edu.my/ejournal. The journal website can also be used for further correspondence on the paper's status.

INTERNATIONAL ADVISORY COMMITTEE

A. Anwar, United States
Abdul Latif Bin Ahmad, Malaysia
Farzad Ismail, USM, Pulau Pinang, Malaysia
Hanafy Omar, Saudi Arabia
Hany Ammar, United States
Idris Mohammed Bugaje, Nigeria
K.B. Ramachandran, India
Kunzu Abdella, Canada
Luis Le Moyne, ISAT, University of Burgundy, France
M Mujtaba, United Kingdom
Mohamed AI-Rubei, Ireland
Mohamed B Trabia, United States
Syed Kamrul Islam, United States
Tibor Czigany, Budapest University of Technology and Economics, Hungary
Yiu-Wing Mai, The University of Sydney, Australia.

Published by:

IIUM Press

IIUM Press,
International Islamic University Malaysia
Jalan Gombak, 53100 Kuala Lumpur, Malaysia
Phone (+603) 6421-5014, Fax: (+603) 6421-6298

Whilst every effort is made by the publisher and editorial board to see that no inaccurate or misleading data, opinion or statement appears in this Journal, they wish to make it clear that the data and opinions appearing in the articles and advertisement herein are the responsibility of the contributor or advertiser concerned. Accordingly, the publisher and the editorial committee accept no liability whatsoever for the consequence of any such inaccurate or misleading data, opinion, or statement.

ISSN 1511 - 788X

9 771511 788008

IIUM Engineering Journal
ISSN: 1511-788X E-ISSN: 2289-7860

IIUM ENGINEERING JOURNAL

Volume 25, Issue 2, July 2024
https:/doi.org/10.31436/iiumej.v25i2

Table of Contents

COPYRIGHT ..... **I**

EDITORIAL ..... **II**

CHEMICAL AND BIOTECHNOLOGY ENGINEERING

2785: BIOMARKERS IDENTIFICATION FROM URINE OF DENGUE PATIENTS THROUGH GCMS FOR INITIAL DEVELOPMENT OF NON-INVASIVE DIAGNOSTIC KIT 1
Azura Amid, Muhamad Shirwan, Aszrin, Norbaiyah, Fatein, Nurul Aulia Zakaria, Nor Zamzila Abdullah, Nik Nur Fatnoon Nik Ahmad

2933: RECENT ADVANCES IN ENHANCED POLYAMIDOAMINE INHIBITORS FOR SILICATE SCALES IN THE PETROLEUM UPSTREAM... 17
Tan Bee Chea, Ismail Mohd Saaid, Shiferaw Regassa Jufar, Suzalina Zainal, Abdelazim Abbas Ahmed Awadelseed

2957: BIOHYDROGEN PRODUCTION IN SEMI-CONTINUOUS SYSTEM USING IMMOBILIZED CELL MEMBRANE 32
Nurul Sakinah Engliman; Jamaliah Md Jahim, Ainul Husna Abdul Aziz, Peer Mohamed Abdul, Abdullah Amru Indera Luthfi, Nur Syakina Jamali

2974: AN EASY-TO-FABRICATE, SUBMERGED CARBON-BASED AIR CATHODE FOR BIOFUEL CELLS 46
Asiah Sukri, Raihan Othman, Norshahida Sarifuddin

2978: EVALUATION AND KINETICS OF TOFU WASTEWATER BIOREACTOR WITH ADDITION OF WATER HYACINTH..... 57
Herawati Budiastuti, Laily Isna Ramadhani, Pratap Pullammanappallil, Ranggi Octaviani Pratiwi, Yuliana Nur Amanah

3090: FUNCTIONAL METAGENOMICS APPROACH FOR DISCOVERY OF NOVEL COLD-ACTIVE PROTEASE FROM THE ANTARCTIC REGION 69
Muhammad Asyraf Abd Latip, Noor Faizul Hadry Nordin, Siti Aisyah Alias, Jerzy Smykla, Faridah Yusof, Mohd Azrul Naim Mohamad

3098: CONTINUOUS FLOW BEAD-MILLING IMPACT ON SULFUR CURING FOR ADVANCED ELASTOMERIC RUBBER COMPOSITES... 83
Mohamad Firdaus Omar, Fathilah Ali, Mohammed Saedi Jami, Azlin Suhaida Azmi, Farah Ahmad

3240: EFFECT OF INCORPORATING CELLULOSE NANOFIBERS AND LEMONGRASS ESSENTIAL OIL IN POLYVINYL ALCOHOL-POLYETHYLENE GLYCOL/GLYCERIN HYDROGEL FOR WOUND DRESSING 99
Nur Huda Syazwani Jafri, Dzun Noraini Jimat, Wan Mohd Fazli Wan Nawawi, Yusilawati Ahmad Nor, Azura Amid

CIVIL AND ENVIRONMENTAL ENGINEERING

3030: MONITORING CHANGE DETECTION OF VEGETATION VULNERABILITY USING HOTSPOTS ANALYSIS 116
Basheer Jasim, Oday Z. Jasim, Amjed N. AL-Hameedawi

3040: RESISTANCE OF RUBBERIZED ASPHALT MIXTURE TO AGING AND MOISTURE DAMAGE..... 130
Ayman Hassan Al-Qudah, Suhana Koting, Mohd Rasdan Ibrahim, Muna Alibrahim, Nishanthini Jegatheesan

3211: GEOMETRIC QUALITY ANALYSIS OF TERRESTRIAL LASER SCANNING DATA FOR INDUSTRIAL USAGE..... 148
Ali Salah J. Al-Saedi, Fanar M. Abed, Imazhim A. Alwan

ELECTRICAL, COMPUTER AND COMMUNICATIONS ENGINEERING

2769: CLASSIFICATION OF C. ANNUUM AND C. FRUTESCENS RIPENING STAGES: HOW WELL DOES DEEP LEARNING PERFORM? 167
Marsyita Hanafi, Siti Mariam Shafie, Zarina Ibrahim

2954: ENHANCED ANTENNA PERFORMANCE AT 3.5 GHZ WITH A COMPACT AND INTELLIGENT REFLECTING SURFACE 179
Firdaus Taqiyuddin Ahamad, Norun Farihah Abdul Malek, Farah Sakiinah Roslan, Norazlina Saidin, Md. Rafiqul Islam, Nidal El Qasem

2963: BRAIN TUMOR SEGMENTATION AND CLASSIFICATION USING CNN PRE-TRAINED VGG-16 MODEL IN MRI IMAGES 196
Gayathri T., Sundeep Kumar K

3044: MEMETIC ALLIGATOR OPTIMIZATION ALGORITHM FOR OPTIMAL THERMOREGULATORY CONTROL IN PIPING SYSTEMS .. 212
Weng-Hooi Tan, Ts. Dr. Noramalina Abdullah, Junita Mohamad-Saleh

2063: OPERATIONAL COST MODELS FOR AN EARTH STATION SYSTEM USING 2-PARALLEL AND 4-PARALLEL CONFIGURATIONS 232

Wan Muhammad Syaamil Wan Aris, Nadirah Abdul Rahim, Mahayaudin M. Mansor, Nur Shazana Abdul Rahman

3119: ATTITUDE UAV STABILITY CONTROL USING LINEAR QUADRATIC REGULATOR-NEURAL NETWORK (LQR-NN).....246
Oktaf Agni Dhewa, Fatchul Arifin, Ardy Seto Priyambodo, Anggun Winursito, Yasir Mohd. Mustafa

3129: BUILDING A GIS MAP FOR FORECASTING THE MIR INDEX IN AN GIANG.....266
Thanh-Nghi Doan, Khanh Tran Thien Nguyen

3279: AREA-BASED RAINFALL RATE MODEL FOR SPECIFIC ATTENUATION IN THE EQUATORIAL REGION.....287
Yasser Asrul Ahmad, Ahmad Fadzil Ismail, Muhammad Nasrin Aqil Abdul Hamid, Mohd Faizal Jamlos

MATERIALS AND MANUFACTURING ENGINEERING

3058: PARTIAL REPLACEMENT OF SILICA NANOPARTICLES IN CEMENT PASTE FOR CO2 REGENERATION CAPTURE APPLICATION 299
Mudrikah Sofia Mahmud, Farah Diana Mohd Daud, Norshahida Sarifuddin, Hafizah Hanim Mohd Zaki, Norhuda Hidayah Nordin, Dr. Nur Farahiyah Mohammad

MECHATRONICS AND AUTOMATION ENGINEERING

2964: A STUDY ON THE TERRESTRIAL MOBILITY OF A SPHERICAL AMPHIBIAN ROBOT 309
Mohd Bazli Bahar, Shahrum Shah Abdullah, Mohd Shahrieel Mohd Aras; Mohamad Haniff Harun, Muhamad Khairi Aripin, Fariz Ali@Ibrahim

3017: EXPLORING THE EFFICACY OF PIEZOELECTRIC-BASED SENSORY SYSTEMS FOR HEART RATE MONITORING IN DIFFERENTIATING STRESS VS. RELAX CONDITIONS.....325
Muhammad Hafiz Zulhilmi Zulkifli, Farahiyah Jasni , Aimi Shazwani Ghazali, Nor Hidayati Diyana Nordin, Shahrul Na'im Sidek, Syamimi Mardiah Shaharum, Fadhlan Hafizhelmi Kamaru Zaman

3029: ENHANCING PROSTHETIC CONTROL: NEURAL NETWORK CLASSIFICATION OF THUMB MUSCLE CONTRACTION USING HD-sEMG SIGNALS 338
Muhammad Mukhlis Suhaimi, Aimi Shazwani Ghazali, Ahmad Jazlan Haja Mohideen, Muhammad Hariz Hafizalshah, Shahrul Na'im Sidek

3094: PERFORMANCE ANALYSIS OF AN INDUCTION MOTOR COUPLED VIF WITH MR FLUID DAMPER.....350
Syed Munimus Salam, Muhammad Mahbubur Rashid

3188: BEE ALGORITHM BASED CONTROL DESIGN FOR TWO-LINKS ROBOT ARM SYSTEMS 367
Russul A. Kadhim, Mina Q. Kadhim, Huthaifa Al-Khazraji, Amjad J. Humaidi

3207: INTEGRATION OF MFCCS AND CNN FOR MULTI-CLASS STRESS SPEECH CLASSIFICATION ON UNSCRIPTED DATASET 381
Nur Aishah Zainal, Ani Liza Asnawi, Ahmad Zamani Jusoh, Siti Noorjannah Ibrahim, Huda Adibah Mohd. Ramli

3250: DESIGN AND DEVELOPMENT OF A SLIDER-CRANK ACTUATED KNEE EXOSKELETON WITH OPTIMIZED MOTION CONTROLLER 396
Mariam MD Ghazaly, Nai Jun An, Law Hin Kwee, Zulkeflee Abdullah, Norhaslinda Hasim, Isa Halim, Nasharuddin Zainal

https://doi.org/10.31436/iiumej.v25i2.2785

BIOMARKER IDENTIFICATION FROM URINE OF DENGUE PATIENTS THROUGH GCMS FOR INITIAL DEVELOPMENT OF NON-INVASIVE DIAGNOSTIC KIT

AZURA AMID[1*], MUHAMAD SHIRWAN ABDULLAH SANI[1], ASZRIN ABDULLAH[2], NORBAIYAH MOHAMED BAKRIM[2], WAN FATEIN NABEILA WAN OMAR[2], NURUL AULIA ZAKARIA[3], NOR ZAMZILA ABDULLAH[4], NIK NUR FATNOON NIK AHMAD[3]

[1]International Institute for Halal Research and Training (INHART), International Islamic University Malaysia, Gombak, Malaysia
[2]Department of Basic Medical Sciences, Kulliyyah of Medicines, International Islamic University Malaysia, 25200 Kuantan, Pahang.
[3]Department of Internal Medicine, Kulliyyah of Medicines International Islamic University Malaysia, 25200 Kuantan, Pahang
[4]Department of Pathology and Laboratory Medicine, Kulliyyah of Medicines, International Islamic University Malaysia, 25200 Kuantan, Pahang

**Corresponding author: azuraamid@iium.edu.my*

(Received: 8 March 2023; Accepted: 20 February 2024; Published online: 15 July 2024)

ABSTRACT: Dengue infection detection methods, namely dengue serology test and Real-Time PCR, are only available in clinical laboratories or healthcare facilities. This is time-consuming and inconvenient for patients. Thus, a non-invasive on-site urinary diagnostic kit for dengue infection that requires no trained personnel for blood extraction would be advantageous and warranted, albeit in developed or underdeveloped nations with limited resources. To develop the on-site diagnostic kit, an identification of biomarkers related to dengue infection is needed. This preliminary study aims to identify potential metabolite biomarkers from the urine of dengue patients using GCMS analysis. Urine samples of patients with serologically confirmed dengue infection were analyzed and compared with healthy volunteers. The study shows that the GCMS approach can identify differences in the urine of dengue patients from healthy volunteers. Heptacosane, Hexadecane, 2,4-bis(1,1-dimethylethyl) phenol, 2-bromooctane, tetradecane, hexyl octyl ester sulfurous acid, 2-benzoyl methyl ester benzoic acid, 2,9-dimethyl decane, and pentadecane were identified from the urine of dengue patients. The identified secretion of alkane may be a suitable candidate for colorimetric assay for the development of a user-friendly, home-screening rapid test kit for the detection of dengue infection.

ABSTRAK: Kaedah pengesanan jangkitan denggi seperti ujian serologi denggi dan PCR masa nyata, merupakan kaedah biasa yang digunakan di makmal klinikal atau pusat kesihatan. Ianya memakan masa dan menyukarkan pesakit. Oleh itu, kit diagnostik air kencing yang tidak invasif yang tidak memerlukan kakitangan terlatih bagi mengekstrak darah adalah sangat berguna dan diperlukan bagi mengesan jangkitan denggi, sama ada di negara maju atau tidak membangun dengan sumber terhad. Bagi membangunkan kit diagnostik ini, pengesanan penanda bio berkaitan jangkitan denggi diperlukan. Kajian awal ini bertujuan bagi mengenal pasti penanda bio metabolit yang berkaitan pada air kencing pesakit denggi dengan menggunakan analisis GC/MS. Sampel air kencing pesakit yang disahkan secara serologi telah dikaji dan dibandingkan dengan sukarelawan yang sihat. Kajian menunjukkan teknik GCMS dapat mengenal pasti perbezaan antara air kencing pesakit denggi dan sukarelawan yang sihat. Heptakosane, Hekzadekane, fenol 2,4-bis(1,1-dimetiletil), 2-bromookten, tetradekana, asid sulfurik ester heksil oktil, asid benzoik metil ester 2-benzoik, 2,9-dimetil

dekana, dan pentadekana telah dikenal pasti terkumpul dalam air kencing pesakit denggi. Rembesan alkana merupakan penanda bio yang sesuai bagi ujian kolorimetrik dalam membangunkan kit ujian pantas diagnostik yang mesra pengguna bagi mengesan denggi di rumah.

KEYWORDS: Biomarkers, metabolites, GCMS, dengue, detection method.

1. INTRODUCTION

Dengue is the most prevalent arboviral infection globally, affecting 400 million people worldwide every year [1]. It is caused by the dengue viruses (DENVs) that are transmitted by its' principal vector, the *Aedes aegypti* mosquitoes [2]. It is endemic in tropical and subtropical areas where the *Aedes sp*. is highly infested. Of note, Asian countries score the highest dengue disease burden internationally [2-4]. It was reported that over the past 50 years, the incidence of dengue has increased by 300% globally [3,4]. This trend is mainly attributed to global warming, population growth, urbanization, globalization, and enhanced human mobility, as well as inadequate mosquito control measures [1-3]. In Malaysia, dengue incidence was reported to be between 69.9 and 93.4 per 1000 people [4].

1.1. Pathophysiology

Dengue infection has a heterogeneous manifestation ranging from asymptomatic to mild, self-limiting febrile illness to severe and life-threatening systemic disease, depending on the viral-host interplay [2,3]. The three stages of dengue include the incubation phase, the febrile phase, and the critical phase [2]. DENV is a member of the Flaviviridae family with four well-described serotypes of dengue viruses: DENV1-4 [1-3]. DENV is a single-stranded RNA virus that encodes for the capsid, pre-membrane, envelope, and seven non-structural proteins, all of which contribute to its virulence and were the target of vaccine development [1, 5]. Infection by one serotype provides long-term immunity towards the specific serotype and some degree of cross-protection against the other serotypes [1]. Surprisingly, this cross-protection was reported to increase the risk of severe dengue in secondary infection by other serotypes [1,3,6]. This antibody-dependent enhancement phenomenon has proven to be detrimental as it augments viral entry into the immune cells, escapes the immune responses, and dysregulates the pro-and anti-inflammatory balance[1-2]. Eventually, this leads to dengue shock syndrome, which could be fatal.

1.2. Dengue Infection Method of Detection

At present, the dengue serology test is the gold standard for confirming the diagnosis by detecting anti-DENV IgM and IgG [7]. Other methods, such as virus isolation and molecular approach using reverse transcriptase-polymerase chain reaction (RT-PCR), have also been used in detecting the presence of the dengue virus in a blood sample [8]. However, all these methods are laboratory-dependent, require blood extraction, and are time-consuming, which may cause a delay in diagnosis. Moreover, the requirement for blood extraction procedures may not be the preference of some patients. Hence, these methods are more practical for research purposes and as diagnostic tools in healthcare facilities. Moving forward, several rapid diagnostic tests have been established to help physicians make quick diagnoses [9]. Although the rapid diagnostic test kit may not be as accurate as the standard test, the kit remains useful for initial patient screening, particularly in limited-resource hospital settings and remote areas [10]. Nonetheless, blood extraction is still required for these rapid diagnostic test kits.

Alternative samples, such as urine and saliva, have been tested [11]. In one study, dengue RNA was detected in serum, urine, and saliva samples using real-time RT-PCR for DENV

detection and quantification [11]. Another study analyzed a larger sample size of urine samples and concluded that RT-PCR and NS1 tests demonstrated an overall sensitivity and specificity of 41.6%/ 14.5% in urine, as compared to 85.4%/ 63.4% in plasma [12]. A year later, the same authors evaluated the use of urine samples in several commercialized rapid diagnostic kits and demonstrated promising data [13]. However, the authors recommended that the sensitivity of the kit be further improved for urine samples before utilizing it in clinical settings. Therefore, the use of urine as the sample for the initial screening method before final detection is promising.

1.3. Dengue Infection Diagnostic Challenges

To date, there is no specific treatment for dengue infection; management is merely supportive [7]. Debilitating complications of dengue infection stresses the importance of early detection and management to reduce morbidity and mortality rates. Diagnostic delay in dengue infection can be related to physician and patient factors. Physicians or travelers from non-endemic regions may have low suspicion of dengue infection, contributing to the delay. Patients' delay in seeking treatment may also contribute to the development of disease complications. Among the factors associated with delayed care-seeking is poverty, which may affect one's behavior despite the available resources [14]. Interestingly, one study in Thailand found that being female is significantly associated with morbidity among dengue patients [15]. These associated factors may imply that the inconvenience of performing the diagnostic test, which is only available in a bigger healthcare facility, causes patients to put off seeking treatment until more serious symptoms develop.

Presently, gold-standard diagnostic tests for dengue involve invasive blood extraction from suspected patients [7–9]. Blood-collecting procedures could be difficult in some dengue patients, especially among those presenting with hemorrhagic shock, children, and newborns. Furthermore, patients with typanophobia or extreme fear of medical procedures involving needles could delay seeking treatment [15]. Hence, an alternative for a non-invasive method using other than blood samples, which require minimal skills to perform, is very much needed. A diagnostic kit using salivary or urinary samples helps to remove the hindrance associated with typanophobia apart from providing options.

Therefore, this study aims to explore the potential of developing a non-invasive urinary diagnostic kit for dengue infection that would be advantageous and warranted, albeit in more developed or underdeveloped nations with limited resources. Together with continuing education to raise awareness of its early symptoms, a sensitive, feasible, and affordable diagnostic tool will enable early initiation of supportive treatment and may reduce morbidity and mortality among dengue patients

2. MATERIALS AND METHOD

All chemicals and reagents purchased are analytical grades.

2.1. Sample collection

Ten urine samples were obtained from patients with serologically confirmed dengue infection who were admitted to IIUM Sultan Ahmad Shah Medical Centre, Kuantan. Patients with a history of chronic diseases or pregnant women were excluded. Forty milliliters of urine samples were collected at 8 a.m. in all consented patients aged 18 - 60 years. An additional 3 urine samples from healthy volunteers were also obtained. The urine was centrifuged for 10 minutes at 3000 revolutions per minute within 2 hours of collection. The supernatant was aliquoted and stored in a -80°C freezer to prevent metabolite changes before analysis. Ethical

approval was obtained from the International Islamic University Malaysia Research Ethics Committee (IREC 2019-187).

The urine samples were registered as C1, C2, C3, D1, D2, D3, D4, D5, D6, D7, D8, D9, and D10 then subjected to freeze-drying activity including main-drying under 0.011 mbar at -60ºC for 48 h and final-drying under 0.011 mbar at -76◦C for another 48 h.

2.2. Sample preparation

An amount of 0.01 g freeze-dried urine samples was derivatized with 0.5 mL of N,O-bis-trimethylsilyl (BSTFA)-trifluoroacetamide-trimethylchlo-rosilane (TMCS) mixture at 99:1 ratio (Macherey–Nagel, Germany) and 0.5 mL hexane to give a 0.01 g/mL urine mixture. The mixture was then heated at 60°C for 30 min before injection of 1 µL urine mixture into the Pegasus High Throughput Gas Chromatography Time-of-Flight Mass Spectrometer (GC-TOF/MS) [16].

2.3. GC/MS Analysis

A volume of 1 µL urine mixture was injected into the 225◦C GC-TOF/MS (Leco, Netherlands) inlet at a 1:10 splitless mode ratio. The urine mixture was evaporated and carried at 1.2 mL/min rate by He gas to the oven at temperature programming: 80◦C for 2 min; raised to 240◦C at 5◦C/min for 5 min, and raised to 300◦C at a rate of 3 ◦C/min for 5 min. The evaporated metabolites of the urine mixture were separated by an HP-5 MS capillary column (30 m × 0.25 mm internal diameter and 0.25 µm film thickness), eluted to 300◦C transfer line of the mass spectrometer (MS), and detected by the MS set at 70 eV electron energy and a mass range of 50–550 m/z. The detected metabolites were compared to the National Institute of Standard (NIST) Library 2017, where compounds with similarities of more than 80% were chosen for multivariate data analysis [17]. At least five replicates of the GCMS analysis were carried out for each sample. The composition of each urine was reported in the form of percentage composition.

2.4. Dataset pre-processing

The percentage composition of each urine sample was imported to the dataset table in XLSTAT 2019 software (Addinsoft, Paris, France). About 65 urines x 46 metabolites in the dataset were pre-processed to facilitate the process of differentiation among the urines. This pre-processing step is applied to reduce the dataset's variation of metabolites. Dataset pre-processing tests consisted of dataset transformation using XLSTAT-Pro (2019) statistical software.

To ensure the dataset follows normal distribution before the PCA, the dataset normality was tested using Lilliefors at α of 0.05. The dataset was transformed using standardize (n-1), standardize (n), center, standard deviation−1 (n-1), standard deviation−1 (n), rescale from 0 to 1, rescale from 0 to 100, and Pareto methods. The transformation of each metabolite was employed to ensure the transformed dataset remained closer to the original dataset. The normal distribution of the transformed dataset was evaluated by the normality test of Lilliefors at α of 0.05 [18].

2.5. Principal component analysis

The PCA was applied to investigate the metabolite contribution to the urines and identify the variance of intercorrelated metabolites at α of 0.05. These objectives were achieved by transforming the dataset into principal components (PCs), which are new sets of independent variables that are displayed in the form of correlation biplot and metabolite plot.

In this study, cumulative variability (CV) of two-dimensional PCs entailing PC1 and PC2 was computed for the metabolite exploratory. The number of PCs selected was based on the highest cumulative variability. The metabolites with FL ≥ |0.750|, |0.500| < FL < |0.749| and FL ≤ |0.499| were considered metabolites with strong, moderate, and weak factor loading, respectively. To determine the contribution of metabolites to the urines, the metabolite and correlation biplots were plotted, and the metabolites that contributed to each urine were examined and further explained. The metabolites with weak FL were removed from the metabolite dataset. Using 21 metabolites with strong and moderate FLs, a new PCA was executed to produce a new metabolite plot and correlation biplot. The FL and correlations of the metabolites were assessed, and the metabolites' apportionment to the urine was examined [18].

2.6. Discriminant analysis

In this study, discriminant analysis (DA) was employed to develop a discriminating model (DM) for the urines. Only the urines of dengue-infected patients were present at F1 > 0, and control urine was present at F1 < 0, which brought the dataset to 26 urines x 21 metabolites as both training and cross-validation datasets. An assessment of the classification of urine in the training dataset was performed. The correct classification of the dengue-infected patients and controls was assessed, and their dissimilarity was explored. Classification at < 100% for the training dataset indicated potential urines and metabolites that may render incorrect discrimination between dengue-infected and controls; hence, these urines and metabolites were removed, and discriminant analysis was repeated. After obtaining 100% correct classification in training datasets, the developed DM was cross-validated using the same dataset to validate the high predictive ability [19]. Then, the significant metabolites that contributed to the DM were identified via a unidimensional test of means equality of the dengue-infected patients and controls.

3. RESULTS AND DISCUSSION

3.1. Chemical composition of urine from dengue-infected patients

The chemical composition of urine from dengue-infected patients is listed in Table 1.

3.2 Metabolite correlations

The Pearson correlation test was chosen to select suitable correlation types to determine correlations between two metabolites. The correlation explores the dataset in investigating the strength (weak, moderate, or strong) and direction (positive and negative) of the linear relationship between two metabolites. Table 2 shows the correlation matrix (CM) of the metabolites in the urine of dengue-infected patients. The CM measures the strength and direction of a linear relationship between two metabolites. The CM indicated that positive and negative correlations among the metabolites at |0.700| < R < |1.000| for strong CM, |0.300| < R < |0.700| for moderate CM, and 3|0.000| < R < |0.300| for weak CM as in generic guideline.

Table 1. Chemical composition of urine from dengue-infected patients

Metabolite	Retention time, s	Composition of metabolites in the urine of dengue-infected patient[1], %												
		C1	C2	C3	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Allyl octyl ester oxalic acid	494.603	nd	nd	nd	0.16	nd	nd	nd	nd	nd	1.83	nd	nd	0.70
2,4,4-trimethyl hexane	495.002	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
3,3-dimethyl hexane	507.319	nd	nd	1.31	nd	nd	nd	1.38	5.72	nd	nd	nd	0.48	1.12
2,3,3-trimethyl pentane	530.824	nd	nd	1.25	nd	nd	nd	0.20	nd	0.63	nd	nd	0.39	nd
2,9-dimethyl decane	542.942	nd	nd	nd	nd	0.85	nd	0.37	1.22	nd	nd	nd	0.26	0.41
Dodecamethyl cyclohexasiloxane	607.052	nd	5.12	2.86	9.55	1.95	8.63	0.64	12.53	2.66	0.44	nd	2.23	nd
2-methyl dodecane	792.872	nd	nd	1.16	nd	nd	nd	nd	nd	nd	nd	nd	0.51	0.15
Tetradecane	792.904	nd	nd	nd	0.64	nd	nd	1.05	nd	5.21	0.88	nd	2.37	1.89
Hexyl pentyl ester sulfurous acid	793.138	nd	nd	1.68	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
1-iodo tetradecane	801.936	nd	nd	nd	nd	nd	nd	nd	1.56	3.60	0.16	nd	1.01	0.28
2,3,5,8-tetramethyl decane	832.878	nd	nd	7.97	nd	4.17	nd	4.66	nd	3.93	9.59	nd	11.15	19.23
2,2,5-trimethyl 3,4-hexanedione	840.846	nd	nd	nd	nd	4.62	nd	nd	0.31	nd	nd	nd	nd	0.20
Hexadecane	840.946	26.00	2.43	0.98	5.08	1.69	6.96	10.12	2.09	6.30	23.06	nd	14.53	11.94
Nonadecane	840.960	6.37	41.26	0.67	4.98	25.92	14.71	13.38	21.71	3.64	2.70	nd	1.71	0.50
4-methyl-1-undecene	840.979	3.13	nd	4.49	22.11	0.34	nd	2.30	3.58	5.82	1.48	nd	3.73	3.69
2,6-dimethyl heptadecane	841.543	2nd	nd	4.63	nd	4.08	nd	1.93	nd	17.28	nd	nd	2.25	0.55
3-isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane	855.985	nd	nd	2.26	9.25	2.42	nd	0.84	1.38	nd	nd	nd	nd	nd
2,4-bis(1,1-dimethylethyl) phenol	864.617	3.89	6.48	30.04	7.06	7.04	3.61	4.62	nd	1.74	2.04	nd	1.22	9.33
Allyl nonyl ester oxalic acid	903.328	0.35	nd	nd	0.22	0.35	4.22	nd	nd	2.10	0.27	nd	3.57	3.44
2,6,11-trimethyl dodecane	903.395	2.41	nd	nd	nd	11.56	nd	nd	nd	1.70	nd	nd	11.06	0.16
1-iodo-2-methylnonane	903.410	nd	8.15	0.73	5.91	2.93	7.76	2.81	0.44	1.95	1.67	nd	1.90	0.58
Tridecane	903.461	3.72	nd	1.57	nd	nd	nd	nd	nd	nd	0.02	nd	1.18	0.12
2-bromooctane	983.540	nd	nd	0.92	nd	nd	nd	0.49	nd	0.28	nd	nd	0.83	0.66
(1,1-dimethylethyl)thio acetic acid	1030.880	nd	nd	3.25	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
2,4-dimethyl undecane	1031.210	nd	1.40	nd	nd	nd	nd	nd	nd	2.11	nd	nd	0.70	0.64
Octyl propyl ester oxalic acid	1098.540	nd	0.64	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	1.52
Hexyl octyl ester sulfurous acid	1128.160	nd	nd	nd	1.67	nd	nd	1.56	0.69	1.31	nd	nd	0.45	0.08
Eicosane	1128.230	5.33	nd	2.52	3.32	6.23	7.52	3.24	4.27	4.77	15.36	nd	7.45	3.91
4,7-dimethyl undecane	1128.280	nd	nd	1.97	5.63	0.38	nd	1.55	4.02	nd	nd	nd	0.35	0.15
Undecane	1128.320	5.04	nd	nd	nd	nd	nd	nd	nd	nd	0.51	nd	0.77	0.23
Dotriacontane	1128.360	nd	nd	10.43	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
1-iodo-2-methylundecane	1128.390	4.78	10.06	9.78	11.68	5.01	35.87	1.18	11.51	1.80	3.01	nd	6.37	12.06
2,6,10,14-tetramethyl heptadecane	1138.920	nd	nd	nd	0.27	1.10	nd	0.52	nd	1.32	0.44	nd	0.34	0.46
Pentadecane	1183.540	3.08	5.82	6.12	4.31	7.70	8.88	16.65	7.79	3.48	10.80	nd	7.53	11.02

Metabolite	Retention time, s	Composition of metabolites in the urine of dengue-infected patient[1], %												
		C1	C2	C3	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
2,4,6-trimethyl octane	1183.700	5.33	1.72	nd	0.82	1.59	nd	3.56	nd	nd	3.08	nd	1.22	2.81
Heptadecane	1183.870	nd	6.25	nd	nd	nd	nd	2.84	9.22	5.76	1.94	nd	2.71	0.55
N,N,O-triacetylhydroxylamine	1195.060	nd	nd	nd	0.35	nd	nd	nd	nd	nd	nd	nd	nd	nd
Hexacosane	1388.580	nd	2.08	nd	nd	nd	nd	1.66	nd	3.92	nd	nd	1.26	nd
2-ethylhexyl isohexyl ester sulfurous acid	1388.610	2.70	nd	nd	nd	nd	nd	3.51	0.48	4.16	1.61	nd	0.15	0.45
1-iodoundecane	1388.630	nd	nd	0.55	0.10	0.85	1.85	nd	nd	0.17	nd	nd	0.73	0.13
2,2,3,3,5,6,6-heptamethyl heptane	1388.650	nd	1.78	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
2-ethylhexyl hexyl ester sulfurous acid	1388.680	nd	nd	2.42	nd	0.84	nd	nd	nd	2.36	nd	nd	nd	nd
Hexatriacontane	1388.750	nd	3.54	nd	nd	nd	nd	0.21	nd	nd	nd	nd	nd	nd
2-methyl heptadecane	1388.940	1.34	nd	0.45	5.89	4.96	nd	10.46	6.15	nd	nd	nd	5.70	6.15
2-benzoyl methyl ester benzoic acid	1392.830	nd	nd	nd	1.02	2.75	nd	2.66	3.35	3.76	8.33	nd	1.25	3.14
Heptacosane	1437.420	6.53	3.26	nd	nd	0.65	nd	5.59	1.98	8.23	10.77	nd	2.66	1.74

Note: [1]nd = not detected

https://doi.org/10.31436/iiumej.v25i2.2785

Table 2. Correlation matrix of metabolites in the urine of dengue-infected patients

Metabolites[1]	Correlation matrix, CM[2,3]																				
	M3	M4	M5	M6	M8	M9	M11	M13	M14	M15	M17	M22	M28	M31	M32	M35	M39	M40	M41	M43	M45	M46
M3	1.000	0.540	0.412	0.242	0.637	0.582	0.402	**0.707**	0.603	0.302	**0.769**	0.344	0.236	0.284	**0.785**	0.511	0.459	**0.753**	0.220	**0.775**	0.542	0.222
M4	0.540	1.000	**0.723**	-0.049	0.627	0.199	0.220	0.410	0.190	0.304	0.575	0.255	-0.042	0.296	**0.783**	0.700	0.552	**0.749**	0.664	0.570	**0.906**	0.374
M5	0.412	**0.723**	1.000	0.070	0.585	0.192	0.286	0.495	0.104	0.570	0.533	0.473	-0.006	0.137	0.514	**0.748**	0.584	0.360	**0.871**	0.149	0.556	0.218
M6	0.242	-0.049	0.070	1.000	0.003	0.109	-0.061	0.140	0.013	-0.040	0.001	**0.737**	**0.742**	-0.046	0.038	0.005	-0.043	-0.019	0.133	0.125	-0.062	-0.040
M8	0.637	0.627	0.585	0.003	1.000	**0.809**	0.175	0.655	**0.713**	0.180	0.555	0.195	-0.019	0.343	**0.858**	**0.768**	**0.771**	**0.741**	0.561	0.433	0.452	0.226
M9	0.582	0.199	0.192	0.109	**0.809**	1.000	0.263	0.661	**0.928**	0.103	0.402	0.137	0.097	0.499	**0.701**	0.534	0.661	0.576	0.110	0.367	0.110	0.295
M11	0.402	0.220	0.286	-0.061	0.175	0.263	1.000	**0.759**	0.285	**0.860**	0.488	-0.048	-0.053	0.373	0.339	0.103	0.201	0.329	-0.052	0.486	0.317	0.313
M13	**0.707**	0.410	0.495	0.140	0.655	0.661	**0.759**	1.000	0.555	**0.705**	0.659	0.177	0.100	0.295	0.672	0.437	0.457	0.619	0.237	0.625	0.331	0.175
M14	0.603	0.190	0.104	0.013	**0.713**	**0.928**	0.285	0.555	1.000	0.072	0.425	0.051	0.035	0.688	0.688	0.503	**0.718**	0.539	0.022	0.353	0.197	0.481
M15	0.302	0.304	0.570	-0.040	0.180	0.103	**0.860**	**0.705**	0.072	1.000	0.393	0.110	-0.001	0.229	0.233	0.229	0.239	0.144	0.247	0.271	0.257	0.258
M17	**0.769**	0.575	0.533	0.001	0.555	0.402	0.488	0.659	0.425	0.393	1.000	0.228	-0.023	0.266	0.692	0.473	0.433	0.663	0.330	0.612	0.584	0.172
M22	0.344	0.255	0.473	**0.737**	0.195	0.137	-0.048	0.177	0.051	0.110	0.228	1.000	0.516	0.053	0.271	0.297	0.194	0.116	0.473	0.135	0.174	0.167
M28	0.236	-0.042	-0.006	**0.742**	-0.019	0.097	-0.053	0.100	0.035	-0.001	-0.023	0.516	1.000	-0.048	-0.012	0.031	-0.045	-0.017	-0.045	0.135	-0.040	-0.058
M31	0.284	0.296	0.137	-0.046	0.343	0.499	0.373	0.295	0.688	0.229	0.266	0.053	-0.048	1.000	0.494	0.370	0.675	0.282	0.086	0.167	0.376	**0.888**
M32	**0.785**	**0.783**	0.514	0.038	**0.858**	**0.701**	0.339	0.672	0.688	0.233	0.692	0.271	-0.012	0.494	1.000	0.680	0.677	**0.916**	0.413	**0.739**	**0.715**	0.415
M35	0.511	0.700	**0.748**	0.005	**0.768**	0.534	0.103	0.437	0.503	0.229	0.473	0.297	0.031	0.370	0.680	1.000	**0.779**	0.542	0.692	0.203	0.561	0.408
M39	0.459	0.552	0.584	-0.043	**0.771**	0.661	0.201	0.457	**0.718**	0.239	0.433	0.194	-0.045	0.675	0.677	**0.779**	1.000	0.485	0.562	0.160	0.467	0.617
M40	**0.753**	**0.749**	0.360	-0.019	**0.741**	0.576	0.329	0.619	0.539	0.144	0.663	0.116	-0.017	0.282	**0.916**	0.542	0.485	1.000	0.243	**0.860**	**0.736**	0.240
M41	0.220	0.664	**0.871**	0.133	0.561	0.110	-0.052	0.237	0.022	0.247	0.330	0.473	-0.045	0.086	0.413	0.692	0.562	0.243	1.000	-0.029	0.471	0.188
M43	**0.775**	0.570	0.149	0.125	0.433	0.367	0.486	0.625	0.353	0.271	0.612	0.135	0.135	0.167	**0.739**	0.203	0.160	**0.860**	-0.029	1.000	0.615	0.158
M45	0.542	**0.906**	0.556	-0.062	0.452	0.110	0.317	0.331	0.197	0.257	0.584	0.174	-0.040	0.376	**0.715**	0.561	0.467	**0.736**	0.471	0.615	1.000	0.458
M46	0.222	0.374	0.218	-0.040	0.226	0.295	0.313	0.175	0.481	0.258	0.172	0.167	-0.058	**0.888**	0.415	0.408	0.617	0.240	0.188	0.158	0.458	1.000

Note:

[1]M3 = Hexadecane; M4 = 1-iodo-2-methylundecane; M5 = Pentadecane; M6 = Undecane; M8 = 2,4,6-trimethyl octane; M9 = 4-methyl-1-undecene; M11 = 1-iodoundecane; M13 = Eicosane; M14 = 2 methyl heptadecane; M15 = Allyl nonyl ester oxalic acid; M17 = 2,4-bis(1,1-dimethylethyl) phenol; M22 = 2,4-dimethyl undecane; M28 = 2,3,3-trimethyl pentane; M31 = 2-methyl dodecane; M32 = 2,3,5,8-tetramethyl decane; M35 = 3,3-dimethyl hexane; M39 = 2,6,10,14-tetramethyl heptadecane; M40 = 2-benzoyl methyl ester benzoic acid; M41 = Allyl octyl ester oxalic acid; M43 = Tetradecane; M45 = 2,9-dimethyl decane; and M46 = 1-iodo-tetradecane.

[2] $|0.000| < R < |0.300|$ = weak correlation, $|0.300| < R < |0.700|$ = moderate correlation and $|0.700| < R < |1.000|$ = strong correlation matrix.

[3]Correlation matrix with the bold value indicated a strong correlation between two metabolites

From Table 3, 2,3,5,8-tetramethyl decane had 7 strong CMs followed by 2,4,6-trimethyl octane and 2-benzoyl methyl ester benzoic acid (6 CMs), hexadecane (3 CMs), and 1-iodo-2-methylundecane (4 CMs). Also, pentadecane, 4-methyl-1-undecene, eicosane, 2-methyl heptadecane, 3,3-dimethyl hexane, 2,6,10,14-tetramethyl heptadecane, tetradecane, and 2,9-dimethyl decane individually had 3 CMs. Among these metabolites, the 2,3,5,8-tetramethyl decane had strong correlation with 2-benzoyl methyl ester benzoic acid (R = 0.916), 2,4,6-trimethyl octane (R = 0.858), hexadecane (R = 0.785), 1-iodo-2-methylundecane (R = 0.783), tetradecane (R = 0.739), 2,9-dimethyl decane (R = 0.715) and 4-methyl-1-undecene (R = 0.701). Although these metabolites rendered strong CM, the CM value only explained the correlation between two metabolites [20] while the CMs among more than two metabolites bring more meaningful information, where the PCA can cater to this purpose [21].

3.2 Metabolite correlations

The Pearson correlation test was chosen to select suitable correlation types to determine correlations between two metabolites. The correlation explores the dataset in investigating the strength (weak, moderate, or strong) and direction (positive and negative) of the linear relationship between two metabolites. Table 2 shows the correlation matrix (CM) of the metabolites in the urine of dengue-infected patients. The CM measures the strength and direction of a linear relationship between two metabolites. The CM indicated that positive and negative correlations among the metabolites at $|0.700| < R < |1.000|$ for strong CM, $|0.300| < R < |0.700|$ for moderate CM, and $3|0.000| < R < |0.300|$ for weak CM as in generic guideline. From Table 3, 2,3,5,8-tetramethyl decane had 7 strong CMs followed by 2,4,6-trimethyl octane and 2-benzoyl methyl ester benzoic acid (6 CMs), hexadecane (3 CMs), and 1-iodo-2-methylundecane (4 CMs). Also, pentadecane, 4-methyl-1-undecene, eicosane, 2-methyl heptadecane, 3,3-dimethyl hexane, 2,6,10,14-tetramethyl heptadecane, tetradecane and 2,9-dimethyl decane individually had 3 CMs. Among these metabolites, the 2,3,5,8-tetramethyl decane had strong correlation with 2-benzoyl methyl ester benzoic acid (R = 0.916), 2,4,6-trimethyl octane (R = 0.858), hexadecane (R = 0.785), 1-iodo-2-methylundecane (R = 0.783), tetradecane (R = 0.739), 2,9-dimethyl decane (R = 0.715) and 4-methyl-1-undecene (R = 0.701). Although these metabolites rendered strong CM, the CM value only explained the correlation between two metabolites [20] while the CMs among more than two metabolites bring more meaningful information, where the PCA can cater to this purpose [21].

Table 3. Metabolites with strong and moderate factor loading in the urine of dengue-infected patients

Metabolite	Designation	Factor loading (FL)[1,2]		
		F1	F2	F3
Hexadecane	M3	**0.8415**[a]	0.1805	-0.0806
1-iodo-2-methylundecane	M4	**0.7572**[a]	-0.1588	0.4861
Pentadecane	M5	**0.6556**[b]	0.0136	**0.6144**[b]
Undecane	M6	0.2016	**0.8113**[a]	-0.1352
Heptacosane	M7	**0.5897**[b]	**0.5315**[b]	-0.1862
2,4,6-trimethyl octane	M8	**0.8207**[a]	-0.2194	0.1018
4-methyl-1-undecene	M9	**0.6737**[b]	-0.1698	-0.4236
Eicosane	M13	**0.7683**[a]	0.0420	-0.1401
2,4-bis(1,1-dimethylethyl) phenol	M17	**0.7467**[b]	-0.0005	0.1296
Dodecamethyl cyclohexasiloxane	M20	0.1414	**0.7147**[b]	-0.1438
2,4-dimethyl undecane	M22	0.4259	**0.6404**[b]	0.1520
2,3,3-trimethyl pentane	M28	0.1722	**0.8061**[a]	-0.2255
2-bromooctane	M29	**0.6308**[b]	0.4901	0.2031
2-methyl dodecane	M31	**0.5367**[b]	-0.4156	-0.4666
2,3,5,8-tetramethyl decane	M32	**0.9265**[a]	-0.1847	0.0067
3,3-dimethyl hexane	M35	**0.7333**[b]	-0.1918	0.3373
Hexyl octyl ester sulfurous acid	M38	0.3339	**0.5490**[b]	0.1871
2,6,10,14-tetramethyl heptadecane	M39	**0.7423**[b]	-0.3650	0.0196
2-benzoyl methyl ester benzoic acid	M40	**0.8040**[a]	-0.1474	0.0605
Tetradecane	M43	**0.6782**[b]	0.1122	-0.1129
2,9-dimethyl decane	M45	**0.6970**[b]	-0.1692	0.3435
Eigenvalue		11.4543	4.0302	3.1970
Variability, %		24.9006	8.7614	6.9500
Cumulative explained variability (CEV), %		24.9006	33.6619	40.6120

Note:
[1]a FL ≥ |0.750| = strong factor loading and b |0.500| < FL < |0.749| = moderate factor loading in the principal component.
[2] Factor loading with bold superscript a indicated strong and moderate factor loading in the principal component.

3.3 Dataset exploration by principal component analysis

The purpose of using PCA in this study was to investigate the metabolite contribution to the urines and identify the variance of intercorrelated metabolites when patients were infected with the dengue virus.

Fig. 1 (a) shows the correlation biplot between the metabolites and urines which explains the metabolite contribution to the individual urine samples. The F1 and F2 axes explained 8.76% and 24.90% of the dataset respectively, which cumulatively explained 33.66% of the total dataset. From the correlation biplot in Fig. 1 (a), C1, C2 and C3 replicate i.e control urines of dengue-free patients were located in the F1 < 0 axes only. Also, D1, D2, D3, and D5 belonged together with these control urines. Fig. 1 (b) depicted that hexatriacontane, 2-ethylhexyl hexyl ester sulfurous acid, dotriacontane, hexyl pentyl ester sulfurous acid and 3-isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane were dominant in five control urines: C1-4, C2-3, C2-4, C3-2 and C3-5. Meanwhile, 2,2,3,3,5,6,6-heptamethyl heptane, 4,4-trimethyl hexane, 2,3,5,8-tetramethyl decane, and N,N,O-triacetylhydroxylamine were dominant in another 10 replicates of the control urines.

https://doi.org/10.31436/iiumej.v25i2.2785

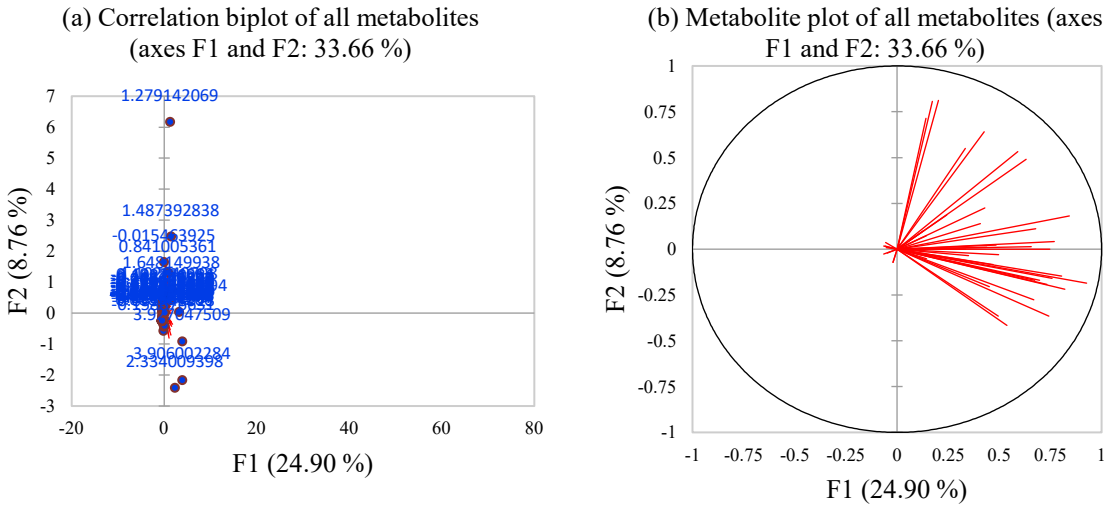


Figure 1. (a) Metabolite plot and (b) correlation biplot of all metabolites in the urine of dengue-infected patients

Only the urines of dengue-infected patients were present at F1 > 0, while no urine controls fell to this side (Fig. 1(a)). Moreover, 37 metabolites contributed to this site and since the urines present in this site are from the dengue-infected patients (Fig. 1(b)), this study proposed these metabolites to be the potential biomarkers in urine to discriminate the dengue-infected patients from the healthy ones. However, Daddiouaissa et al.(2021) proposed the selection of significant metabolites as biomarkers based on the calculated factor loading (FL) in the PCA; hence, this study selected and discussed metabolites with strong (FL ≥ |0.750|) and moderate (|0.500| < FL < |0.749|) only.

Table 3 exhibited the PCA with three PCs for the urine metabolites consisting of 21 significant metabolites ($p < 0.05$), respectively. Since the PCA is a two-dimensional PC, only PC1 and PC2 had cumulative explained variability (CEV) of 33.66% with an eigenvalue (EV) of 4.03 for the whole dataset. Within these PCs, 15 metabolites have a simultaneous and positive correlation among them in PC1, while 6 metabolites had simultaneous and positive correlation among them in PC2. Among them, hexadecane, 1-iodo-2-methylundecane, 2,4,6-trimethyl octane, eicosane, 2,3,5,8-tetramethyl decane, and 2-benzoyl methyl ester benzoic acid had strong factor loading (FL) in PC1 while the undecane and 2,3,3-trimethyl pentane had strong FL in PC2 (FL ≥ |0.750|), indicating that these metabolites were dominant in the urine of dengue-infected patients due to two individual potential factors.

The pentadecane, heptacosane, 4-methyl-1-undecene, 2,4-bis(1,1-dimethylethyl) phenol, 2-bromo octane, 2-methyl dodecane, 3,3-dimethyl hexane, 2,6,10,14-tetramethyl heptadecane, tetradecane and 2,9-dimethyl decane had moderate FL in PC1 while heptacosane, dodecamethyl cyclohexasiloxane, 2,4-dimethyl undecane and hexyl octyl ester sulfurous acid had moderate FL in PC2 (|0.500| < FL < |0.749|).

3.4 Discriminant of urines of dengue-infected patients and control

Table 1 shows the classification matrix of training and cross-validation datasets of urine of dengue-infected patients and controls by discriminant analysis. In this study, the establishment of

https://doi.org/10.31436/iiumej.v25i2.2785

discriminating model 1 (DM1) applied metabolites with strong and moderate factor loading from the PCA in the training dataset which rendered 90.91% and 100.00% correct classification for urines of dengue-infected patient and controls, respectively. The average of correct classification was 96.15% for a total of 26 urines. The correct classification < 100% was due to DM1 employing 21 metabolites entailing strong and moderate FLs, whereas the metabolites with moderate FL may have reduced the classification ability of DM1. Moreover, the location of D7-4 and D9-3 urines from the PCA, which was nearer to the cluster of urine controls, contributed to 96.15% correct classification. These reasons were supported by the location of both cluster centroids, as depicted in Fig. 2(a), which indirectly reduced the correct classification of the cross-validation dataset to 76.92% for the same number of total urines. The DM1 provided information on which metabolites were the most significant ones via F-statistic value and p-value. Since the DM1 was established at α of 0.05; therefore, selecting metabolites with F-statistic < 5 and p-value < 0.05, and removing D7-4 and D9-3 urines from the DM2 in Table 4 shows the 100% correct classification of the training dataset for the dengue-infected patient, controls, and average classification for a total of 24 urines. This result indicated that employing metabolites with F-statistic < 5 and p-value < 0.05, and urines except D7-4 and D9-3 had improved the correct classification of urines of dengue-infected patients and controls. This was evident in Fig. 2(b), which depicted the cluster centroids of dengue-infected patients and controls were further separated as compared to although C2-4 fell beyond the hoteling border. Additionally, correct classification for cross-validation in DM2 had also improved to 87.50% by using the same treatment as compared to DM1.

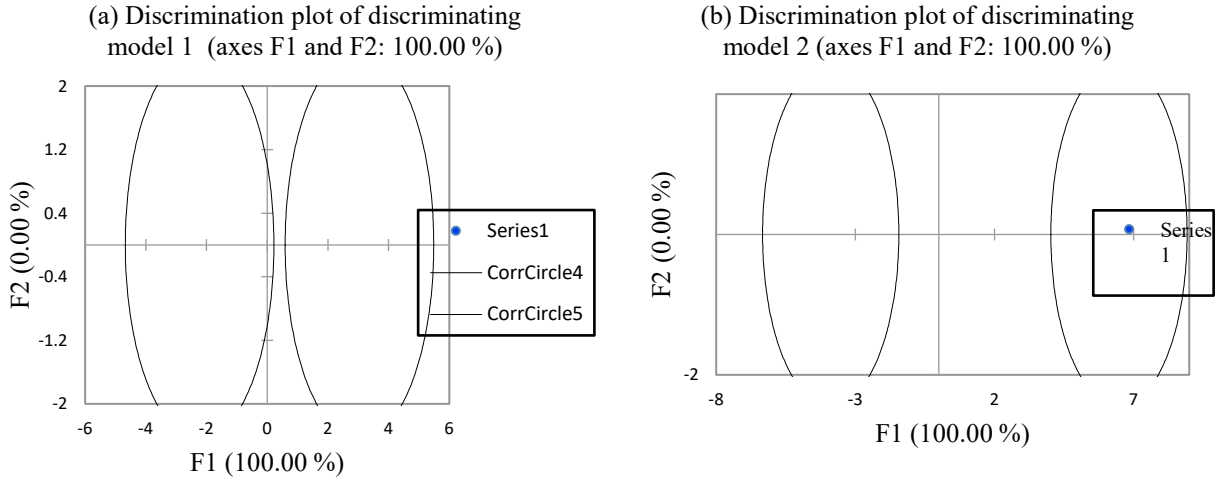


Figure 2. Discrimination plot of (a) discriminating model 1 and (b) discriminating model 2

Based on the F-statistic < 5 and p-value < 0.05, the DM2 had selected nine significant metabolites that could be proposed as potential biomarkers to assist in identifying dengue-infected patients. The nine metabolites' significance rankings were as follows: Heptacosane > Hexadecane > 2,4-bis(1,1-dimethylethyl) phenol > 2-bromooctane > tetradecane > hexyl octyl ester sulfurous acid > 2-benzoyl methyl ester benzoic acid > 2,9-dimethyl decane > pentadecane. Among these metabolites, six of them were alkane, one phenol, and two acid groups. Bouatra and co-workers[22] found the presence of 2,4-bis(1,1-dimethylethyl) phenol and benzoic acid group in a pool of human urine; however, they did not associate the presence of this metabolite with the dengue-infected patient.

https://doi.org/10.31436/iiumej.v25i2.2785

Table 4. Classification matrix of training and cross-validation datasets of urine of dengue-infected patients and controls by discriminant analysis

Discriminating model (DM)	**Treatment**	**Dataset**	**Correct classification, %**	**Number of urines and p-values of Fisher distance[1]**		**Total urine**	**Ranking of significant metabolites contributed to the DM (p < 0.05)2**
				The urine of a dengue-infected patient	Urine of control		
Discriminating model 1 (DM1)	Application of metabolites with strong and moderate factor loading from the PCA	Training dataset					Hexadecane (33.5959,) > Heptacosane (22.5566) > 2-bromooctane (12.0689) > Tetradecane (11.2645) > 2,4-bis(1,1-dimethylethyl) phenol (10.8625) > 2-benzoyl methyl ester benzoic acid (8.2477) > Hexyl octyl ester sulfurous acid (6.9570) > 2,9-dimethyl decane (6.1448) > Pentadecane (5.4461)
		The urine of a dengue-infected patient	90.91	10 (1)	1 (< 0.0011)	11	
		Urine of control	100.00	0 (< 0.0011)	15 (1)	15	
		Average	96.15	10	16	26	
		Cross-validation dataset					
		The urine of a dengue-infected patient	54.55	6 (1)	5 (< 0.0011)	11	
		Urine of control	93.33	1 (< 0.0011)	14 (1)	15	
		Average	76.92	7	19	26	
Training dataset The urine of a dengue-infected patient Urine of control Average Cross-validation dataset The urine of a dengue-infected patient Urine of control Average	(1) Application of metabolites with (a) strong and moderate factor loading from the PCA and (b) F-statistic < 5 and p-value < 0.05 only. (2) Removal of urine sample D7-4 and D9-3 100.00 100.00 100.00 66.67 100.00 87.50	Training dataset					Heptacosane (41.1045, < 0.0001) > Hexadecane (35.4330, < 0.0001) > 2,4-bis(1,1-dimethylethyl) phenol (13.6746, 0.0013) > 2-bromooctane (11.9207, 0.0023) > Tetradecane (10.6397, 0.0036) > Hexyl octyl ester sulfurous acid (9.5052, 0.0054) > 2-benzoyl methyl ester benzoic acid (8.9872, 0.0066) > 2,9-dimethyl decane (8.2742, 0.0088) > Pentadecane (6.7105, 0.0167)
		The urine of a dengue-infected patient	100.00	9 (1)	0 (< 0.0011)	9	
		Urine of control	100.00	0 (< 0.0011)	15 (1)	15	
		Average	100.00	9	15	24	
		Cross-validation dataset					
			66.67	6 (1)	3 (< 0.0011)	9	
			100.00	0 (< 0.0011)	15 (1)	15	
			87.50	6	18	24	

Note:
[1] Calculated p-value of Fisher distance < 0.05 indicated two clusters were significant difference.
[2] Values in parenthesis were F-statistic value and p-value

The production of alkane in the dengue-infected patient suggests that the host cells are experiencing oxidative stress as mentioned by Gil and co-workers [23], which then leads to lipid peroxidation [24] and alkane production [25]. Alkane produced was then secreted by the host through urine supported by Smith and co-workers [26] when they identified more than 200 volatiles in human urine, which belonged to different chemical classes such as alcohols, ketones, hydrocarbons, pyrroles, furans, aldehydes, terpenes, sulfur-containing compounds (isocyanates, sulfides), and O- and N-heterocyclic compounds. Courageot and co-workers [27] reported that the dengue (DEN) virus can trigger the host cells to undergo apoptosis in a cell-dependent manner and it is confirmed by Di Meo and co-workers[28] that dengue viruses caused apoptosis through lipid peroxidation. Suggesting that the alkane identified in our study is a result of dengue virus infection.

The 2,4-bis(1,1-dimethylethyl) phenol is a common toxic secondary metabolite produced by organisms and functions in ensuring endocidal regulation [29]. Moreover, Leila and co-workers[30] reported that the 2,4-bis(1,1-dimethylethyl) phenol exhibited significant anti-coxsackievirus B-3 (CVB-3) and anti-herpes virus type 2 (HSV-2) activities. Both studies supported the presence of a high amount of 2,4-bis(1,1-dimethylethyl) phenol in this study. Furthermore, Brynildsen and co-workers[31] through their study found that 2,4-bis(1,1-dimethylethyl) phenol also increased microbial ROS production that can potentiate killing the parasites.

4. CONCLUSION

This study shows that the GCMS approach can identify metabolite biomarkers from the urine of dengue patients. Heptacosane, Hexadecane, 2,4-bis(1,1-dimethylethyl) phenol, 2-bromooctane, tetradecane, hexyl octyl ester sulfurous acid, 2-benzoyl methyl ester benzoic acid, 2,9-dimethyl decane, and pentadecane are the potential biomarkers for dengue virus detection using a human urine sample.

ACKNOWLEDGMENT

The authors would like to thank the International Islamic University Malaysia for providing the facilities for this research.

REFERENCES

[1] Wilder-Smith A, Ooi EE, and Wills B. (2019) Dengue. The Lancet, 393(10169):350–363.

[2] Whitehorn J, and Simmons CP. (2011) The pathogenesis of dengue. Vaccine, 29(42). doi:10.1016/j.vaccine.2011.07.022

[3] Martina BEE. (2014) Dengue pathogenesis: A disease driven by the host response. Science Progress, 97(3). doi:10.3184/003685014X14049173153889

[4] Woon YL, Hor CP, Lee KY, Mohd Anuar SFZ, Mudin RN, Sheikh Ahmad MK, Komari S, Amin F, Jamal R, Chen WS, Goh PP, Yeap L, Lim ZR, Lim TO. (2018) Estimating dengue incidence and hospitalization in Malaysia, 2001 to 2013. BMC Public Health, 18(1). doi:10.1186/s12889-018-5849-z

[5] Nasar S, Rashid N, Iftikhar S. (2020) Dengue proteins with their role in pathogenesis, and strategies for developing an effective anti-dengue treatment: A review. Journal of Medical Virology, 92(8) doi:10.1002/jmv.25646

[6] Halstead SB. (2012) Controversies in dengue pathogenesis. Paediatrics and International Child Health, 32(SUPP1). doi:10.1179/2046904712Z.00000000045

[7] Ministry of Health. Clinical Practice Guideline. Management of Dengue Infection in Adults Revised Second Edition. (2014) https://www.moh.gov.my/moh/attachments/5502.pdf

[8] Muller DA, Depelsenaire ACI, Young PR. (2017) Clinical and laboratory diagnosis of dengue virus infection. Journal of Infectious Diseases, 215. doi:10.1093/infdis/jiw649

[9] Kikuti M. Cruz JS, Rodrigues MS, Tavares AS. Paploski IAD, Silva MMO, Santana PM, Tauro LB, Silva GAOF. Campos GS. et al. (2019) Accuracy of the SD BIOLINE Dengue Duo for rapid point-of-care diagnosis of dengue. PLoS ONE, 14(3). doi:10.1371/journal.pone.0213301

[10] Krishnananthasivam S, Fernando AN, Tippalagama R, Tennekoon R. De Man J, Seneviratne D, Premawansa S, Premawansa G, De Silva AD. (2015) Evaluation of a Commercial Rapid Test Kit for Detection of Acute Dengue Infection. The Southeast Asian journal of tropical medicine and public health, 46(4).

[11] Poloni TR, Oliveira AS, Alfonso HL, Galvo LR, Amarilla AA, Poloni DF, Figueiredo LT, V. Aquino H. (2010) Detection of dengue virus in saliva and urine by real time RT-PCR. Virology Journal, 7. doi:10.1186/1743-422X-7-22

[12] Andries AC, Duong V, Ong S, Ros S, Sakuntabhai A, Horwood P, Dussart P. Buchy P. (2016) Evaluation of the performances of six commercial kits designed for dengue NS1 and anti-dengue IgM, IgG and IgA detection in urine and saliva clinical specimens. BMC Infectious Diseases, 16(1). doi:10.1186/s12879-016-1551-x

[13] Andries AC, Duong V, Ly S, Cappelle J, Kim KS, Lorn TP, Ros S, Ong S, Huy R, Horwood P. et al. (2015) Value of Routine Dengue Diagnostic Tests in Urine and Saliva Specimens. PLoS Neglected Tropical Diseases, 9(9). doi:10.1371/journal.pntd.0004100

[14] Pawar R. (2015) Dengue Diagnosis: Challenges and Opportunities. Immunochemistry & Immunopathology, 1(1). doi:10.4172/2469-9756.1000105

[15] Cooper R. (2017). Diagnostic and statistical manual of mental disorders (DSM). Knowledge Organization, 44(8). doi:10.5771/0943-7444-2017-8-668

[16] Sani MSA, Bakar J, Rahman RA, Abas F. (2020) Effects of coated capillary column, derivatization, and temperature programming on the identification of Carica papaya seed extract composition using GC/MS Analysis. Journal of Analysis and Testing, 4:23–34. doi:10.1007/s41664-020-00118-z

[17] Daddiouaissa D, Amid A, Shirwan M, Sani A, and Ahmed AEM. (2021). Evaluation of metabolomics behavior of human colon cancer HT29 cell lines treated with ionic liquid graviola fruit pulp extract. Journal of Ethnopharmacology, 270:113813. doi:10.1016/j.jep.2021.113813

[18] Ismail AM, Sani MSA, Azid A, Zaki NNM, Arshad S, Tukiran NA, Abidin SASZ, Samsudin MS, Ismail A. (2021) Food forensics on gelatine source via ultra-high-performance liquid chromatography diode-array detector and principal component analysis. SN Applied Sciences, 3(79):1–19. doi:10.1007/s42452-020-04061-7

[19] Andrada MF, Vega-Hissi EG, Estrada MR, Garro Martinez JC. (2015). Application of k-means clustering, linear discriminant analysis and multivariate linear regression for the development of a predictive QSAR model on 5-lipoxygenase inhibitors. Chemometrics and Intelligent Laboratory Systems, 143:122–129. doi:10.1016/j.chemolab.2015.03.001

[20] Ratner B. (2009) The correlation coefficient: Its values range between +1/−1, or do they? Journal of Targeting, Measurement and Analysis for Marketing, 17(2):139–142. doi:10.1057/jt.2009.5

[21] Jannat B, Ghorbani K, Kouchaki S, Sadeghi N, Eslamifarsani E, Rabbani F, Beyramysoltan S. (2020) Distinguishing tissue origin of bovine gelatin in processed products using LC/MS technique in combination with chemometrics tools. Food Chemistry, 319:126302. doi:10.1016/j.foodchem.2020.126302

[22] Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P. et al. (2013). The Human Urine Metabolome. PLoS ONE, 8(9). doi:10.1371/journal.pone.0073076

[23] Gil L, Martínez G, Tápanes R, Castro O, González D, Bernardo L, Vázquez S, Kourí G, Guzmán MG. (2004) Oxidative stress in adult dengue patients. American Journal of Tropical Medicine and Hygiene, 71(5). doi:10.4269/ajtmh.2004.71.652

[24] Klassen P, Biesalski HK, Mazariegos M, Solomons NW, Fürst P. (2004). Classic dengue fever affects levels of circulating antioxidants. Nutrition, 20(6). doi:10.1016/j.nut.2004.03.016

[25] Saxena R. (2014) Arthritis as a Disease of Aging and Changes in Antioxidant Status. In: Aging: Oxidative Stress and Dietary Antioxidants, doi:10.1016/B978-0-12-405933-7.00005-6

[26] Smith S, Burden H, Persad R, Whittington K, De Lacy Costello B, Ratcliffe NM, Probert CS. (2008). A comparative study of the analysis of human urine headspace using gas chromatography-mass spectrometry. Journal of Breath Research, 2(3). doi:10.1088/1752-7155/2/3/037022

[27] Courageot MP, Catteau A, Desprès P. (2003) Mechanisms of Dengue virus-induced cell death. Advances in Virus Research, 60. doi:10.1016/S0065-3527(03)60005-9

[28] Di Meo S, Reed TT, Venditti P, Victor VM. (2016) Harmful and Beneficial Role of ROS. Oxidative Medicine and Cellular Longevity, 2016: 7909186. doi:10.1155/2016/7909186

[29] Zhao F, Wang P, Lucardi RD, Su Z, Li S. (2020) Natural sources and bioactivities of 2,4-di-tert-butylphenol and its analogs. Toxins, 12(1). doi:10.3390/toxins12010035

[30] Leila A, Lamjed B, Roudaina B, Najla T, Taamalli A, Jellouli S, Mokhtar Z. (2019). Isolation of an antiviral compound from Tunisian olive twig cultivars. Microbial Pathogenesis, 128. doi:10.1016/j.micpath.2019.01.012

[31] Brynildsen MP, Winkler JA, Spina CS, MacDonald IC, Collins JJ. (2013). Potentiating antibacterial activity by predictably enhancing endogenous microbial ROS production. Nature Biotechnology, 31(2). doi:10.1038/nbt.2458

https://doi.org/10.31436/iiumej.v25i2.2933

RECENT ADVANCES IN ENHANCED POLYAMIDOAMINE INHIBITORS FOR SILICATE SCALES IN THE PETROLEUM UPSTREAM

BEE CHEA TAN[1,2*], ISMAIL MOHD SAAID[1,2], SHIFERAW REGASSA JUFAR[1,2], SUZALINA ZAINAL[3], ABDELAZIM ABBAS AHMED AWADELSEED[4]

[1]*Petroleum Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia*
[2]*Centre of Research in Enhanced Oil Recovery, Institute of Hydrocarbon Recovery, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia*
[3]*PETRONAS Research Sdn Bhd, Bangi, Malaysia*
[4]*Wood PLC Group, Scotland, United Kingdom*

**Corresponding author: tan_22000121@utp.edu.my*

(Received: 17 July 2023; Accepted: 24 March 2024; Published online: 15 July 2024)

ABSTRACT: Chemical flooding is regarded as a promising enhanced oil recovery technique to recover more hydrocarbon from reservoirs. However, the dissolution of quartz minerals in a highly alkaline environment poses the risk of silicate scaling near the production well region from the mixing of two different waters. Commercial scale inhibitors are effective, but they are also harmful to the environment. This paper aims to provide insights into current advances in environment-friendly or "green" scale inhibitors for petroleum upstream. Previous research works have demonstrated that green chemicals are effective in mitigating silicate, carbonate, and sulfide scales. Polyamidoamine or amide-based inhibitors have been widely investigated in recent literature due to several advantages. The addition of anionic compounds in these inhibitors enhanced scale inhibition efficiency by roughly 10%. Nevertheless, the reported findings were deliberated for industrial wastewater treatment. Meanwhile, understanding the performance of polyamidoamine or amide-based scale inhibitors in petroleum upstream is inadequate to a certain extent. The formation process of silicate scales inside a reservoir is rather complicated by looking at the influence of water salinity, composition of brine, temperature, pressure, and rock type. Hence, it is essential to study and develop green scale inhibitors that are effective and environmentally friendly to meet increasingly stringent disposal regulations in the petroleum industry.

ABSTRAK: Pembanjiran kimia merupakan teknik pemulihan minyak. Ia berpotensi dalam memperoleh lebih banyak hidrokarbon dari takungan. Namun, pelarut mineral kuarza dalam persekitaran beralkali tinggi memberi risiko penumpukan silikat berhampiran kawasan takungan pengeluaran. Ia disebabkan oleh pencampuran dua jenis cecair berbeza. Perencat penumpukan silikat komersial adalah berkesan, tetapi sangat berbahaya pada alam sekitar. Kajian ini bertujuan bagi menambahbaik kemajuan perencat silikat mesra alam terkini atau perencat silikat hijau bagi bidang saliran petroleum. Kajian terdahulu telah membuktikan bahawa bahan kimia mesra alam adalah berkesan dalam pengurangan penumpukan silikat, karbonat dan sulfida. Perencat poliamidoamina atau perencat bersumber amida telah dikaji secara meluas dalam beberapa kajian sejak kebelakangan ini kerana kelebihannya yang banyak. Penambahan sebatian anionik dalam perencat ini mampu meningkatkan keberkesanan perencat silikat sebanyak 10%. Namun, laporan kajian terdahulu adalah khusus bagi rawatan sisa air industri. Sementara itu, pemahaman tentang prestasi perencat silikat bersumberkan poliamidoamina atau perencat bersumber amida dalam saliran petroleum masih tidak mencukupi. Proses pembentukan penumpukan silikat dalam takungan

adalah agak rumit berdasarkan faktor saliniti air, komposisi air garam, suhu, tekanan dan jenis batuan. Oleh itu, kajian dan pembangunan berkesan tentang perencat silikat mesra alam adalah penting bagi memenuhi peraturan pelupusan sisa yang semakin ketat dalam industri petroleum.

KEYWORDS: *Green Scale Inhibitor, PAMAM, Silicate, Formation Damage, and ASP Flooding.*

1. INTRODUCTION

Global oil demand is projected to rise and is expected to reach its peak either in the year 2024 at 98.1 million barrels per day or 103.2 million barrels per day in the year 2035 [1]. In addition, conventional oilfields are maturing, and the impact of waterflooding will wear off eventually. This could result in the rise of field implementation of tertiary oil recovery in the coming years. Tertiary recovery is the stage where advanced techniques are used to recover the remaining hydrocarbons from a reservoir. It is commonly known as enhanced oil recovery (EOR) in the petroleum upstream sector. These techniques include microbial injection, steam injection, gas flooding, chemical flooding, and even hybrid methods.

Chemical flooding is of special interest for its success in cost-effectively recovering significant amounts of residual hydrocarbon from matured oilfields. The common chemicals used are polymers, surfactants, and alkalis. They can be injected either singly or in combination according to the reservoir's characteristics and limitations [2, 3]. The combination of all three chemicals is referred to as alkaline surfactant polymer (ASP) flooding. The combination of alkali and surfactant in a chemical injection slug is to achieve overall interfacial tension reduction of fluids, meanwhile, the polymer is used to achieve better sweep efficiency in the reservoir. Compared to other chemical EOR types, ASP flooding could improve oil recovery as much as 45% in laboratory experiments and up to 28% in oilfield tests [4].

Although ASP flooding is deemed as a promising EOR technique, it has raised concern for silicate scaling. This could cause problems in terms of formation damage and flow assurance, thus hindering the productivity of a well. Scale can be deposited at any point of the petroleum production system, from near the wellbore region up to the surface facilities [5]. Several oilfields have reported the presence of silicate scale in their production systems. For instance, the silicate content in one of the Daqing's production wells was reported to be about 70% in scaling peak stage due to ASP flooding [6]. Meanwhile in Alberta, silicate scale was found inside pumps, pipelines, and facilities after one year of ASP injection [7]. Hence, it is important to manage oilfield scales when adopting an alkaline-based chemical flooding.

Chemical treatment using scale inhibitors is an effective approach to maintain well productivity by retarding or minimizing the risk of scaling at the production wellbore and its proximity region [5, 8]. For oilfield operations, phosphonates and phosphino-polyacrylates are the two commonly used scale inhibitors due to good thermal stability. Unfortunately, they are not intended to treat silicate scales. In fact, they are primarily used to inhibit carbonate, sulfate, and phosphate scales [9,10]. Besides that, the inherent biodegradability of phosphonates were reported to be less than 40% in 28 days, while phosphino-polyacrylates were reported as non-biodegradable [11].

Owing to rising environmental concerns and discharge restrictions, environment-friendly or "green" scale inhibitors were developed for sustainable alternatives. Carboxymethyl inulin (CMI), polyaspartic acid (PASP), polyepoxysuccinic acid (PESA), and chitosan are some

https://doi.org/10.31436/iiumej.v25i2.2933

examples of well-accepted green inhibitors in the industry. This paper provides an insight into the current state of green scale inhibitors and their challenges. Special attention is given to highlight the limitations of polyamidoamine-based formulations and alternative components for their enhancement in mitigating silicate scales. The formation of silicate scales due to ASP flooding and the mechanism of chemical scale inhibition are also described in this paper. The green chemicals used for scale inhibition are reviewed according to their effectiveness and mechanisms in reported laboratory works. The effectiveness of amine-based chemicals such as polyamidoamine dendrimers is also discussed accordingly.

2. FORMATION OF SILICATE SCALE

The alkaline slug in ASP injectant is often at pH 11 and above [12]. The excessive aqueous alkali would dissolve the quartz mineral of sandstone rock surface when they are in contact for sufficient time at elevated temperature and pressure, such as reservoir conditions [13]. Unfortunately, the dissolution of quartz is an irreversible chemical reaction at high pH environment. The dissolved quartz mineral is present in the form of monomeric silica in solution ($Si(OH)_3O^-$ Na^+) that is usually soluble and stable at high pH reservoir conditions [14].

As the alkaline solution (pH ≥11) that contains monomeric silica flows towards the production well, the solubility of monomeric silica decreases due to pH reduction when gradually mixing with formation water or connate water that is usually in the range of pH 6 to pH 8 [15]. The soluble monomeric silica begins to polymerize and form colloidal silica when pH drops below 10.5. When the pH drops further to 8.5, the colloidal silica polymerizes further and forms insoluble amorphous silicate scale [16, 17]. In addition, the monomeric silica also tends to form metallic silicates in the presence of excessive divalent cations in connate water, such as magnesium (Mg^{2+}) and calcium (Ca^{2+}) ions [18, 19]. As such, Fig. 1 illustrates the mixing of two different waters in a production well.

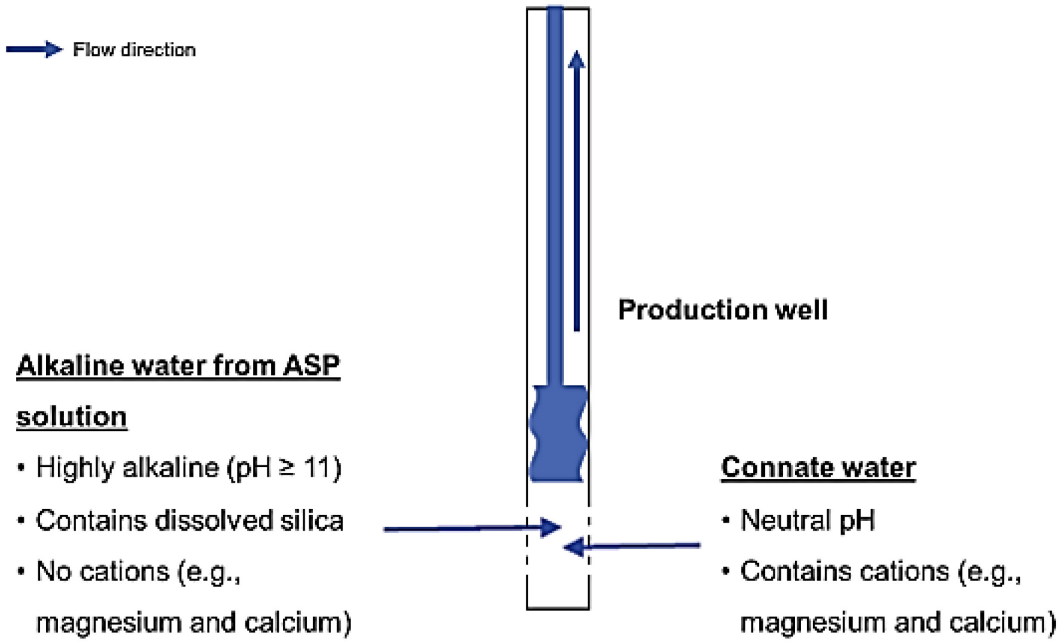


Fig. 1. Mixing of ASP solution and connate water in a production well, modified from Arensdorf et al. [20].

Based on the process described above, the pH change in alkaline water and divalent cations in connate water are the two important factors that influence the final form of the silicates at constant temperature and pressure. The research findings on the effects of pH and divalent cations that contribute to silicate scale formation are summarized in Table 1.

Table 1: Experimental findings of factors that influence the final form of silicate scales

Factor	Main Findings	Reference
pH	The experiment was carried out using fixed concentration of magnesium and silicon ions by varying the final mixture's pH. Amorphous silica and magnesium silicate were identified in the beginning at pH 8.5. Infrared spectrum revealed the tendency of magnesium silicate scale formation increased with increasing pH. However, the reduction of pH on silicate scale formation is not conducted in this study.	[17]
	This review article mentioned that silica scale formation usually takes place at pH less than 8.5 whereas magnesium silicate scale forms at pH 8.5 and above if water contained significant magnesium cations.	[16]
Divalent Cations	Test brine is composed of monovalent ions, divalent ions, and silicon. The ScaleSoftPitzer™ software predicts that calcium silicate and calcium/magnesium silicate are the two dominant scales with saturation index of 17.51 and 10.35, respectively. The environment conditions were set at 95°C temperature, 14.7 psia pressure, and brine with pH 10.3.	[18]
	Different concentrations of silicon, magnesium, and calcium ions in the final brine mixture were studied. Pure silicon brine of pH 5 turned cloudy and polymerized after 22 hours. At the same pH condition, severe white precipitates were formed in the beginning when magnesium and calcium ions were added in silicon brines. The same reaction occurred when the final mixture was adjusted to pH 8.5 and pH 11. Imaging showed that these precipitates were amorphous solid silicate scales with traces of microcrystals.	[19]

3. MECHANISMS OF CHEMICAL SCALE INHIBITION

In comparison to mechanical scale removal, chemical scale inhibitors are preferred in oilfield practice due to several advantages. The chemicals are usually added upstream of the problematic region and must be constantly present in the scaling water to inhibit scales from precipitating and growing [21]. A scale inhibitor can work in one or several ways in accordance with different stages of scale formation. With the aid of Fig. 2, the following five mechanisms of scale inhibition are disseminated [5, 8]:

- Chelating, whereby the inhibitor ions control water hardness and scaling ion deposits.
- Interfering nucleation process, whereby the inhibitor ions reach and to disrupt scaling ion clusters, thus preventing further growth of the existing clusters.
- Retarding crystal growth, whereby the inhibitor ions diffuse over the crystal surface and active sites through strong affinity, thus promoting the formation of small crystals to reduce supersaturation of the solution. It would be an advantage if the inhibitor ions

are absorbed into the crystal to make the crystal become soft, friable, and easy to remove.

- Modifying crystal surface, whereby the inhibitor ions adsorb strongly onto the crystal surface and prevent attachment of crystals.
- Dispersing flocculation, whereby the inhibitor ions modify the kinetics of nucleation and growth of scale on surfaces.

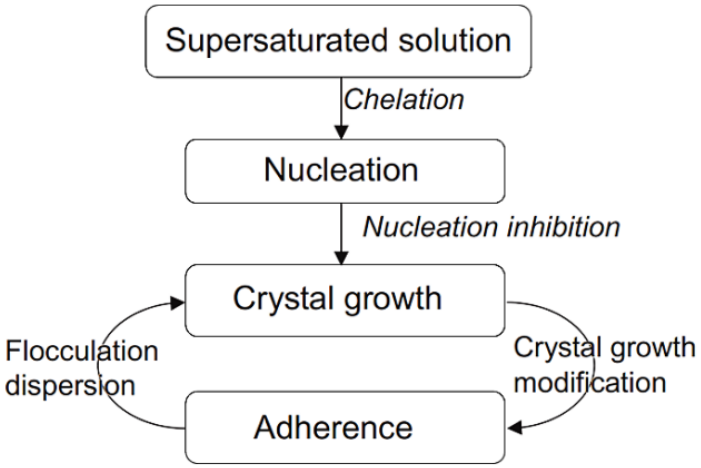


Fig. 2. Scale inhibition mechanisms (on arrow) at different stages of scale formation (in boxes), adapted from Wu [5].

It is essential to determine the inhibition mechanisms for successful application. This will allow the chemical inhibitors to work more efficiently for specific scale types, hence leading to optimum industrial operations. Nonetheless, there are some challenges in determining the precise mechanisms. In reality, scales can exhibit diverse mechanisms and evolve over time due to variation in operational conditions. This may degrade the performance of inhibitors in comparison to laboratory work. As a result, bridging the knowledge gap between research studies and practical application could benefit the formulation and development of existing inhibitors for sophisticated scales.

4. GREEN SCALE INHIBITORS FOR SILICATES

Fink (2021) suggested that an ideal scale inhibitor should possess low toxicity and be highly biodegradable [12]. Because of the increasing environmental awareness and stringent disposal regulations, environment-friendly or “green” scale inhibitors are continuously developed and improvised according to scale types and reservoir conditions. Over the past 10 years, research on green scale inhibitor is drawing the interest of stakeholders. As provided in Fig. 3, the number of research articles on green scale inhibitor has increased four-fold in Scopus and six-fold in Web of Science databases.

https://doi.org/10.31436/iiumej.v25i2.2933

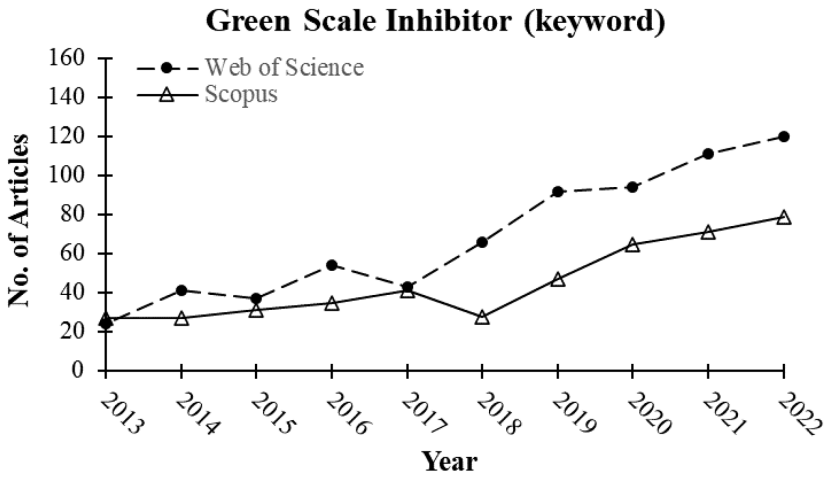


Fig. 3. Number of research articles on green scale inhibitor from year 2013 to 2022, reproduced from databases of Scopus [22] and Web of Science [23].

Numerous polymeric inhibitors have been investigated for their effectiveness against silicate scaling. Among others, cationic polymers and neutral polymers were found effective for silicate scale inhibition [24]. Green scale inhibitors such as amine-based polymers [25-28], modified resin [29], and grafted polymethacrylates [30, 31] were tested in the recent works. The performance and mechanism of these green scale inhibitors are summarized in Table 2.

Table 2: Summary of the performance and mechanism of green scale inhibitors from reported laboratory experiments

Green Chemicals	Max. Scale Inhibition	Parameters	Mechanism	Reference
PER-PAMAM G1	Silicate: ~ 85% Carbonate: ~ 85%	Conc.: 5–120 ppm pH: 5–9 Base: Pure water Temp.: 30–80°C Time: 10–30 hrs	Crystal surface modification	[25]
PER-PAMAM G2	Silicate: ~ 79% Carbonate: ~ 79%			
PER-PAMAM G1 + PAPEMP	Silicate: ~ 95% Carbonate: ~ 96%	Conc.: 80 + (5–45) ppm pH: 7 Base: Pure water Temp.: 40°C, 80°C Time: 10–30 hrs	N/A	
PAMAM-0 G-Eth	Silicate: 80%	Conc.: 10–80 ppm pH: 5–9 Base: Pure water Temp.: 30°C Time: 12–24 hrs	Crystal growth retardation	[26]
PAMAM-0 G-Prop	Silicate: 83%			
PAMAM-0 G-But	Silicate: ~ 81%			
PAMAM-0 G-Hex	Silicate: 77%			
AA/AT/DE + PESA	Silicate: ~ 90%	Conc.: (20–70) + (5–50) ppm pH: 7 Base: Pure water Temp.: Ambient Time: 24–96 hrs	Nucleation interference and crystal growth retardation	[27]

https://doi.org/10.31436/iiumej.v25i2.2933

Green Chemicals	Max. Scale Inhibition	Parameters	Mechanism	Reference
CPEG + TOMAC	Silicate: N/A	Conc.: 2000 ppm + (1–5)% pH: 3–9 Base: N/A Temp.: 90°C Time: 1 hr	Nucleation interference	[28]
GA-type resin	Silicate: 80%	Conc.: ~1440 ppm pH: 4–10 Base: Pure water Temp.: 25°C Time: 2–130 hrs	Nucleation interference	[29]
OH-type resin	Silicate: 5%			
PEGPHOS-LOW	Silicate: ~ 84% Carbonate: ~ 98% Sulfide: ~ 95%	Conc.: 500–5000 ppm pH: 8.5 Base: Pure water Temp.: Ambient Time: 3 hrs	Chelation and flocculation dispersion	[30]
HOMOPEG	Silicates: ~ 12% Sulfides: < 5% Carbonate: ~ 5%	Conc.: 100–400 ppm pH: 8.5 Base: Pure water Temp.: Ambient Time: 3 hrs	Inactive	[31]
HOMOPHOS	Silicates: ~ 15% Sulfides: ~ 20% Carbonate: ~ 95%		Chelation	
PEGPHOS-LOW	Silicates: 81% Sulfides: ~ 75% Carbonate: < 10%			
PEGPHOS-HIGH	Silicates: 69% Sulfides: ~ 30% Carbonate: ~ 30%			

PER-PAMAM G1 = first generation polyamidoamine dendrimer with pentaerythritol-acrylonitrile adduct initiator core; PER-PAMAM G2 = second generation polyamidoamine dendrimer with pentaerythritol-acrylonitrile adduct initiator core; PAPEMP = polyamine polyether methylene phosphonic acid; PAMAM-0 G-Eth = zeroth generation amine-terminated dendrimer synthesized with 1,2-ethylenediamine; PAMAM-0 G-Prop = zeroth generation amine-terminated dendrimer synthesized with 1,3-diaminopropane; PAMAM-0 G-But = zeroth generation amine-terminated dendrimer synthesized with 1,4-butanediamine; PAMAM-0 G-Hex = zeroth generation amine-terminated dendrimer synthesized with 1,6-diaminohexane; AA = polycation adipic acid; AT = amine-terminated polyether D230; DE = diethylenetriamine copolymer; PESA = polyepoxysuccinic acid; CPEG = chitosan-polyethylene glycol composite; TOMAC = tri octyl methyl ammonium chloride; GA= gallic acid; OH = hydroxide; HOMOPEG = polyethylene glycol grafted polymethacrylate; HOMOPHOS = methyl phosphonic acid grafted polymethacrylate; PEGPHOS-LOW = polymethacrylate grafted with methyl phosphonic acid and lower short-chain polyethylene glycol ratio; and PEGPHOS-HIGH = polymethacrylate grafted with methyl phosphonic acid and higher short-chain polyethylene glycol ratio.

It can be inferred from Table 2 that effective scale inhibition can be achieved with low concentrations of green chemicals. They also showed good thermal stability which is also suggested in terms of ideal scale inhibitor [12]. On top of that, scale inhibition was improved further with the chemical additive of polyamine polyether methylene phosphonic acid (PAPEMP) and polyepoxysuccinic acid (PESA). Other than silicate scales, these green scale inhibitors were also examined for their effectiveness on carbonate and sulfide scales, which are commonly encountered in oilfields. Furthermore, it can also be deduced that several green scale inhibitors could work in two scale inhibition mechanisms. As a result, green scale inhibitors could be utilized as alternatives for phosphorous commercial chemical inhibitors, especially amine-based inhibitors.

https://doi.org/10.31436/iiumej.v25i2.2933

4.1. Polyamidoamine (PAMAM) Dendrimer Inhibitors

Polyamidoamine (PAMAM) dendrimers as silicate scale inhibitors were investigated extensively in recent works. It has been demonstrated from earlier research that PAMAM inhibitors with amine group (–NH_2) terminal were effective in inhibiting silicate scale formation because its ammonium cations would bind with silicate anions in a solution. However, the formation of the PAMAM-silica composite was evident in the reported research. This situation has deteriorated the performance of PAMAM inhibitors over time [32-35]. Consequently, recent studies have modified the existing PAMAMs to improve their inhibition performance.

Sun et al. (2021) have synthesized the zeroth generation PAMAM with reactive cores of different carbon chain lengths to inhibit silica growth. The study was intended to investigate the inhibitory effect of PAMAM inhibitors with different carbon chain lengths. The study observed that the best scale inhibition performance was PAMAM-0 G-Prop, the synthesized PAMAM that is made up of medium carbon chain length of reaction cores. In addition, these synthesized PAMAMs also altered the morphology of the silica scale [26]. The molecular structure of the PAMAM-0 G-Prop is illustrated in Fig. 4. It has eight amine groups, –NH_2 and –NH, which are distributed evenly in the carbon chains.

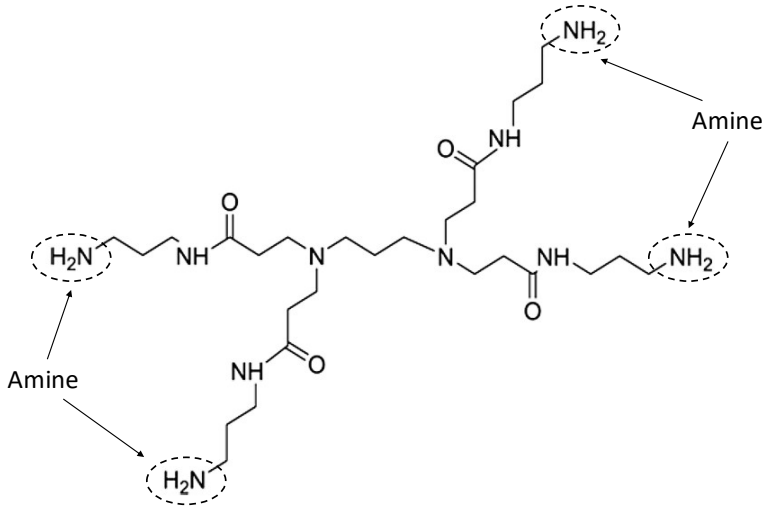


Fig. 4. Synthesized zeroth generation polyamidoamine dendrimer with 1,3-diaminopropane (PAMAM-0 G-Prop), reproduced from Sun et al. [26].

Furthermore, Shi et al. (2022) modified the initiator core of first and second-generation PAMAMs from ethylenediamine to pentaerythritol-acrylonitrile adduct (PER-PAMAMs). The study aimed to diversify the existing PAMAM scale inhibitors. The study found that PER-PAMAM G1 exhibited better silicate scale inhibition and yielded the best inhibition performance when 15 ppm of PAPEMP was added at the tested conditions. Likewise, the morphology of the silicate scale was also altered due to PER-PAMAM inhibitors [25]. The molecular structures of PER-PAMAM G1 are provided in Fig. 5. It also consists of eight amine groups but with pentaerythritol-acrylonitrile adduct initiator core, such that four oxygen atoms are in the molecule's core.

https://doi.org/10.31436/iiumej.v25i2.2933

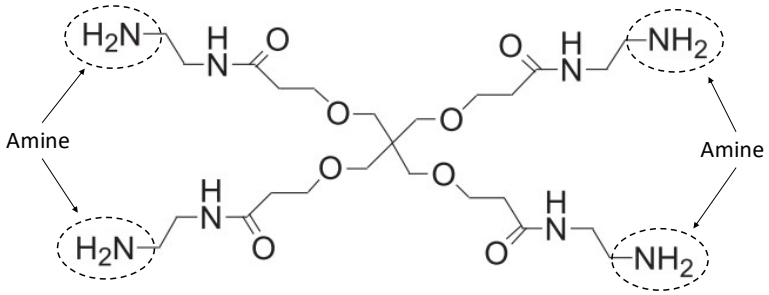


Fig. 5. Synthesized first-generation polyamidoamine dendrimer with pentaerythritol-acrylonitrile adduct initiator core (PER-PAMAM G1), reproduced from Shi et al. [25].

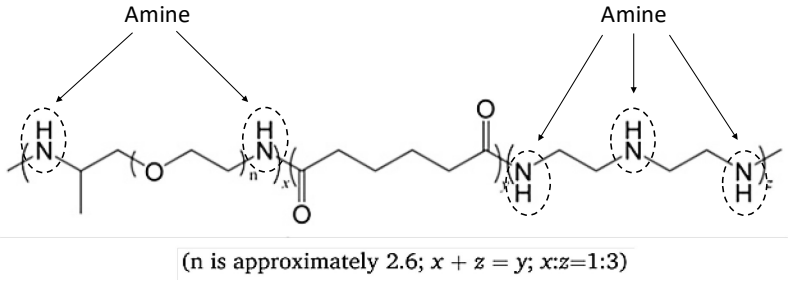
Based on both studies, the length of the carbon chain has more influence than the molecular core itself. Apparently, PAMAM dendrimer with medium carbon chain length is the optimum inhibitor for industrial water systems. The silicate scale inhibition test was conducted using distilled water at neutral pH and temperature up to 40°C. Unfortunately, this condition does not represent the typical oil reservoir. Thus, it is uncertain regarding the effectiveness of these inhibitors if they were undertaken for saline water, a wider range of pH, as well as greater temperatures and pressures. Moreover, PAMAM chemicals are expensive due to their specialty. Consequently, other amine-terminated polymeric chemicals shall also be assessed and improvised for silicate scale inhibition.

4.2. Amine-terminated AA/AT/DE Inhibitor

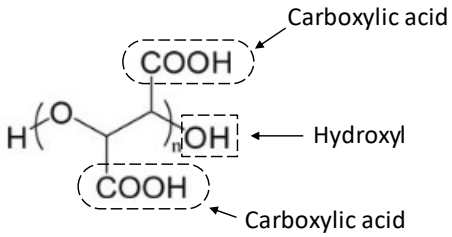
Another amine-based inhibitor, AA/AT/DE, was also tested for silicate scale inhibition. It is made up of adipic acid (AA), amine-terminated polyether D230 (AT), and diethylenetriamine (DE) at a molar ratio of 4:1:3, respectively. In the study, 50 ppm of AA/AT/DE inhibitor stabilized 500-ppm silicate at the tested conditions. Similar to the PAMAM inhibitor, a small amount of polymer floc appeared in the solution because the AA/AT/DE inhibitor was cationic at pH 7. The addition of polyepoxysuccinic acid (PESA) into the AA/AT/DE inhibitor has prevented floc formation and improved inhibition as well. For the same silicate concentration, 40-ppm AA/AT/DE added with 40-ppm PESA was the most optimal inhibitor formulation [27].

The chemical structures of AA/AT/DE inhibitor and PESA are provided in Fig. 6 (a) and Fig. 6 (b). The figure shows that AA/AT/DE inhibitor carries $–NH_2$ functional groups, while PESA carries carboxyl (–COOH) as well as hydroxyl (–OH) functional groups. Previous work also revealed that PESA alone was not effective towards silica scale formation under the tested conditions [27].

Based on the previous work, this inhibitor was tested to inhibit silica for a reverse osmosis membrane at room temperature and in distilled water [27]. Therefore, it remains uncertain whether AA/AT/DE inhibitors could perform similarly under reservoir conditions that have higher temperatures and pressures. The compatibility of formation water and AA/AT/DE inhibitor is not known either. Besides that, the study has not reported whether this inhibitor is feasible above neutral pH, which may be useful in treating magnesium silicate or aluminum silicate.



(a) AA/AT/DE



(b) PESA

Fig. 6. Chemical structures of (a) adipic acid/ amine-terminated polyether D230/ diethylenetriamine copolymer (AA/AT/DE) inhibitor and (b) polyepoxysuccinic acid (PESA), reproduced from Tan et al. [27].

4.2. Chitosan-Polyethylene Glycol Composite (CPEG) Inhibitor

Chitosan is a biopolymer that carries both $–NH_2$ and –OH functional groups, meanwhile, polyethylene glycol (PEG) is a neutral polymer that carries –OH groups only. The amine group in chitosan possesses adsorption capacity to eliminate anions from aqueous solutions, and the addition of PEG is believed to improve the adsorption capacity of chitosan [36]. The change in chitosan's structure due to PEG addition is illustrated in Fig. 7 below. The addition of PEG polymer, represented in shaded circles, seems to strengthen chitosan's structure without changing its number of functional groups.

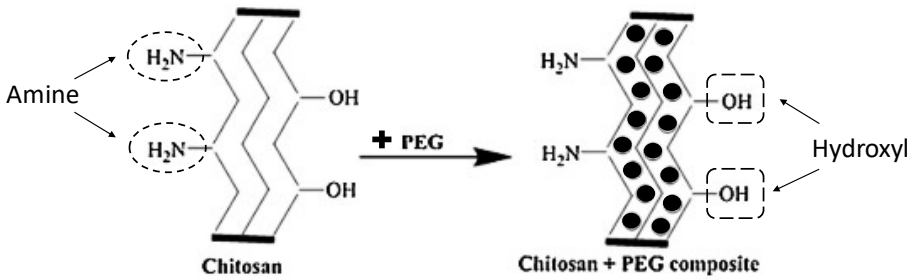


Fig. 7. Structural change in chitosan due to polyethylene glycol (PEG), reproduced from Rajeswari et al. [36].

This composite of chitosan and polyethylene glycol (CPEG) was studied in silicate scale inhibition. Hayati et al. (2021) reported that the highest adsorption capacity of CPEG was achieved at pH 6 with approximately 7.5 mg/g. Moreover, the additive of 1% tri octyl methyl ammonium chloride (TOMAC) emulsifier in CPEG has significantly increased the adsorption capacity of CPEG to 72.6 mg/g [28].

Nevertheless, the effectiveness of CPEG and CPEG-TOMAC inhibitors on silicate scale inhibition was not measured, and a mere 20 ppm of silicate was tested in the mentioned work. Moreover, the paper is focused on the elimination of silicic acid ($H_2SiO_4^-$) in geothermal fluid from the Dieng Geo Dipa Power Plant in Indonesia. Hence, these polymeric inhibitors are yet to be examined in the petroleum industry in which the composition of the formation water can be different from the geothermal fluid. Sodium chloride is a dominant compound in most formation water. However, the composition of the geothermal fluid is not revealed in the paper for comparison.

5. ENHANCED POLYAMIDOAMINE SCALE INHIBITORS

The advantages of polyamidoamine (PAMAM) dendrimers have gained interest in the research field. They have relatively large surface functional groups, exhibit greater biocompatibility due to amine groups, and exhibit decent solubility in water. These characteristics have positioned PAMAMs as environment-friendly or "green" scale inhibitors. Besides that, they also have good adsorption and encapsulation functions for metallic cations, such as cadmium and nickel [37, 38]. Even though they are effective in inhibiting amorphous silica scale at low concentrations, they tend to produce insoluble PAMAM-silica composite in water because of the incomplete interaction of anionic silica and cationic inhibitor.

Table 3: Summary of investigated anionic compounds for PAMAM scale inhibitor's enhancement before the year 2018

Anionic Compound	Best Formulation	Max. Scale Inhibition	Condition	Reference
PGLU	5000 ppm (+ 10 ppm of PAMAM-2) with volume ratio of 2:3	Silicate: ~70%	pH: N/A Base: Synthetic brine Temp.: N/A Time: 12–24 hrs	[39]
SAA	Synthesized PAMAM G4.0 with mass ratio 1.0	Carbonate: 92.8%	pH: 7 Base: N/A Temp.: 40–80°C Time: 10 hrs	[40]
PAM-co-AA	20 ppm (+ 40 ppm of PAMAM-2)	Silicate: 6.28[a]	pH: 7 Base: Pure water Temp.: N/A Time: 12–72 hrs	[41]
PAA (low molecular weight)				
PAA (high molecular weight)		Silicate: 6.16[a]		
CMI		Silicate: 6.75[a]		

[a] Defined as the number of silicate molecules stabilized as found experimentally divided by the estimated silicate molecules stabilized per molecule of dendrimer.

PGLU = pteroyl-L-glutamic acid; SAA = sodium acrylate; PAM-co-AA = poly(acrylamide-co-acrylate); PAA = polyacrylate; CMI = carboxymethyl inulin; PAMAM-2 = second generation polyamidoamine dendrimer; and PAMAM G4.0 = fourth generation polyamidoamine dendrimers.

It remains controversial whether more sophisticated scale types could be produced and deposited at any point in time. Subsequently, an anionic compound is added to the cationic PAMAM inhibitor to enhance the inhibition performance. Apart from the aforementioned PAPEMP and PESA, Table 3 sets out other anionic compounds that have been investigated for PAMAM inhibitor enhancement in the last decade. These include pteroyl-L-glutamic acid [39], polyacrylates [40], carboxymethyl inulin, and poly(acrylamide-co-acrylate) [41]. In the case of distilled water, a small volume of anionic compounds was sufficient to enhance PAMAM inhibitors. The mass ratios of cationic to anionic compounds were found to be 1:1, 2:1, and 1:018 in the reported works. In contrast, a greater concentration of anionic compounds may be needed for reservoir conditions. For instance, 5000 ppm of anionic pteroyl-L-glutamic acid was added with a small volume of cationic PAMAM inhibitors [39]. Its inhibitory mechanism was briefly described; however, the characteristics of the formulation itself are not presented in the paper. It is apparent from Table 2 and Table 3 that not much literature employed PAMAM in silicate scale inhibition studies. This may hold back eventual pilot testing due to insufficient data.

6. SUMMARY AND FUTURE DIRECTIONS

Silica or silicate ions are negatively charged upon dissolution in water. Thus, polymeric inhibitors with positive change are believed to be effective in impeding scale formation by attracting negatively charged scaling ions. Many positively charged or cationic inhibitors usually carry at least an amine functional group in their chemical structures. Polyamidoamine (PAMAM) dendrimers have gained popularity in scale inhibition studies due to larger surface areas of active sites, good water solubility, and better biocompatibility. Research works also reveal that PAMAM dendrimers inhibit many types of scale such as silica, silicate, as well as carbonate. Unfortunately, it has been observed in some works that PAMAM dendrimers tend to form insoluble composites with silica in water. This has called for the addition of anionic compounds in PAMAM inhibitors and their effectiveness in scale inhibition has also been enhanced because of this. Nevertheless, the enhanced PAMAM inhibitors are yet to be adopted in oilfields and treatment plants. Other than cost, the working mechanism of these enhanced PAMAM inhibitors is still not comprehensive. The optimum mass ratio and type of anionic compound are also important to the effectiveness of PAMAM dendrimers. For future oilfield applications, it is essential to research further the development of “green” scale inhibitors that have a balanced cationic-anionic property and performance on par with commercial scale inhibitors, such as phosphonates and phosphine-polyacrylates.

ACKNOWLEDGMENTS

The authors would like to extend their sincere appreciation to Yayasan Universiti Teknologi PETRONAS (YUTP 015PBC-019) for supporting this research. The authors would also like to express gratitude to the Petroleum Engineering Department and Centre of Research in Enhanced Oil Recovery for providing the resources and infrastructure. Centre of Graduate Studies.

REFERENCES

[1] Tveiten OG. (2023) Guest editorial: recycling CO_2 for EOR? Why not? J Pet Technol., 75(2):10-13. doi:10.2118/0223-0010-JPT

[2] Gbadamosi AO, Junin R, Manan MA, Agi A, Yusuff AS. (2019) An overview of chemical enhanced oil recovery: recent advances and prospects. Int Nano Lett., 9(3):171-202. doi:10.1007/s40089-019-0272-8

[3] Ragab A, Mansour EM. (2021) Enhanced oil recovery: Chemical flooding. In Geophysics and Ocean Waves Studies. Edited by Essa KS, Risio MD, Celli D, Pasquali D. London, IntechOpen; pp 1-13. doi: 10.5772/intechopen.90335

[4] Sarbast R, Salih N, Préat A. (2022) A critical overview of ASP and future perspectives of NASP in EOR of hydrocarbon reservoirs: potential application, prospects, challenges and governing mechanisms. Nanomaterials, 12(22):4007. doi:10.3390/nano12224007

[5] Wu X. (2018) Formation damage by inorganic deposition. In Formation Damage During Improved Oil Recovery: Fundamentals and Applications. Edited by Yuan B, Wood DA. Oxford, Elsevier; pp 217-242. doi: 10.1016/B978-0-12-813782-6.00005-1

[6] Sun C, Guo H, Li Y, Jiang G, Ma R. (2020) Alkali effect on alkali-surfactant-polymer (ASP) flooding enhanced oil recovery performance: two large-scale field tests' evidence. J Chem., 2020:2829565. doi:10.1155/2020/2829565

[7] Hunter K, McInnis L, Ellis-Toddington T, Kerr S. (2013) The use of modeling and monitoring to control scale in Alberta ASP floods. In Proceedings of the SPE Enhanced Oil Recovery Conference: 2-4 July 2013; Kuala Lumpur. doi:10.2118/165285-MS

[8] Mpelwa M, Tang SF. (2019) State of the art of synthetic threshold scale inhibitors for mineral scaling in the petroleum industry: a review. Pet Sci., 16(4):830-849. doi:10.1007/s12182-019-0299-5

[9] Malaie K, Shojaei O, Iranpour S, Taherkhani Z. (2021) Crystal growth inhibition of gypsum under normal conditions and high supersaturations by a copolymer of phosphino-polycarboxylic acid. Heliyon, 7(1):e06064. doi:10.1016/j.heliyon.2021.e06064

[10] Li J, Tang M, Ye Z, Chen L, Zhou Y. (2017) Scale formation and control in oil and gas fields: a review. J Dispersion Sci Technol., 38(5):661-670. doi:10.1080/01932691.2016.1185953

[11] Olajire AA. (2015) A review of oilfield scale management technology for oil and gas production. J Pet Sci Eng., 135:723-737. doi:10.1016/j.petrol.2015.09.011

[12] Fink J. (2021) Scale inhibitors. In Petroleum Engineer's Guide to Oil Field Chemicals and Fluids. 3rd edition. Edited by Fink J. Oxford, Elsevier; pp 351-391. doi: 10.1016/B978-0-323-85438-2.00007-4

[13] Yuan B, Wood DA. (2018) Overview of formation damage during improved and enhanced oil recovery. In Formation Damage During Improved Oil Recovery: Fundamentals and Applications. Edited by Yuan B, Wood DA. Oxford, Gulf Professional Publishing; pp 1-20. doi: 10.1016/B978-0-12-813782-6.00001-4

[14] Sheng JJ. (2016) Formation damage in chemical enhanced oil recovery processes. Asia-Pac J Chem Eng., 11(6):826-835. doi:10.1002/apj.2035

[15] Al Kalbani M, Jordan M, Mackay E, Sorbie K, Nghiem L. (2021) Modelling the impact of alkaline-surfactant and alkaline-surfactant-polymer flooding processes on scale precipitation and management. J Pet Sci Eng., 205:108777. doi:10.1016/j.petrol.2021.108777

[16] Saadi MA, Haddabi NA. (2019) The root of silica scale formation and its remedy. Glob J Eng Sci., 3(3):1-3. doi:10.33552/gjes.2019.03.000565

[17] Sazali RA, Sorbie KS, Boak LS. (2015) The effect of pH on silicate scaling. In Proceedings of the SPE European Formation Damage Conference and Exhibition: 3-5 June 2015; Budapest. doi:10.2118/174193-MS

[18] Lu H, Liu Y, Watson B. (2021) Silicate scale inhibitor evaluation and applications. In Proceedings of the CORROSION 2021: 19-30 April 2021; Virtual. Retrieved from https://onepetro.org/NACECORR/proceedings/CORR21/4-CORR21/D041S017R001/463741

[19] Sazali RA, Sorbie KS, Boak LS, Al Badri NS, Veny H, Hamid FHA, Abidin MZZ. (2022) Silicate scaling formation: impact of pH in high-temperature reservoir and its characterization study. Bull Chem React Eng Catal., 17(3):661-682. doi:10.9767/bcrec.17.3.15290.661-682

[20] Arensdorf J, Hoster D, McDougall D, Yuan M. (2010) Static and dynamic testing of silicate scale inhibitors. In Proceedings of the International Oil and Gas Conference and Exhibition in China: 8-10 June 2010; Beijing. doi:10.2118/132212-MS

[21] Bai Y, Bai Q. (2019) Subsea corrosion and scale. In Subsea Engineering Handbook. 2nd edition. Edited by Bai Y, Bai Q. Oxford, Elsevier; pp 455-487. doi: 10.1016/B978-0-12-812622-6.00017-8

[22] Scopus Database [https://www.scopus.com/term/analyzer.uri?sort=plf-f&src=s&sid=568bdc83411ca1d7415b595f258ca945&sot=a&sdt=a&sl=74&s=TITLE-ABS-KEY%28green+scale+inhibitor%29+AND+PUBYEAR+%3e+2012+AND+PUBYEAR+%3c+2024&origin=resultslist&count=10&analyzeResults=Analyze+results]

[23] Web of Science Database [https://www.webofscience.com/wos/woscc/analyze-results/67b0eceb-8dd5-488f-871b-2970f6a768ad-89304a51]

[24] Spinthaki A, Demadis KD. (2020) Chemical methods for scaling control. In Corrosion and Fouling Control in Desalination Industry. Edited by Saji VS, Meroufel AA, Sorour AA. Switzerland, Springer Nature; pp 307-342. doi: 10.1007/978-3-030-34284-5_15

[25] Shi F, Zhuang X, Cui C, Zhang S. (2022) Synthesis, characterization and scale inhibition performance evaluation of novel dendrimers with the initiator core of pentaerythritol derivative. Desalination, 528:115632. doi:10.1016/j.desal.2022.115632

[26] Sun Y, Yin X, Chen Z, Yang W, Chen Y, Liu Y, Zuo Y, Li L. (2021) Use of polyaminoamide dendrimers starting from different core-initial molecules for inhibition of silica scale: experiment and theory. Colloids Surf A Physicochem Eng Asp., 613:126095. doi:10.1016/j.colsurfa.2020.126095

[27] Tan M, Fang L, Zhang B, Zhao Y, Meng X, Li F. (2021) Synergistic inhibition effect and mechanism of polycation and polyanion on colloidal silica. Colloids Surf A Physicochem Eng Asp., 610:125701. doi:10.1016/j.colsurfa.2020.125701

[28] Hayati N, Humaida H, Siswanta D. (2021) Adsorption of silicate anions from geothermal brine using chitosan-polyethylene glycol composite to prevent silica scaling on the Dieng Geo Dipa geothermal energy system. Indones J Chem., 21(5):1063-1071. doi:10.22146/ijc.49248

[29] Bai S, Han J, Du C, Ding W. (2019) Selective removal of silicic acid by a gallic-acid modified resin. J Water Reuse Desalin., 9(4):431-441. doi:10.2166/wrd.2019.030

[30] Spinthaki A, Kamaratou M, Skordalou G, Petratos G, Tramaux A, David G, Demadis KD. (2021) A universal scale inhibitor: a dual inhibition/dispersion performance evaluation under difficult brine stresses. Geothermics, 89:101972. doi:10.1016/j.geothermics.2020.101972

[31] Spinthaki A, Kamaratou M, Skordalou G, Petratos G, Petrou I, Tramaux A, David G, Demadis KD. (2021) Searching for a universal scale inhibitor: a multi-scale approach towards inhibitor efficiency. Geothermics, 89:101954. doi:10.1016/j.geothermics.2020.101954

[32] Neofotistou E, Demadis KD. (2004) Silica scale inhibition by polyaminoamide STARBURST® dendrimers. Colloids Surf A Physicochem Eng Asp., 242(1-3):213-216. doi:10.1016/j.colsurfa.2004.04.067

[33] Neofotistou E, Demadis KD. (2004) Use of antiscalants for mitigation of silica (SiO_2) fouling and deposition: fundamentals and applications in desalination systems. Desalination, 167:257-272. doi:10.1016/j.desal.2004.06.135

[34] Mavredaki E, Neofotistou E, Demadis KD. (2005) Inhibition and dissolution as dual mitigation approaches for colloidal silica fouling and deposition in process water systems: functional synergies. Ind Eng Chem Res., 44(17):7019-7026. doi:10.1021/ie0501982

[35] Demadis KD. (2005) A structure/function study of polyaminoamide dendrimers as silica scale growth inhibitors. J Chem Technol Biotechnol., 80(6):630-640. doi:10.1002/jctb.1242

[36] Rajeswari A, Amalraj A, Pius A. (2015) Removal of phosphate using chitosan-polymer composites. J Environ Chem Eng., 3(4):2331-2341. doi:10.1016/j.jece.2015.08.022

[37] Ren B, Wang K, Zhang B, Li H, Niu Y, Chen H, Yang Z, Li X, Zhang H. (2019) Adsorption behavior of PAMAM dendrimers functionalized silica for Cd(II) from aqueous solution: experimental and theoretical calculation. J Taiwan Inst Chem Eng., 101:80-91. doi:10.1016/j.jtice.2019.04.037

[38] Sohail I, Bhatti IA, Ashar A, Sarim FM, Mohsin M, Naveed R, Yasir M, Iqbal M, Nazir A. (2020) Polyamidoamine (PAMAM) dendrimers synthesis, characterization and adsorptive removal of nickel ions from aqueous solution. J Mater Res Technol., 9(1):498-506. doi:10.1016/j.jmrt.2019.10.079

[39] Mahat SQA, Saaid IM. (2017) Development of anionic-cationic inhibitors for mitigating silicate scales during ASP flooding. J Pet Sci Eng., 149:701-706. doi:10.1016/j.petrol.2016.10.061

[40] Chen F, Wu K, Liu Y, Huang H, He K. (2014) Polyaminoamide dendrimers surface-modified with anionic terminal groups for use as calcium carbonate scale inhibitors. J Polym Eng., 34(8):733-738. doi:10.1515/polyeng-2014-0130

[41] Demadis KD, Neofotistou E. (2007) Synergistic effects of combinations of cationic polyaminoamide dendrimers/anionic polyelectrolytes on amorphous silica formation: a bioinspired approach. Chem Mater., 19(3):581-587. doi:10.1021/cm062370d

https://doi.org/10.31436/iiumej.v25i2.2957

BIOHYDROGEN PRODUCTION IN SEMI-CONTINUOUS SYSTEM USING IMMOBILIZED CELL MEMBRANE

NURUL SAKINAH ENGLIMAN[1], AINUL HUSNA ABDUL AZIZ[1], PEER MOHAMED ABDUL[2], ABDULLAH AMRU INDERA LUTHFI[2], JAMALIAH MD JAHIM[2], NUR SYAKINA JAMALI[3]

[1]Department of Chemical Engineering and Sustainability, Kulliyyah of Engineering, International Islamic University. Malaysia (IIUM), 53100 Kuala Lumpur, Malaysia
[2]Research Centre for Sustainable Process Technology (CESPRO), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), 436000 Selangor, Malaysia
[3] Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia

**Corresponding author: sakinahengliman@iium.edu.my*

(Received: 9 August 2023; Accepted: 10 May 2024; Published online: 15 July 2024)

ABSTRACT: Hydrogen is considered to be the fuel of the future because of its high energy content (122 kJ/g), and water is the only byproduct of its use. Moreover, the production of hydrogen via fermentation of organic wastes is carbon neutral. This study was conducted to evaluate the performance of immobilized cells on PVDF membrane for biohydrogen production using a sequencing batch reactor by varying the hydraulic retention times (HRT) of the system and to compare the efficiency between suspended and attached systems on the production of biohydrogen. It was found that the biohydrogen fermentation performance was improved in a semi-continuous system, especially with immobilized cells. The optimum HRT that supports the highest biohydrogen yield was for an HRT of 12 hours, where the performance of hydrogen production was improved and in which the maximum hydrogen yield was achieved at 2.43 mol H_2/mol and maximum hydrogen production rate (HPR) of 2.46 L H_2/L.d as compared to other HRT for both systems. Therefore, the result of this study can be applied as the benchmark for scaling up the process.

ABSTRAK: Hidrogen boleh dianggap sebagai sumber tenaga penting pada masa hadapan kerana kapasiti tenaga yang tinggi (122 kJ/g) dan hanya air terhasil dari tindak balas hidrogen. Tambahan, sisa pengeluaran hidrogen melalui proses fermentasi sisa organik adalah bersifat semula jadi. Kajian ini dijalankan bagi mengkaji prestasi sel tidak bergerak pada membran PVDF bagi penghasilan biohidrogen menggunakan reaktor kelompok turutan dengan mengubah sistem masa pengekalan hidraulik (HRT) dan dengan membuat perbandingan kecekapan antara sistem yang tergantung dan sistem yang bersambung pada penghasilan biohidrogen. Dapatan kajian mendapati prestasi fermentasi diperbaharui di bawah sistem separa turutan terutama dengan sel tidak bergerak. Nilai optimum HRT yang mempunyai hasil biohidrogen tertinggi adalah pada ketika HRT 12 jam di mana prestasi penghasilan hidrogen dapat diperbaharui dan menghasilkan hidrogen tertinggi pada 2.43 mol H2/mol dan kadar penghasilan hidrogen maksimum (HPR) pada 2.46 L H2/L.d berbanding sistem HRT lain pada kedua-dua sistem. Oleh itu, dapatan kajian ini boleh digunakan sebagai penanda aras bagi kenaikan proses.

KEYWORDS: *biohydrogen, immobilized cell membrane, semi-continuous system*

1. INTRODUCTION

Hydrogen is considered the future's alternative fuel to replace fossil fuels. It offers a lot of possibilities to generate a valuable renewable energy carrier that is cleaner and has high-energy content. Biological hydrogen production is more favorable as compared to conventional physico-chemical methods of hydrogen production such as steam reforming. This is due to its potential as a carbon-neutral and cost-effective fuel that is inexhaustible and versatile, as it can use various types of carbon sources for biological hydrogen production [1-3]. Besides, biological hydrogen production is more feasible for practical application, which includes the integration with fuel cell technologies, since it can produce high rates of hydrogen at lower capital cost [4, 5].

Biological hydrogen production can be categorized into four processes: photo fermentation, dark fermentation, bio-photolysis, and sequential fermentation (an integrated process between dark and photo fermentation) commonly used in hydrogen generation [6]. However, dark or anaerobic fermentation is more favorable as it is much simpler and inexpensive, which makes it suitable for small-scale production facilities [7, 8]. It has been proven that both pure cultures and mixed cultures can be used for dark fermentation, and they are able to utilize a wide range of carbon sources, including waste [9, 10].

Despite its great potential, biological hydrogen production still cannot compete with the thermochemical processes of generating hydrogen as the latter process is more stable and reliable to support the growing demand for energy [11]. Thus, current research on biological hydrogen production, particularly anaerobic processes, has focused on improving the hydrogen conversion and volumetric production rates. One of the attractive ways to achieve a stable and high conversion rate is through the immobilization processes of hydrogen-producing bacteria. A previous study found that immobilized cells are able to increase hydrogen production by up to 20% compared to free cells or suspended cells [13]. Similarly, the stability of hydrogen production with higher hydrogen production was also proven with the increasing value between 15 to 20% when microbial cells were immobilized on the support material compared to suspended cells [3,13,14]. This is because immobilization helps retain the bacteria inside the reactor and prevents it from washing out, especially in a continuous system as compared to a suspended cell system [13,15,24].

A continuous system is preferable to a batch system since it generates a continuous and stable production of hydrogen gas and, thus, improves productivity over batch systems [15,17]. Basically, there are two types of reactors that are commonly used for continuous and semi-continuous hydrogen production. The first type is the suspended system, where the bacteria is suspended and might form bacterial flocks in the bottom of the reactor. Mixing is needed to achieve a homogenous system and ensure that the substrate is well-distributed throughout the reactor. The second type is attached systems, in which the bacteria grow on the surface of support material and form biofilm. The bacteria will remain retained in the system during the effluent's flushing-out process and thus maintain the active biomass inside the reactor.

Various modes of operation have been used for fermentative hydrogen production, including batch, semi-continuous, and continuous systems. It is known that batch hydrogen fermentation will produce lower hydrogen production rates than continuous or semi-continuous modes. Semi-continuous mode, with proper dilution rates, can enhance the fermentation rate. Valdez-Vazquez et al. [18] found that when the dilution rate was controlled during the fermentation of municipal solid waste, the degradation of non-degradable waste improved, and in the meantime, biogas production increased. Lin and Chou [17] found that a sequencing batch reactor run at a hydraulic retention time (HRT) of 8 hours with the organic

loading rate (OLR) of 80 kg COD/l was able to produce 2.8 mol of hydrogen (39 mmol H2/ kg COD/d). They concluded that the hydrogenic activity of mixed bacteria was HRT dependent, with the lower HRT having the effect of deteriorating hydrogen productivity. In addition, control of pH was important to ensure a stable operation of the reactor, as pH can be an indicator for monitoring bacterial activity. The advantages of sequencing batch reactors include greater biomass retention (hence the ability to decouple solids retention time (SRT) from HRT), a higher degree of process flexibility with respect to changes in organic loading rate (OLR), relative ease of operation and lower capital investment. The semi-continuous mode operation of this process is also considered more feasible for potential real applications and commercialization [15, 20].

Nowadays, most fermentative hydrogen production is conducted continuously. Continuous stirred tank reactors (CSTR) are widely used for fermentative hydrogen production. Bacteria inside the CSTR accumulate inside the reactor to form a suspension and reduce the mass transfer resistance as compared to the batch system. However, HRT still plays a crucial factor in CSTR operations. Since higher OLR will reduce the efficiency of the fermentation process inside the CSTR, suitable HRT needs to be determined, or biomass washout will deteriorate the hydrogen production rates [18-21].

Later, in order to achieve higher cell density retained inside the reactor, an immobilized system was developed either through entrapment [23,24] or attachment [13,24]. The studies found that cells immobilized were able to increase the hydrogen production rate with decreasing HRT, which would not be achievable under the suspended CSTR system. Cell immobilization on support material improved mass transfer efficiency and cell stability. A longer operation time will allow the formation of granules, where the bacteria formed a dense flock that settled in the bottom of the reactor, reducing the cell washout problem while improving the substrate conversion to hydrogen [25]. The operational conditions, such as pH and HRT, still play an important role in the process, especially when the reactor needs to be maintained for a long time in order to achieve process stability, which can be a major economic barrier in terms of maintenance [26,27].

Even though the continuous system promotes a better hydrogen production process due to process stability, high organic content, and ability to run at shorter HRT, the performance of each mode of operation is still dependent on other factors such as type of substrate used, type of reactor, and operational condition selected. However, in order to develop a new fermentation process, such as the integrated fermentation process, it is found that fermentative hydrogen production via batch system, particularly by sequencing batch mode, offered some advantages, such as better control of the microbial population due to the cyclic operation [28, 29]. Thus, the objective of this study was to evaluate the potential use of immobilized cells on PVDF membrane for energy production using a sequencing batch reactor. The influence of hydraulic retention times (HRT) (36, 24, and 12 h) and the comparison between suspended and attached systems on the production of hydrogen were especially investigated.

2. METHODOLOGY

2.1 Inoculum Preparation

The immobilized membrane with bacterial cells attached to it was developed with some modifications from Engliman et al. [5]. The membrane underwent the acclimatization step using an HRT of 2 days via repeated batch cultivation in a 500ml modified bottle reactor. Nitrogen gas was purged periodically to maintain the anaerobic condition inside the reactor.

https://doi.org/10.31436/iiumej.v25i2.2957

The pH of the fermentation medium was only set initially at pH 5.5, and the seed was cultivated at 55℃ inside the shaking water bath at 150 rpm. The microbial cell was acclimatized under sequencing batch mode, and the experiments were run at an HRT of 2 days with a 500 ml working volume. Sequencing batch operation was carried out by the removal of 50% of the culture medium and replacement with a new medium after each cycle. The process cycle was continued until the hydrogen production rate achieved consistency. Samples of fermentation products were taken during each cycle and sent for analysis. The seed was acclimatized until a stable and consistent hydrogen production was achieved before being applied to the semi-continuous reactor.

2.2 Fermentation Medium

The synthetic media was used as a fermentation medium in this study. The growth medium used in this study consisted of glucose (10 g/l) as the main carbon source and was supplemented with other nutrients such as yeast extract (3 g/l), peptone (5 g/l), sodium chloride (1 g/l), sodium acetate (1 g/l) and Wolfe's mineral solution (1ml/l). This growth medium will be used throughout the study of the effect of HRT in sequential batch systems.

2.3 Experimental Set-Up for Semi-Continuous Fermentation Systems

Fig. 1 shows the schematic diagram of the semi-continuous system used in this study. The fermentation experiments were carried out using 400ml synthetic medium as a feedstock, with 15% inoculum seed added into a 500ml modified Scott Duran bottle. The temperature was controlled at 60ºC using a shaking water bath. The pH was only maintained initially at pH 5.5 using 1M NaOH and 1M HCl. The total volume of gas collected was measured by using a gas meter (MiliGascounter, Ritter, Germany). Another tube was connected to the cylinder for biogas sampling and gas composition analysis.

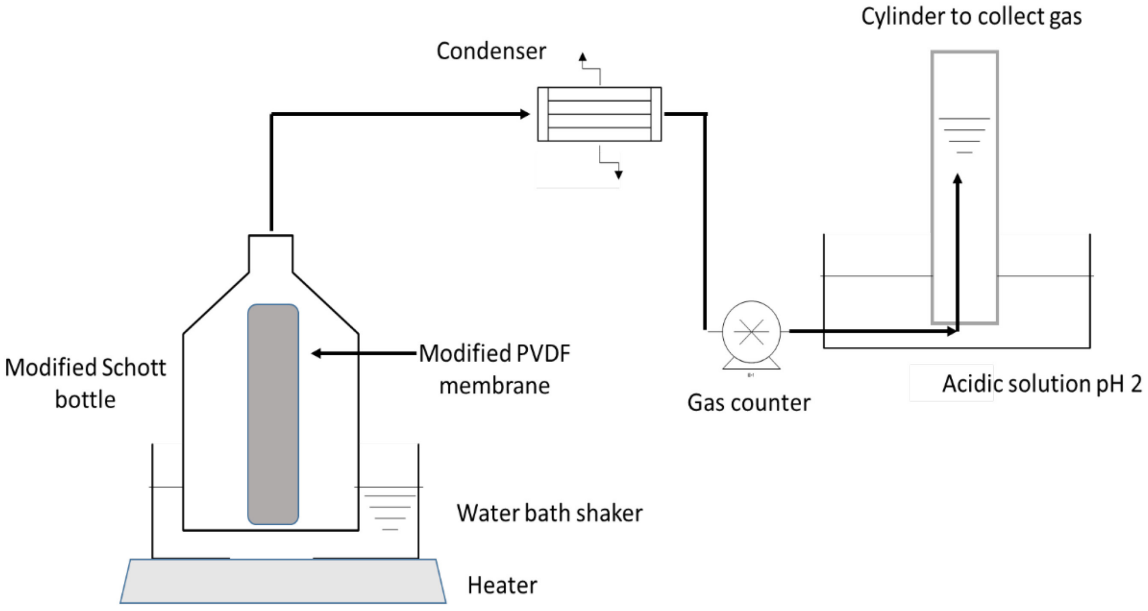


Fig. 1. Schematic diagram of the experimental set-up for this study

2.4 Experimental Operation

The experiments were operated in sequential batch mode using HRT of 72, 36, and 24 hours, which corresponded with the feed flow rate of 167, 334, and 500 ml per day, respectively. The influent feed consisted of 10 g/l glucose that was purged with nitrogen

gasses for 15 minutes to ensure an anaerobic condition inside the reactor. The feeding cycle was calculated using Eq. (1) according to the respective HRT (Table 1). The analyses were conducted in triplicate to ensure the consistency of the experiment.

$$Feeding\ Cycle\ \left(\frac{cycle}{day}\right) = \frac{24\ hour\ per\ day \times 1\ cycle}{HRT\ (h)} \quad (1)$$

Table 1. Operational parameters involved in this study

HRT (hr)	Feeding cycle (cycle/day)
72	3
36	2
24	1
12	0.5

2.5 Hydrogen Concentration Analysis

The percentage of hydrogen and carbon dioxide present in the evolved gas was determined using gas chromatography. The gas chromatograph (model SRI 8600C, USA) consisted of two detectors: a helium ionization detector (HID) and a thermal conductivity detector (TCD). High purity helium gas (99.99%) was used as the carrier gas at 25 mL/min. A volume of 1 mL of the gas samples was taken using a 1-mL gas-tight syringe, and the samples were injected into the GC immediately at a temperature of 43ºC and a pressure of 2.7 psi initially, then followed by a ramping of 30ºC per minute, which was maintained for 10 min once the temperature reached 220ºC [5].

2.6 Energy Analysis

The hydrogen energy production rate (EPR), KJ/L/d was calculated using Eq. (2) [30].

$$EPR = \frac{HPR}{22.4} \times HVH2 \quad (2)$$

where HPR is the hydrogen production rate (L/L/d) calculated from the experimental result, and $HVH2$ is the heating value of hydrogen (286 J/mmol).

3. RESULTS AND DISCUSSION

3.1 Effect of HRT on Immobilized Membranes Compared to Suspended Systems

The modified bottle reactors with a working volume 500ml were operated via sequential batch mode for fermentative hydrogen production. The effects of hydraulic retention time (HRT) on biogas production rate, biogas composition, and soluble metabolites produced were evaluated throughout the fermentation process. Four different HRTs, including 72, 36, 24, and 12 hours, were evaluated in order to determine the stability of the continuous process using the designated system. The experiment was carried out at 60ºC with the initial pH of the fermentation medium being set at pH 5.5 and a sugar concentration of 10g/l. The experiment was started with acclimatization of the seed before changing the HRT value. The HRT was changed once the daily biogas and biohydrogen production reached a pseudo-steady state. The feeding strategy used for this purpose is shown in Table 1.

https://doi.org/10.31436/iiumej.v25i2.2957

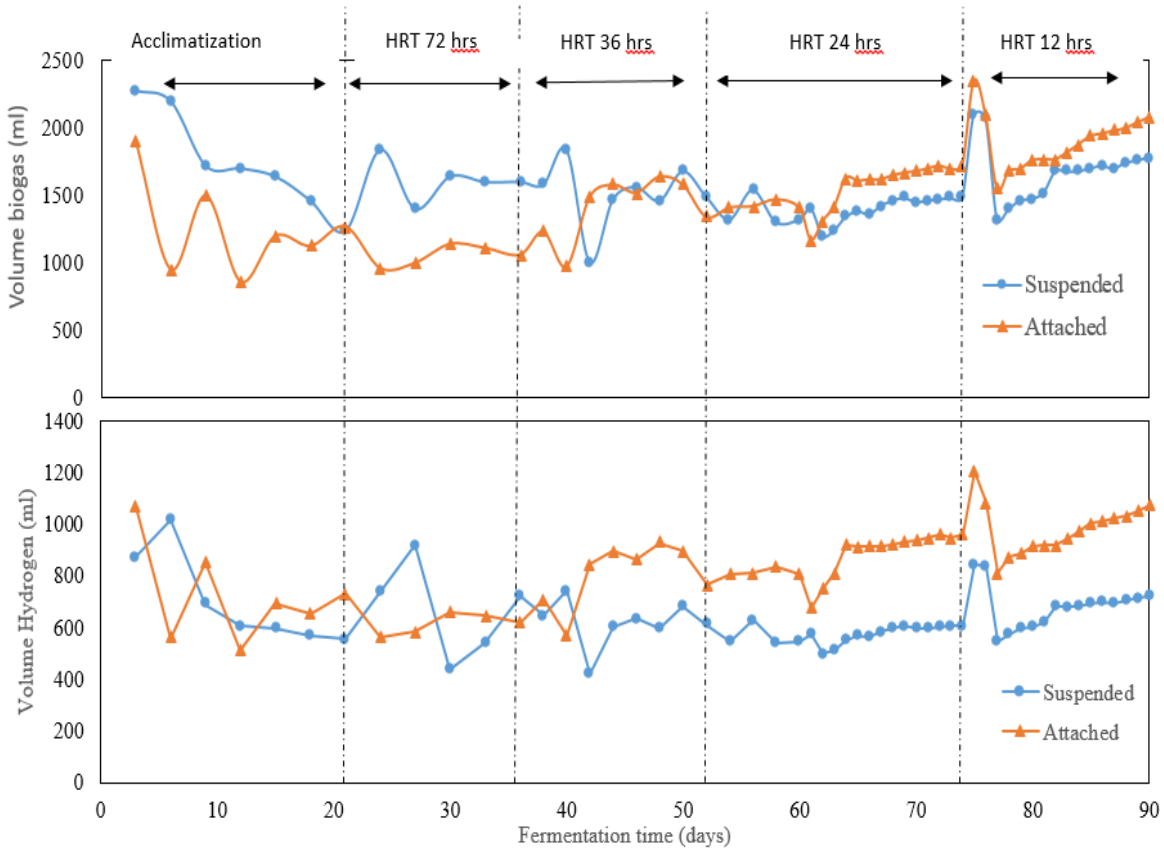


Fig. 2. Biogas and hydrogen production of two different systems when different HRTs were used (HRT of 72, 36, 24, and 12 hours, respectively)

Fig. 2 shows the biogas and hydrogen produced for two different systems: the suspended system, where bacteria was left to be suspended inside the reactor, and the attached system, in which the bacteria was immobilized on the PVDF membrane. It was found that during the early stage of each HRT value, the biogas and biohydrogen produced fluctuated due to the shift in dominant bacteria governing the fermentation process. After an average of three feeding cycles, the biogas and biohydrogen production gradually reached a quasi-steady state with negligible difference from previous cycles. The maximum hydrogen produced was found during the HRT of 12 hours, which was about 1035 ± 26 ml of hydrogen produced daily with the composition of 48% of hydrogen for the attached system. For the suspended system, the HRT of 12 hours showed better hydrogen production than other HRT, with the average daily production during HRT 12 hours being 706 ± 13 ml of hydrogen (average hydrogen composition was 42%). The trend shows that as the HRT became shorter, the hydrogen produced increased up to 10% from the previous HRT. However, the concentration of hydrogen produced in the biogas was not affected significantly by the changes in HRT, still reaching quite high values in a range of 37 – 48% when HRT was decreased from 72 hours to 12 hours (Table 2).

The performance of both systems was evaluated based on hydrogen production rate (HPR) and hydrogen yield (HY). The results showed that the were no significant differences in HPR during HRT 72 and 36 hours, but the HPR started to improve during the middle of HRT 24 hours and kept increasing at HRT 12 hours (Fig. 3). A similar trend was observed from both systems, but attached system shows better HPR during HRT 12 hours, which is about 2.68 L of H2/L of influent/day. However, the hydrogen yield (HY) for all HRT showed little

difference between the suspended and attached systems, which might be due to the reduction of sugar utilization by the bacteria. The highest hydrogen yields for both systems were achieved during late HRT 12 hours, which were 1.41 mol H2/mol glucose for the suspended and 2.43 mol H2/mol glucose for the attached system. Thus, the attached system with bacteria immobilized on the surface of the PVDF membrane showed better performance during the fermentative hydrogen production under thermophilic conditions. The designated immobilization system proved that it can be used to improve hydrogen production in continuous systems.

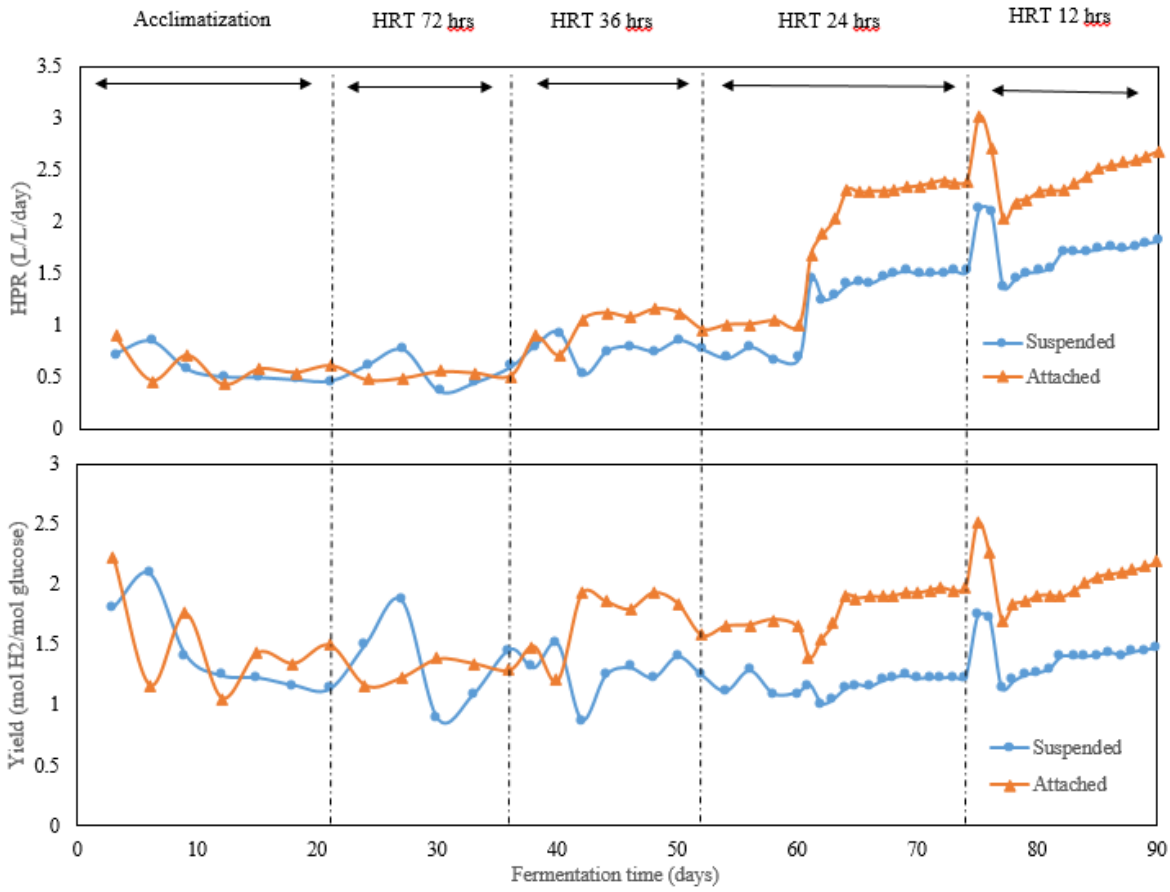


Fig. 3. Performance of suspended and attached systems under different HRT values.

Even though the immobilization system might have a limitation in terms of mass transfer within the system, this limitation can be overcome by modification of support materials used for immobilization systems. The modification of support materials can contribute to improved surface roughness, porosity, and hydrophobicity of the support materials [14]. In this study, the PVDF membrane used had already undergone surface modification that improved its performance to promote better cell immobilization as well as biofilm formation [5,13]. Biofilm formation produced along a cell-surface/interface attachment, known as extracellular polymeric substances (EPS), was responsible for keeping cells together in a three-dimensional structure. EPS mediates in mass transfer through biofilm adsorption of xenobiotics, metals, or inorganic ions and provides physical support for the formation of biofilm [21,23]. The biofilm helped sustain cell viability which would prevent the cells being washed out from the system, thereby enhancing cell density as well as encouraging a consistent biohydrogen production in the system. Similar observations were also obtained from previous studies done by Lutpi et al. [35] and Jamali et al. [24] using granulated activated carbon (GAC) as support materials for cell immobilization to improve hydrogen

production in which the highest hydrogen yield of 5.6 mol H2/ mol substrate and maximum hydrogen production rate of 4.3 mmol H2/l.h and 1.77 mol H2/mol substrate consumed and hydrogen production rate (HPR) of 2.0 mmol H2/l.h were achieved via cells immobilization system respectively.

3.2 Effect of HRT on pH and Sugar Consumption

Later, the effects of HRT on pH and sugar utilization were observed. As summarized in Tables 2a and 2b, it was found that the final pH for all HRT in both systems was in a range of 4.86 to 5.07. Surprisingly, the lowest final pH resulted in the highest hydrogen production rate for the system. This can be seen during the HRT 12 hours for both systems, the final pH dropped to pH 4.9 for the suspended system, but the HPR increased up to 1.71 liters of hydrogen per liter influent per day. Similarly, for the attached system, where the pH dropped to 4.86, but the HPR value improved to 2.46 liters of hydrogen per liter of influent per day. A previous study by van Ginkel et al. [26] also observed similar results and reported that the maximum hydrogen production was achieved as long as the pH depletion did not occur drastically so as to inhibit the whole process.

The results also showed that at the decreased HRT of as low as 12 hours, the accumulation of total volatile fatty acids was increased, and the pH value was dropped. This trend was similar for both systems, and this is probably due to the fermentation process having already reached an unstable state. The depletion of pH value is the biggest indicator for the instability of the process. Furthermore, the sudden pH drop will cause a slow reactor recovery rate and may lead to system failure [27]. However, since the initial pH selected for the process was the optimum value to support the growth of the bacteria, the bacteria were able to adapt to the environmental changes slowly, and thus, no rapid acid by-product was produced. A rapid drop in fermentation pH indicates that the hydrogen was produced rapidly, causing rapid production of acid by-products that increased to inhibitory levels and, thus, spontaneously depleted the buffering capacity [26,27].

On the other hand, the decrease in HRT value affected the sugar uptake by the microbes. The initial glucose concentration was fixed at 10 g/l, and it was found that as the HRT decreased, the substrate conversion increased, but this trend was applicable only for the attached system, whereas for the suspended system, the substrate conversion decreased when the HRT decreased from 72 hours to 12 hours. This situation happened due to the stability and adaptation of the bacterial community inside the system. Since the bacteria was already stable inside the system, it was able to adapt and maintain itself against any changes in hydraulic dilution due to the decreased HRT. However, to conclude this statement, it is better to relate the event with the hydrogen yield and final biomass concentration [28,29]. For the suspended system, hydrogen yield decreased with decreasing HRT, like the biomass concentration (VSS, g/l) (Table 2a), which might be because of the rapid loss of bacterial cells due to dilution when the HRT was shorter. However, for the attached system, where bacteria were immobilized on the PVDF membrane surface, the hydrogen yield was improved when the HRT was shortened to 12 hours, while the bacterial composition (VSS, g/l) was still high inside the reactor (Table 2b). The retention of bacterial cells inside the reactor improved the substrate utilization rate, and thus, more hydrogen could be produced in this designated system. The shorter HRT was not so efficient when applied to the suspended system, as the bacterial growth was unable to compete with the dilution effect due to the rapid rate of feeding, which led to cell washout. Therefore, among these two systems, the best substrate utilization rate was achieved at HRT 12 hours for the attached system, which is about 91% sugar conversion with a hydrogen yield of 2.48 mol H2/mol glucose.

https://doi.org/10.31436/iiumej.v25i2.2957

Table 2: a) Fermentation parameters in various HRT for the suspended system

HRT (hrs)	H_2 content (%)	Final pH	VSS (g/l)	Substrate Consumption (%)	Yield (mol H_2/mol glucose)	Energy production rate, EPR (KJ/L/d)	HPR (L/L/d)
72	37	4.97 ± 0.13	3.52 ± 0.33	89 ± 0.7	1.21 ± 0.36	7.35 ± 1.8	0.58 ± 0.14
36	40	5.07 ± 0.11	3.30 ± 0.24	89 ± 0.6	1.53 ± 0.16	9.61 ± 1.3	0.75 ± 0.10
24	40	4.98 ± 0.13	3.16 ± 0.11	89 ± 0.5	1.88 ± 0.07	18.45 ± 1.1	1.45 ± 0.09
12	42	4.90 ± 0.18	2.95 ± 0.18	84 ± 0.8	1.62 ± 0.16	21.81 ± 2.6	1.71 ± 0.21
HRT (hrs)	**Volatile fatty acids (VFAs)**				**Soluble Metabolites molar ratio**		
	HAc (mM)	**HBu (mM)**	**HPr (mM)**	**HFr (mM)**	**B/A**	**TVFA (mM)**	
72	20.9 ± 0.18	11.4 ± 0.14	8.8 ± 0.37	7.6 ± 0.51	0.65	48.7 ± 1.2	
36	21.2 ± 0.22	10.2 ± 0.31	12.5 ± 0.11	9.0 ± 0.08	0.48	52.9 ± 0.72	
24	21.9 ± 0.09	12.0 ± 0.41	10.3 ± 0.21	10.7 ± 0.15	0.43	54.9 ± 0.86	
12	26.2 ± 0.70	12.7 ± 0.12	11.7 ± 0.08	7.6 ± 0.28	0.49	58.2 ± 0.93	

Table 2: b) Fermentation parameters in various HRT for the attached system (immobilized PVDF membrane)

HRT (hrs)	H_2 content (%)	Final pH	VSS (g/l)	Substrate Consumption (%)	Yield (mol H_2/mol glucose)	Energy production rate, EPR (kJ/L/d)	HPR (L/L/d)
72	39	5.00 ± 0.09	3.96 ± 0.22	88 ± 0.6	1.41 ± 0.32	7.24 ± 1.6	0.57 ± 0.13
36	41	4.92 ± 0.05	4.11 ± 0.41	87 ± 2.4	1.70 ± 0.21	13.00 ± 1.5	1.02 ± 0.12
24	45	4.94 ± 0.09	4.08 ± 0.33	89 ± 0.3	1.85 ± 0.18	28.60 ± 2.7	2.24 ± 0.21
12	48	4.86 ± 0.17	3.50 ± 0.27	91 ± 0.9	2.48 ± 0.16	31.45 ± 3.1	2.46 ± 0.24
HRT (hrs)	**Volatile fatty acids (VFAs)**				**Soluble Metabolites molar ratio**		
	HAc (mM)	**HBu (mM)**	**HPr (mM)**	**HFr (mM)**	**B/A**	**TVFA (mM)**	
72	22.9 ± 0.27	9.7 ± 0.24	9.1 ± 0.15	8.6 ± 0.20	0.42	50.3 ± 0.86	
36	24.3 ± 0.17	10.5 ± 0.06	8.2 ± 0.12	9.2 ± 0.09	0.43	52.2 ± 0.44	
24	23.9 ± 0.07	12.4 ± 0.21	10.4 ± 0.24	7.1 ± 0.22	0.52	53.8 ± 0.74	
12	27.7 ± 0.24	16.3 ± 0.33	9.1 ± 0.11	6.4 ± 0.16	0.59	59.5 ± 0.84	

https://doi.org/10.31436/iiumej.v25i2.2957

Energy production analysis shows that the amount of energy produced was increased proportionally with the increase of HPR (Table 2). The energy generated for the optimum HRT of 12 hours for both systems was about 21.81 KJ H2/L/d for the suspended and 31.45 KJ H2/L/d for the attached system. This indicates that both systems achieved efficient hydrogen production with stable biomass retained inside the reactor, even though the suspended system suffered a small decrease in hydrogen yield under the high dilution rate.

3.3 Effect of HRT on Metabolites Production

Every fermentation process not only produced the desired product, but it also produced biproducts which might subsequently inhibit the process. It was found that the major metabolites produced from these two systems were similar, with acetate and butyrate being the dominant metabolites. Surprisingly, there was no production of ethanol in both reactors, but there was the appearance of a small amount of formic acid in the fermentation effluent. This is probably due to the catalytic hydrogenation of carbon dioxide that accumulated inside the reactor, especially when the fermentation achieved a homogenous state [31]. Besides that, as the initial bacterial seeds used for this study contained a mixed culture of bacteria from the effluent of the CSTR reactor of POME. Cyanobacteria species were probably present inside the seed. Cyanobacteria can oxidize sugar to form formic acid [33].

In Table 2, the total volatile fatty acids (TVFA) increased as HRT decreased from 72 hours to 12 hours, and the patterns were similar for both suspended and attached systems. Within the production of major acids in hydrogen fermentation, acetate and butyrate contributed about 40 – 45% and 19 – 23%, respectively, for the suspended system, while about 44 – 50% of acetate and 19 – 23% of butyrate were produced for the attached system. Ethanol was not detected in any effluent of different HRT for both systems. The high production of acetate and butyrate without any production of ethanol indicates that the system was able to produce hydrogen efficiently, as both acids were correlated with hydrogen production [31,32,34].

The B/A ratio (HAc/HBu) is one of the indicators for hydrogen fermentation, especially in anaerobic processes. Production of acetate and butyrate will always be accompanied by hydrogen generation, as shown in Eq. 3 and 4 [34].

$$C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 2CO_2 + 4H_2 \quad (3)$$

$$C_6H_{12}O_6 \rightarrow C_3H_7COOH + 2CO_2 + 2H_2 \quad (4)$$

These equations show that the production of acetic acid led to the creation of more hydrogen than butyric acid. Thus, a lower butyric to acetic acid (B/A) ratio was desired for higher hydrogen yield and productivity [36,37]. At optimum HRT, HRT 12 hours, about 26.2 mmol/L acetate and 12.7 mmol butyrate were produced in a suspended system with a B/A ratio of 0.49 accompanied by high HPR and yield. Meanwhile, for the attached system, about 27.7 mmol/L acetate and 16.3 mmol/L butyrate available in the effluent with the B/A ratio of 0.59 resulted in the maximum hydrogen yield and productivity.

Even though the major metabolite produced was the main metabolite for hydrogen production, the accumulation of propionic acid may lead to lower hydrogen. This is because some of the substrates might consume the accumulated hydrogen produced for further reduction to propionic acids (Equation 5). This happened when the condition inside the reactor suffered a low pH, high hydrogen partial pressure, and imbalance in the catabolic reaction of bacteria involved in the NAD+/NADH ratio [33,34]. Increasing NADH level during the fermentation led to butyric acid formation; in order to balance the NAD+/NADH ratio, propionic acid was spontaneously produced to generate enough NAD+, thus resulting in

https://doi.org/10.31436/iiumej.v25i2.2957

a shift in the metabolic pathway as the bacteria inside the reactor needed to adapt to the changes happening around them [33]. Therefore, it is suggested to constantly remove any forms of stressors from the system to ensure that the high yield and productivity of the fermentation process can be maintained.

$$C_6H_{12}O_6 + 2H_2 \rightarrow 2CH_5COOH + 2H_2O \quad (5)$$

The soluble metabolites ratios can serve as qualitative indicators of substrate metabolism as well as metabolic pathways involved in fermentative hydrogen production. The linear correlation between the B/A ratio with respect to HPR and hydrogen yield is illustrated in Fig. 4.

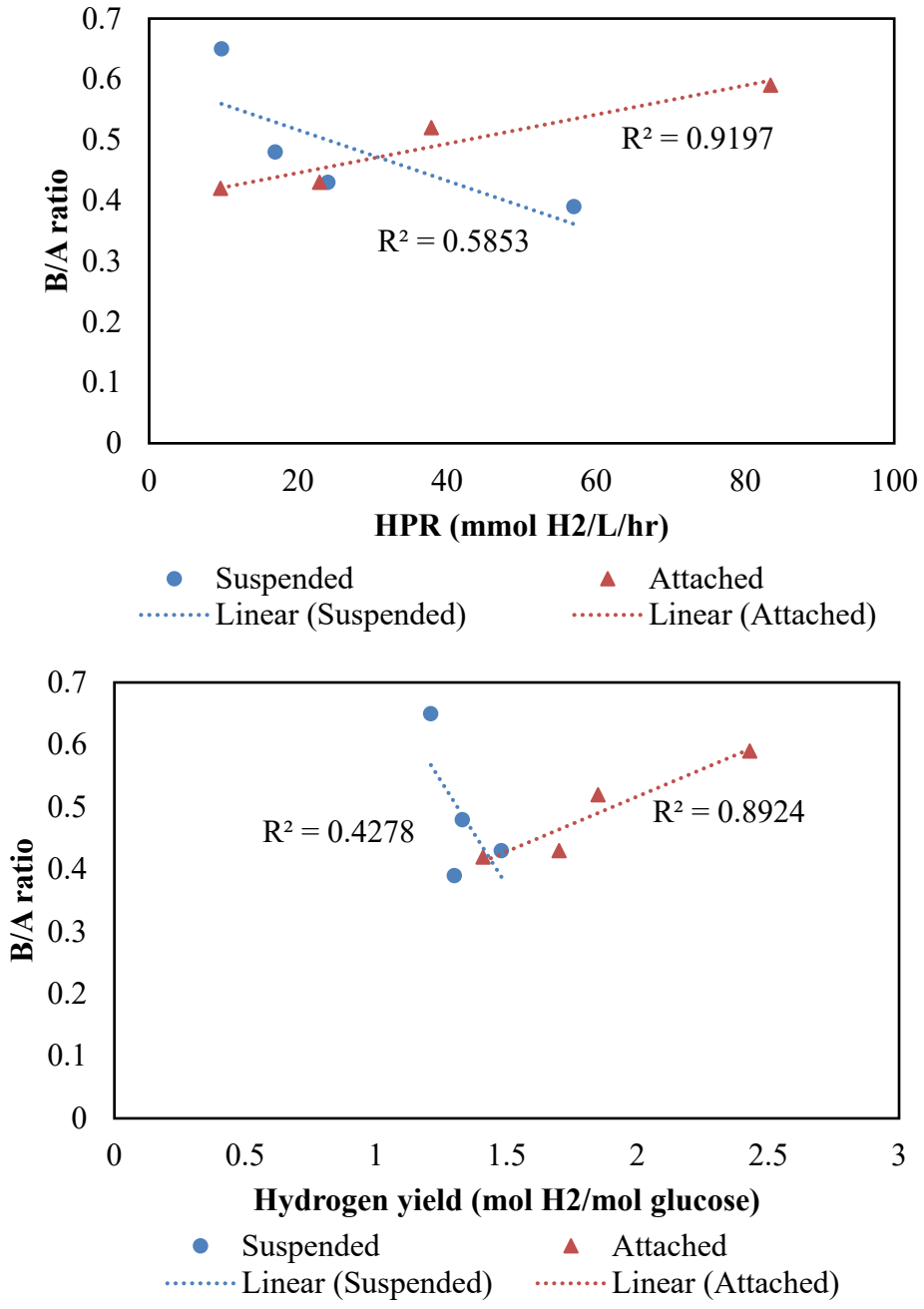


Fig. 4: Linear correlation between B/A ratio with respect to HPR and hydrogen yield

The attached system showed a good linear correlation with respect to HPR and yield with R2 = 0.92 and 0.89, respectively. This indicates that the increment of the B/A ratio near 1 showed that the metabolic pathways favored the acetic pathways for the attached system. However, for the suspended system, the decrement of the linear correlation showed that the production of hydrogen was not so efficient, which correlated with the smallest hydrogen yield and HPR. Therefore, it is proven that the attached system provided better bacterial retention and stability of the fermentation process, which was able to improve the fermentation process and, hence, increase yield and productivity.

4. CONCLUSION

To sum up, it was found that the biohydrogen fermentation performance was improved under a semi-continuous system, especially with immobilized cells. From the observation, it was found that the immobilized system promotes better performance than the suspended system. This is due to immobilized cell technology offering numerous advantages, including higher cell density, higher metabolic activity, prevention of interfacial inactivation, better productivity, and protection against changing environmental conditions. Under the immobilization system, the hydrogen yield increased to 54%, while HPR increased to 64% when compared with the suspended system. For all systems, it was found that the optimum HRT was during HRT 12 hours, where the performance of hydrogen production was improved as compared to other HRTs and in which the maximum hydrogen yield was achieved at 2.43 mol H2/mol and maximum hydrogen production rate (HPR) of 2.46 L H2/L.d. However, it is suggested that further observation, especially using a much lower HRT, is needed to conclude the performance of these two systems.

ACKNOWLEDGEMENT

The author would like to thank the Ministry of Higher Education Malaysia for granting the fund for this research under the Fundamental Research Grant Scheme, FRGS/1/2019/TK07/UIAM/02/1 (FRGS19-105-0714) and to the International Islamic University, specifically to the Department of Chemical Engineering & Sustainability for research facilities.

REFERENCES

[1] Hallenbeck P.C., (2009). Fermentative hydrogen production: principles, process, and prognosis. *International Journal of Hydrogen Energy; 34: 7379*-89

[2] Sinha P, Pandey A., (2011). An evaluative report and challenges for fermentative biohydrogen production. International Journal of Hydrogen Energy. 36:7460-7478.

[3] Maarof, R.M., Md Jahim, J., Azahar, A.Z., Abdul, P.M., Masdar, M.S., Nordin, D., Abd. Nasir, M.A., (2018). Biohydrogen production from palm oil mill effluent (POME) by two stage anaerobic sequencing batch reactor (ASBR) system for better utilization of carbon sources in POME. International Journal of Hydrogen Energy, 1-12

[4] Prawit Kongjan, Sompong O-Thong, Irini Angelidaki, (2011) Biohydrogen production from desugared molasses (DM) using thermophilic mixed cultures immobilized on heat treated anaerobic sludge granules, Int. J. Hydrogen Energy 36 (21), 14261–14269, https://doi.org/10.1016/j.ijhydene.2011.06.130.

[5] Engliman N.S., Jahim J.M., Abdul P.M., Ling T.P., Tan J.P., Ong C.B., (2019) Effectiveness of fouling mechanism for bacterial immobilization in polyvinylidene fluoride membranes for biohydrogen fermentation, Food Bioprod. Process 120; 48–57, https://doi.org/10.1016/j.fbp.2019.12.004.

[6] Tasnim Sahrin, N., Shiong Khoo, K., Wei Lim, J., Shamsuddin, R., Musa Ardo, F., Rawindran, H., Hassan, M., Kiatkittipong, W., Alaaeldin Abdelfattah, E., Da Oh, W., et al. (2022). Current perspectives, future challenges and key technologies of biohydrogen production for building a carbon–neutral future: A review. Bioresource Technology 364.

[7] Sompong O-Thong, Poonsuk Prasertsan, Nils-Kåre Birkeland, (2009). Evaluation of methods for preparing hydrogen-producing seed inocula under thermophilic condition by process performance and microbial community analysis, Bioresour. Technol. 100 (2), 909–918, https://doi.org/10.1016/j.biortech.2008.07.036.

[8] Singh L., Wahid Z.A., (2015). Enhancement of hydrogen production from palm oil mill effluent via cell immobilization technique. International Journal of Energy Resource;39:215-22.

[9] Wu K.J., Chang J.S., (2007). Batch and continuous fermentative production of hydrogen with anaerobic sludge entrapped in a composite polymeric matrix. Process Biochemistry; 42:279-84.

[10] Palazzi E, Fabiano B, Perego P., (2000). Process development of continuous hydrogen production by Enterobacter aerogenes in a packed column reactor. Bioprocess Engineering; 22:205 - 13.

[11] Bao, M. D., Su, H. J., & Tan, T. W. (2013). Dark fermentative bio-hydrogen production: Effects of substrate pre-treatment and addition of metal ions or L-cysteine. Fuel, 112, 38–44. doi:10.1016/j.fuel.2013.04.063

[12] Fuess, L. T., Fuentes, L., Bovio-Winkler, P., Eng, F., Etchebehere, C., Zaiat, M., & Augusto Oller do Nascimento, C. (2021). Full details on continuous biohydrogen production from sugarcane molasses are unraveled: Performance optimization, self-regulation, metabolic correlations and quanti-qualitative biomass characterization. Chemical Engineering Journal, 414, 128934. doi:10.1016/j.cej.2021.128934

[13] Jamali NS, Md Jahim J. (2016). Optimization of thermophilic biohydrogen production by microflora of palm oil mill effluent: cell attachment on granular activated carbon as support media. Malays J Anal Sci.; 20:1437- 1446.

[14] Zur J., Wojcieszynska D., Guzik U. (2016). Metabolic Responses of Bacterial Cells to Immobilization. *Molecules*, 21, 958; doi:10.3390/molecules21070958

[15] R. Vilela, F.T. Saia, G.B. Gregoracci, R. Duarte, P. Andrade, B. van der Zaan, A. Langenhoff, M.H.R.Z. Damianovic, (2019). Hydrogen production in reactors: The influence of organic loading rate, inoculum and support material, Int. J. Hydrogen Energy 44 (50), 27259–27271, https://doi.org/10.1016/j.ijhydene.2019.08.180

[16] Chen, C., Lin, C., Chang, J., (2001). Kinetics of hydrogen production with continuous anaerobic cultures utilizing sucrose as the limiting substrate. *Appl Microbiol Biotechnol*;57:56-64.

[17] Lin, C.Y., Chou, C.H., (2004). Anaerobic hydrogen production from sucrose using an acid-enriched sewage sludge microflora. *Eng. Life Sci*. 4, 66–70.

[18] Valdez-Vazquez, I., Rios-Leal, E., Esparza-Garcia, F., Cecchi, F., Poggi-Varaldo, H.A., (2005). Semi-continuous solid substrate anaerobic reactors for H_2 production from organic waste: Mesophilic versus thermophilic regime.*Int. J. Hydrogen Energy* 30, 1383–1391.

[19] Han, W., Hu, Y. Y., Li, S. Y., Li, F. F., & Tang, J. H. (2016). Biohydrogen production from waste bread in a continuous stirred tank reactor: A techno-economic analysis. Bioresource Technology, 221, 318- 323. doi:10.1016/j.biortech.2016.09.055

[20] Qin, Z., Qin, Q., & Yang, Y. (2013). Continuous Biohydrogen Production with CSTR Reactor under High Organic Loading Rate Condition. In Advanced Materials Research (Vols. 864–867, pp. 225–228). Trans Tech Publications, Ltd. https://doi.org/10.4028/www.scientific.net/amr.864-867.225

[21] Seengenyoung, J. , O-Thong, S. and Prasertsan, P. (2014) Comparison of ASBR and CSTR reactor for hydrogen production from palm oil mill effluent under thermophilic condition. Advances in Bioscience and Biotechnology, 5, 177-183. doi: 10.4236/abb.2014.53022.

[22] Sekoai, P. T., Awosusi, A. A., Yoro, K. O., Singo, M., Oloye, O., Ayeni, A. O., … Daramola, M. O. (2017). *Microbial cell immobilization in biohydrogen production: a short overview. Critical Reviews in Biotechnology, 38(2), 157–171.* doi:10.1080/07388551.2017.1312274

[23] Seelert T, Ghosh D, Yargeau V. Improving biohydrogen production using Clostridium beijerinckii immobilized with magnetite nanoparticles. Appl Microbiol Biotechnol. 2015;99:4107–4116.

[24] Jamali, N. S., Rashidi, N. F. D., Jahim, J. M., Thong, S. O.-, Jehlee, A., & Engliman, N. S. (2019). Thermophilic biohydrogen production from palm oil mill effluent: Effect of immobilized cells on granular activated carbon in fluidized bed reactor. Food and Bioproducts Processing. doi:10.1016/j.fbp.2019.07.012

[25] Venkata Mohan, S., Vijaya Bhaskar, Y., Sarma PN., (2007). Biohydrogen production from chemical wastewater treatment in biofilm configured reactor operated in periodic discontinuous batch mode by selectively enriched anaerobic mixed consortia. *Water Res*;41: 2652–64.

[26] Ginkel, v. S., Sung S, Lay JJ., 2001. Biohydrogen production as a function of pH and substrate concentration. *Environ Sci Technol*;35:4726–30.

[27] Zagrodnik R, Laniecki M. The role of pH control on biohydrogen production by single stage hybrid dark- and photo-fermentation. Bioresour Technol. 2015 Oct;194:187-95. doi: 10.1016/j.biortech.2015.07.028. Epub 2015 Jul 14. PMID: 26196419.

[28] Tiago Borges Ferreira, Gabriel Catucci Rego, Lucas Rodrigues Ramos, Camila Aparecida de Menezes, Laís Am´erico Soares, Isabel Kimiko Sakamoto, Maria Bernadete Amˆancio Varesche, Edson Luiz Silva, HRT control as a strategy to enhance continuous hydrogen production from sugarcane juice under mesophilic and thermophilic conditions in AFBRs, Int. J. Hydrogen Energy 44 (36) (2019) 19719–19729, https://doi.org/10.1016/j.ijhydene.2019.06.050.

[29] Su, C., Liu, Y., Yang, X., & Li, H. (2020). *Effect of hydraulic retention time on biohydrogen production from glucose in an internal circulation reactor. Energy & Fuels.* doi:10.1021/acs.energyfuels.9b033

[30] Kumar G, Sen B, Sivagurunathan P, Lin C-Y. (2016). High rate hydrogen fermentation of cello-lignin fraction in de-oiled jatropha waste using hybrid immobilized cell system. Fuel;182:131e40

[31] Albert, J., Wölfel, R., Bösmann, A., Wasserscheid, P., 2012. Selective oxidation of complex, water-insoluble biomass to formic acid using additives as reaction accelerators. *Energy & Environmental Science*. 5 (7): 7956.

[32] Thauer, R.K., K. Jungermann & K. Decker 1977. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews* 41(1): 100-180

[33] Sivagurunathan, P., Sen, B., Chiu-Yue, L., 2014. Overcoming propionic acid inhibition of hydrogen fermentation by temperature shift strategy. *International Journal of Hydrogen Energy*. 39: 19232-19241.

[34] Suksong, W., P. Kongjan, P. Prasertsan, T. Imai & O. Sompong 2016. Optimization and microbial community analysis for production of biogas from solid waste residues of palm oil mill industry by solid-state anaerobic digestion. *Bioresource technology* 214: 166-174.

[35] Lutpi NA, Jahim JM, Mumtaz T, Abdul PM, Nor MTM. Physicochemical characteristics of attached biofilm on granular activated carbon for thermophilic biohydrogen production. RSC Adv 2015;5:19382e92.

[36] Chu C-Y, Wu S-Y, Shen Y-C. Biohydrogen production performance in a draft tube bioreactor with immobilized cell. Int J Hydrogen Energy 2012;37:15658e65.

[37] Zhang S, Lee Y, Kim T-H, Hwang S-J. Effects of OLRs and HRTs on hydrogen production from high salinity substrate by halophilic hydrogen producing bacterium (HHPB). Bioresour Technol 2013;141:227e32.

https://doi.org/10.31436/iiumej.v25i2.2974

AN EASY-TO-FABRICATE, SUBMERGED CARBON-BASED AIR CATHODE FOR BIOFUEL CELLS

ASIAH SUKRI[1], RAIHAN OTHMAN[1*], NORSHAHIDA SARIFUDDIN[2]

[1]Department of Science in Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[2]Department of Manufacturing and Materials Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

**Corresponding author:* raihan@iium.edu.my

(Received: 25 August 2023; Accepted: 1 January 2024; Published on-line:15 July 2024)

ABSTRACT: An air cathode (AC) is one of the main components in an AC-based biofuel cell (BFC). Its cost accounts for nearly half of the total cost of the cell as specific requirements must be met for it to perform as an effective site for the oxygen (O_2) reduction reaction (ORR) to occur. In most applications where the AC is totally submerged in the electrolyte, air or O_2 is bubbled throughout the electrolyte during the entire discharge operation to enhance the cell performance. This is because the dissolved O_2 (DO) concentration is merely one-third of the O_2 concentration in ambient air. Unfortunately, this approach increases the overall complexity and cost of the system. Therefore, this present study developed an effective, easy-to-fabricate AC for use under totally submerged and unaerated conditions. The design principle of the proposed AC is a balance between the hydrophobicity and hydrophilicity of the components used, i.e., the combination of a carbon felt, an interwoven carbon fiber sheet, and a nickel mesh. All the cathode components were snugly fitted merely using the mechanical pressure of the cylindrical BFC holders. The fabricated AC was assembled in a zinc-air BFC employing fungal microbes *Phanerochaete chrysosporium* (*P. chrysosporium*). When tested at a constant current of 1.0 mA under unaerated, uncontrolled ambition conditions, the zinc-air BFC discharge lasted 42 days with an average operating voltage of 200 mV. Under these conditions and even without the inclusion of any catalytic material, the cell performance met the operating requirements of the low-powered remote sensing devices. Therefore, the proposed easy-to-fabricate submerged air electrode has demonstrated its viability for use in BFCs.

ABSTRAK: Katod udara (AC) merupakan salah satu komponen utama bagi sel fuel bio (BFC) berasaskan AC. Hampir separuh daripada keseluruhan kos sel berpunca dari katod udara (AC). Ini kerana beberapa keperluan khusus perlu dipenuhi bagi menyediakan tapak tindak balas penguraian (ORR) oksigen berlaku. Bagi kebanyakan aplikasi biasa di mana AC terendam sepenuhnya dalam larutan elektrolit, udara atau oksigen dialirkan secara berterusan sepanjang operasi discaj bertujuan menggandakan prestasi sel. Ini disebabkan kepekatan oksigen terlarut adalah hanya sepertiga daripada kepekatan oksigen di dalam udara sekitar. Malangnya, kaedah ini akan menambah kompleksiti dan kos. Oleh itu, kajian ini bertujuan membina AC yang efektif, bersifat mudah-pasang bagi aplikasi terendam sepenuhnya dan tanpa pengudaraan. Prinsip asas pada cadangan rekabentuk katod udara AC ini adalah bagi mengimbangi ciri hidrofobik dan hidrofilik komponen yang digunakan, iaitu kombinasi fabrik karbon, lapisan serat karbon terjalin dan jaringan nikel. Kesemua komponen katod ini terangkum kemas menggunakan tekanan mekanikal pada selinder pemegang BFC. AC ini kemudiannya diguna pakai di dalam BFC zink-udara yang terkandung mikrob kulat *Phanerochaete chrysosporium* (*P. chrysosporium*). Apabila diuji dengan arus tetap 1.0 mA dalam keadaan tanpa udara dan sekitaran tanpa kawalan, sel BFC zink-udara mampu bertahan selama 42 hari dengan purata voltan operasi sebanyak 200 mV. Dalam kondisi ini dan walau

tanpa sebarang unsur pemangkin, prestasi sel memenuhi keperluan operasi peranti penderiaan jauh bertenaga rendah. Oleh itu, katod udara yang dibangunkan bagi aplikasi elektrod terendam penuh dan bersifat mudah-pasang ini telah berhasil memenuhi keperluan bagi kegunaan BFC.

KEYWORDS: *Air-cathode biofuel cell; carbon-based air cathode; zinc-air biofuel cell; microbial fuel cell; Phanerochaete chrysosporium (P. chrysosporium)*

1. INTRODUCTION

Biofuel cells (BFC) are a potential source of green energy as they are clean, renewable, and can be derived from waste, such as municipal wastewater and agricultural waste [1,2]. However, the amount of energy discharged by BFCs is low and far from providing reasonably priced and energy-balanced operations. Multiple studies on increasing the energy balance of BFCs have yielded system designs that are manifold, complex, and costly. For instance, the capital cost of an air cathode (AC)-based BFC for treating municipal wastewater is estimated to be 30 times higher than that of existing activated-sludge systems [3].

The AC is an essential component in an AC-based BFC and accounts for nearly 45% of the overall capital cost of the entire system [4]. Carbon-based ACs, such as carbon felt, carbon cloth, carbon rods, and carbon fiber sheets, are commonly used as they have a large surface area and good electrochemical properties [5]. Catalysts are often used in an AC to increase oxygen (O_2) affinity and decrease the activation energy required for the O_2 reduction reaction (ORR) to occur, which, in turn, increases the amount of energy discharged by the AC [6]. Although platinum-based ACs enable BFCs to discharge energy at significantly higher densities [7], they are very expensive. Cobalt oxides and manganese oxides-based ACs are cheaper alternatives [8,9]. An effective AC also requires a stable triple interface, which is the electrolyte (liquid), the O_2 (gas), and the catalyst/current collector (solid) interface at which the ORR occurs. However, although this feature is unique to ACs, it is rarely discussed when reporting the performance of an AC. An AC's airside is usually porous and coated with a semipermeable, hydrophobic membrane, such as Teflon™, while the electrolyte side is denser and often dispersed with a catalyst. This ensures a good balance between hydrophobicity and hydrophilicity in an AC and, more crucially, prevents flooding, which occurs when an electrolyte accumulates in the porous airside of the AC and hinders the transportation of O_2 to the active sites of the catalyst or conductor [10,11]. Commercially available ACs are expensive due to the distinctive features mentioned above.

This present study designed and fabricated an effective and easily produced carbon-based AC for use in BFCs operating under submerged and unaerated conditions. The proposed AC comprised layers of CF and CFS fortified by a nickel mesh (NM) without the use of a catalyst. All the cathode components were cut into the desired sizes and snugly assembled using the mechanical pressure of the cylindrical holders in the BFC. The proposed AC was tested in a zinc-air microbial fuel cell (MFC) under a submerged condition [12]. Unlike an open-air AC, a submerged AC relies solely on the dissolved O_2 (DO) in the electrolyte. The DO concentration is one-third of the O_2 concentration in ambient air. Therefore, in most BFC applications with a submerged AC, the electrolyte is saturated with O_2 by bubbling air throughout the cell discharge operation. However, this approach decreases the existing low energy gain yields (output/cost) of BFCs.

https://doi.org/10.31436/iiumej.v25i2.2974

2. MATERIALS AND METHODS

2.1. Design considerations and fabrication of the air cathode

Submerged aquatic plants rely on DO and dissolved inorganic carbon (DIC) for respiration and photosynthesis [13]. There are two major challenges concerning submerged aquatic plants as compared to leafy plants exposed to atmospheric air, namely (i) significantly lower DO and DIC concentrations and (ii) profound underwater gas exchange resistance [13]. Hydrophobic leaves are a special adaptation feature or mechanism for submerged aquatic plants. A thin gas film forms on the surface of these leaves due to the hydrophobic nature of its surface. This gas film reduces gas exchange resistance caused by the diffusion boundary layer (DBL) [14,15]. For a similar transport distance across a DBL and the same concentration gradient, the gas exchange under water is 104-fold slower than in air [13]. Due to the large water-gas interfaces between DBL and thin gas film, more O_2 can be collected inside the thin gas film before diffusing into the leaves' stomata [14].

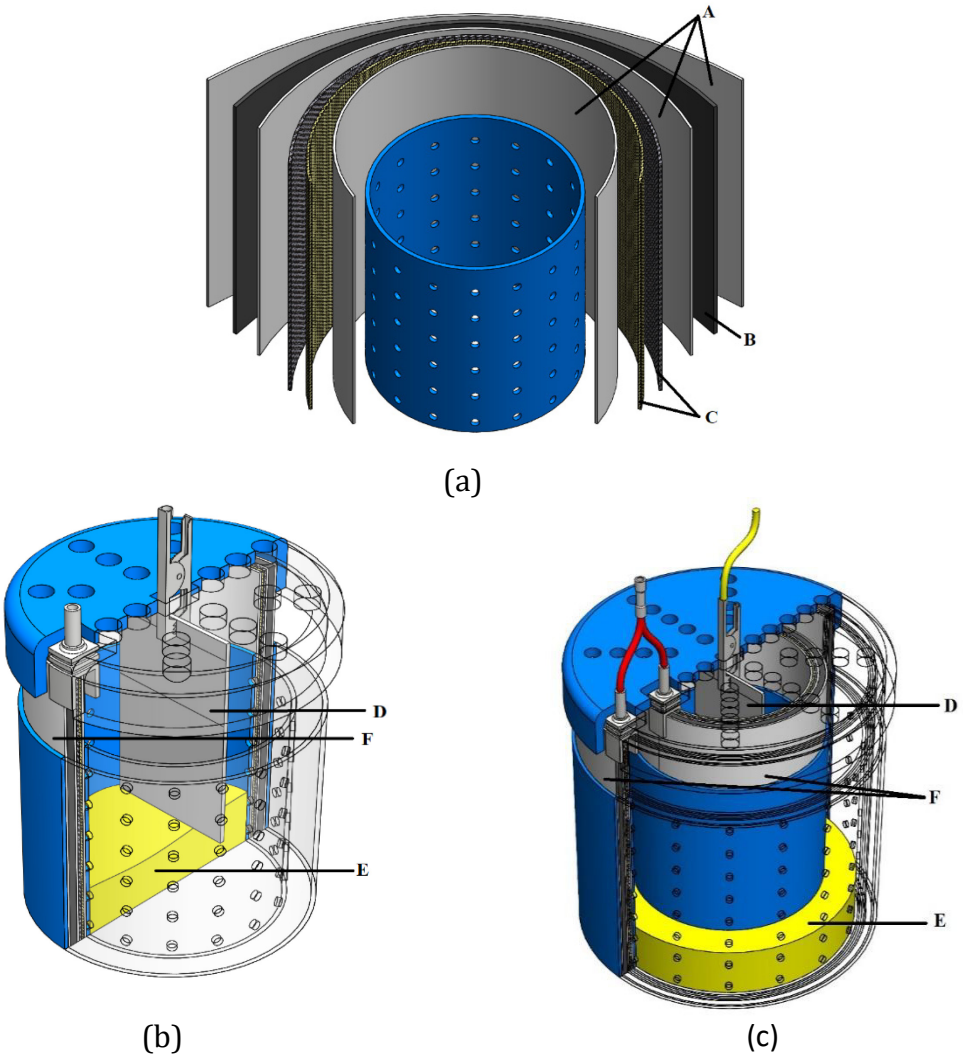


Fig. 1. Schematic illustrations of the (a) components of the AC, (b) the Zn-air MFC, and (c) the bipolar Zn-air MFC, which uses two ACs.

Legend: A – NM; B – CF; C – 2-ply CFS; D - Zn anode; E – Fungus-cultivated EFB; F – AC

https://doi.org/10.31436/iiumej.v25i2.2974

Therefore, hydrophobicity must be considered when designing a totally submerged AC to ensure its effective functioning. The AC that this present study proposed comprises NM, CF, and CFS. The NM served as both structural support and the current collector for the external circuit, the sponge-like nature of the CF served as an air trap while its intrinsic hydrophobicity enabled the formation of a gas film, and the CFS provided electron carriers for the ORR to the external circuit as it is compact, conductive, and has a large surface area.

The holder of the AC comprised two polyvinyl chloride (PVC) tubes of different diameters (height = 45 mm, ϕ_1 = 58 mm, ϕ_2 = 50 mm) with a perforated inner insert. All the cathode components were cut into the desired sizes and wedged between the two cylindrical PVC tubes. This ensured that a uniform and uniaxial pressure was placed on the multiple layers of the AC, thereby ensuring that the exact same properties were maintained for each test run. An inner insert (ϕ_2) with a different diameter could be used to apply a different uniaxial pressure. Fig. 1a illustrates the overall arrangement of the cathode, which was CF and 2-ply carbon sheets inserted between sheets of NM. As seen in Fig. 1c, a bipolar configuration was later used to increase the surface area of the AC. For this configuration, two concentric cylindrical holders of different diameters were used. The performance of the proposed AC was benchmarked against that of commercially available E4/E4A EFL AC strips that had been cut into identical 3 cm by 3 cm pieces.

2.2. Characterisation of the air cathode structure

A JEOL® JSM-6700F field emission scanning electron microscope (FESEM; Tokyo, Japan) was used to observe the CF and CFS surface morphologies of the proposed AC and the E4 AC. ImageJ was used to analyze the hydrophobic contact angle of the CF.

2.3. Fabrication and the operating principles of the Zn-air MFC

The membrane-less, single-chamber cylindrical BFC was filled with 250 ml of 24 g/L unbuffered potato dextrose broth (PDB) electrolyte. The microbes were cultured from 5 g of dried *Phanerochaete chrysosporium* (*P. chrysosporium*) in an organic substrate of 2 g of oil palm-derived empty fruit bunches (EFBs). The fungus-cultivated EFBs resided at the bottom of the BFC enclosure. Further details may be obtained from [15]. A 40 mm by 100 mm Zn anode strip was placed in the middle of the MFC enclosure.

The ORR that occurs on an AC is biocatalytic. When an AC is supplied with lignin-rich EFBs, the *P. chrysosporium*-derived fungal microbes secrete ligninolytic enzymes that predominantly contain laccase, which has a specific affinity for O_2 molecules as its electron acceptor. Therefore, while the fungal microbes degrade the lignocellulosic wall, the laccase catalyzes the reduction of molecular O_2 [16].

$$O_2 + 4H^+ + 4e^- \xrightarrow[laccase]{} 2H_2O \quad (1)$$

As such, a bio-catalyzed electrochemical BFC can be produced by pairing a Zn anode with an AC in a medium rich with *P. chrysosporium* hyphae.

2.4. Electrochemical analysis

A Neware® BTS 4000 battery tester (Shenzhen, China) was used to test the power density and discharge capacity of the Zn-air MFC containing the proposed AC under a 1 mA constant current.

https://doi.org/10.31436/iiumej.v25i2.2974

3. RESULTS AND DISCUSSION

A galvanostatic discharge test has been used to evaluate the bioelectrochemical system's capacity to produce electricity utilizing the proposed submerged air electrode. It is a direct method to assess the system's ability to maintain the charge transfer rate between the electroactive species and the conducting electrode, as well as the stability of the charge transfer processes [12]. Fig. 2 shows the discharge capacity of the proposed AC at a constant current of 1 mA. The Zn-air MFC containing the proposed AC discharged an average operating voltage of 200 mV for 42 days. The total output energy estimated based on the area under the discharge plot was 211 mWh. The performance of the proposed AC was benchmarked against that of a commercially available E4 AC. The E4 AC was able to discharge an average operating voltage of 280 mV for 53 days for a total output energy of 403 mWh. Obviously, being incorporated with a manganese-based catalyst, the E4 air electrode could discharge two times more energy than the proposed non-catalytic AC.

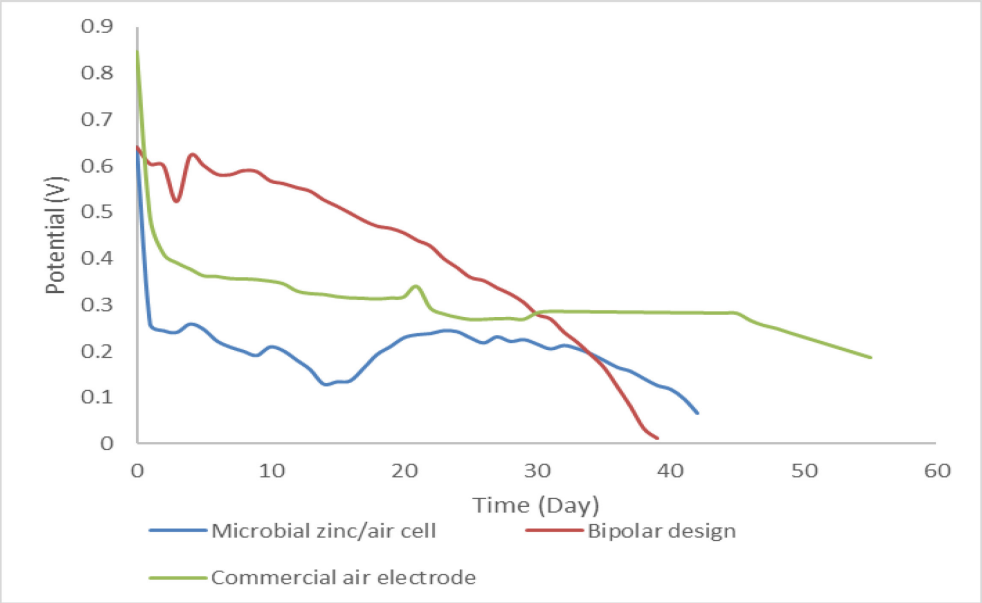


Fig. 2. An evaluation of the AC based on the Zn-air MFC discharge profile at 1 mA

Enhancing the AC surface area has been identified as one of the most critical factors to increase the volumetric power density of an AC-based MFC [17]. It was reported that doubling the AC surface could increase the output by 62%. However, this option is very costly and infeasible for a catalyst-based AC. A bipolar configuration was then used to increase the proposed non-catalytic AC's surface area, significantly improving its discharge profile (Fig. 2). More specifically, increasing the surface area of the proposed AC by 2.5 times enabled it to discharge more energy than the E4 AC for the first 31 days. Furthermore, although the enhanced proposed AC stopped discharging energy on Day 40, its total energy output was 396 mWh, which was only 1.7% lower than that of the E4 AC.

The proposed AC was tested under the following conditions: (i) totally submerged in an unaerated electrolyte so as to solely rely on DO, which constitutes one-third of the O_2 concentration in ambient air [18]; (ii) solely rely on the biocatalyst secreted by the freely suspended fungal *P. chrysosporium*, while the E4 AC was incorporated with a manganese catalyst, and (iii) discharged at a higher amount of current (1 mA) as compared to the commonly reported discharge load of 0.6 to 0.7 mA [12]. In the abovementioned test conditions, the Zn-air MFC containing the proposed submerged, non-catalytic AC

https://doi.org/10.31436/iiumej.v25i2.2974

continuously discharged 1 mA of current for over a month under unaerated and uncontrolled ambient surroundings. These conditions are essential for a bio-electrochemical cell to serve as a stand-alone energy source for low-powered remote sensing devices. First, an open-air AC would induce electrolyte evaporation from the cell as air access must be allowed, whereas for a submerged air electrode, it solely depends on DO, and only a tiny hole is sufficient to maintain the air/electrolyte gas exchange. Second, the cell must be able to operate in a self-sustaining mode for an extended duration without the input of external systems/energy, such as electrolyte aeration or intermittent substrate replenishment [17]. Third, a typical low-powered sensor requires about 2 mA at start-up and hundreds of μA in steady state [19]. Fig. 3 shows the power density profile of the Zn-air MFC containing the proposed submerged AC. The cell registered a maximum power output (1.17 mW) at 4.2 mA, which is way beyond the operating requirements of the low-powered sensors. Fig. 3 also compares the power density profile of the commercial AC. The maximum power density was 1.61 mW at a discharge current of 2.7 mA, which is much higher than the proposed submerged AC. However, the proposed submerged AC possessed a much higher limiting current density.

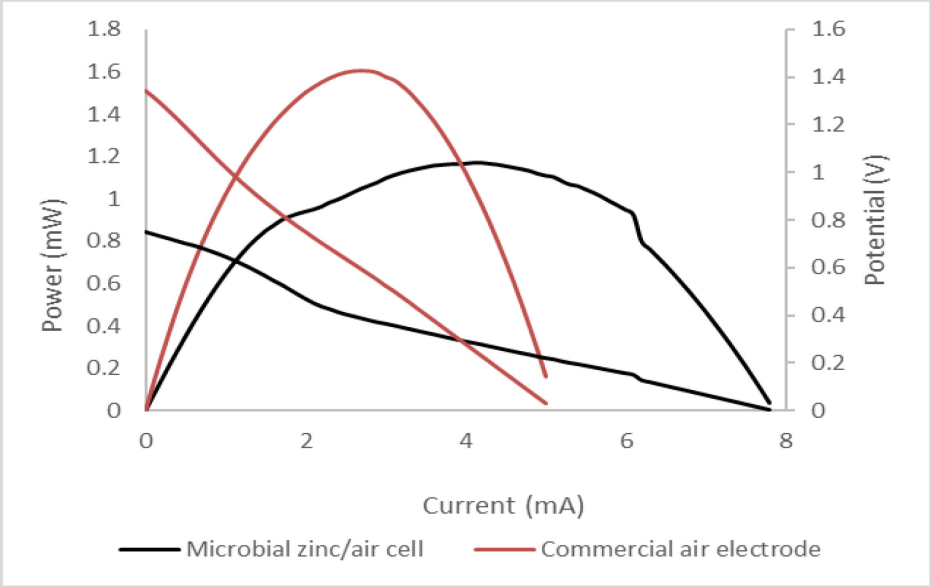


Fig. 3. A comparison of the power and polarization profiles of the Zn-air MFC using the proposed submerged AC and a commercial AC.

Among the notable demonstrations of self-sustaining MFC employing the AC is reported by Khaled et al. [19]. In battery mode (i.e., a single supply of sodium acetate substrate, without replenishment), the MFC power stack could operate a low-powered temperature sensor for 20 days. The power stack comprised two MFCs and was incorporated with a power management unit (PMU). The PMU comprised a DC/DC converter, capacitors to filter MFC output voltage fluctuations, and a mechanical switch. Each MFC alternately powered the sensor in a 48-hour cycle. The MFC comprised a single chamber with 700 ml capacity, inoculated with industrial wastewater and fed with 0.7 g of sodium acetate. The anode was a stainless-steel mesh with a projected area of 100 cm^2. The AC was a 120 cm^2 carbon cloth (30% Teflon treated, 0.5 mg/cm^2 platinum (Pt) catalyst loading). The biofilm on the AC was developed separately. With a total Pt catalyst loading of 60 mg, the MFC registered a maximum power point of about 1.1 mW/m^2 at around 3.5 mA and 0.9 V. Conversely, the Zn-air MFC employing the non-catalytic AC produced a maximum power point of 50 mW/m^2 at 4.2 mA and 0.3 V. The immediate consequence of a non-catalytic AC is the huge potential drop due to the high activation energy, which, however, is compensated by its ability to sustain a higher current.

https://doi.org/10.31436/iiumej.v25i2.2974

An AC generally comprises a conductive base and a gas diffusion layer (GDL). If a catalyst is required, it is incorporated between these two components. The conductive base supplies the electrons required for the ORR to occur from the external circuit or anode, while the GDL is where O^2 diffusion occurs between the electrolyte and the active sites [20,21]. As such, GDLs are designed to be hydrophobic to prevent cathode flooding [20]. The GDL is usually called the "airside," while the conductive base is called the "waterside." The interface between the GDL and the conductive base is a triple interface, i.e., the electrolyte (liquid), O_2 (air), and the catalyst/current collector (solid), at which the ORR occurs. Therefore, it is crucial to stabilize the triple interface to maintain the efficacy and functions of an AC. Most metal-air cells contain an open-air AC, where one side is exposed to air. However, in a seawater aluminum/air cell, the AC is fully submerged in the liquid electrolyte. In such cases, cell operation solely depends on the DO in the electrolyte, which is available at significantly lower concentrations than the amount of O_2 in ambient air [18]. A hydrophobic GDL now performs another important function of facilitating the exchange of O_2 between the cathode and the electrolyte at the gas-liquid interface [22,13]. Air cathode (AC)-based BFCs often use submerged ACs [23,24]. Although O_2 is bubbled throughout the operation of most cell configurations to increase the amount of DO available, a well-designed AC could further improve cell performance.

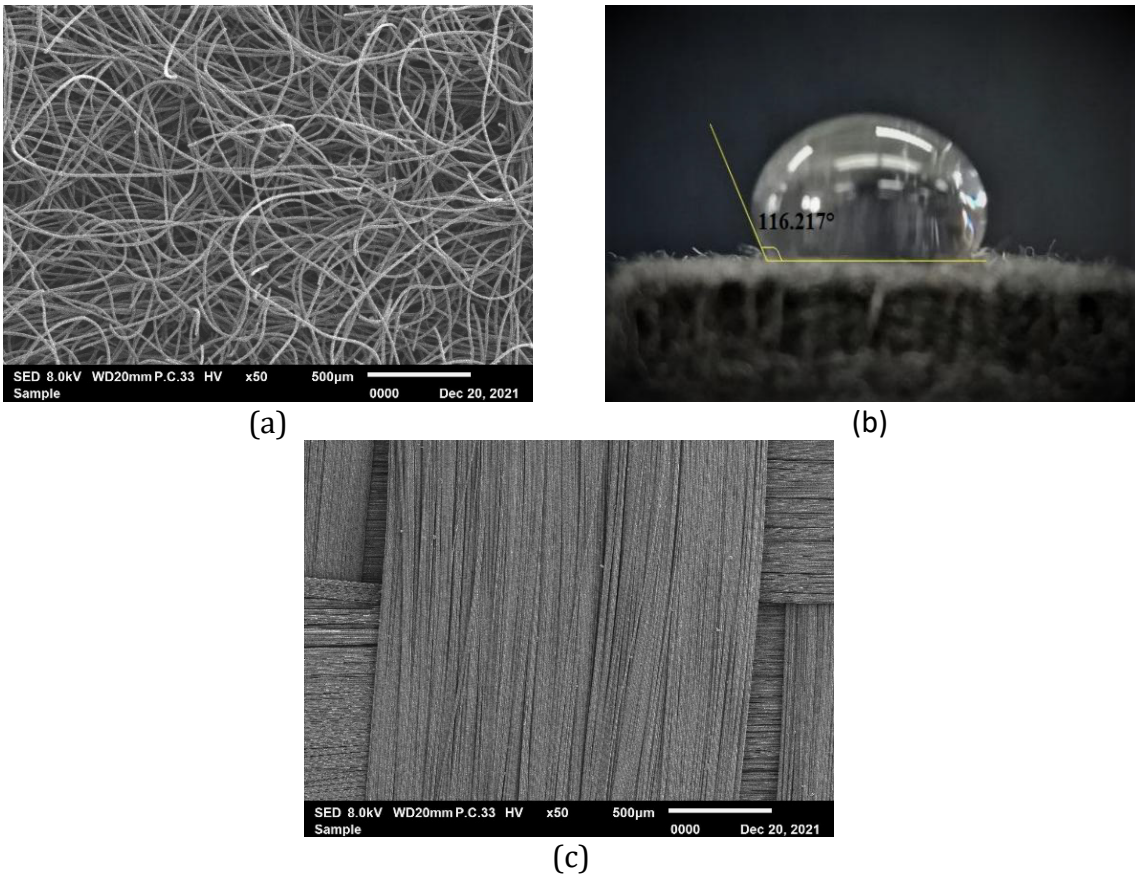


Fig. 4. (a) An SEM image of the carbon fiber depicting extended CF forming the spongy and porous matrix, (b) a digital image depicting the hydrophobic nature of CF; and (c) an SEM image of the CFS depicting the interwound bundles of carbon fiber forming a compact structure.

https://doi.org/10.31436/iiumej.v25i2.2974

The proposed AC functioned effectively and supported the design estimations. The SEM results indicated that CF contains loose and long carbon fibers that form a porous and spongy texture (Fig. 4a), which enables it to entrap air. Fig. 4b depicts the hydrophobic nature of CF, where an H_2O droplet is placed on it. Therefore, when the CF is fully submerged in the electrolyte, its hydrophobicity should enhance O_2 exchange with the electrolyte. Carbon fiber sheets (CFS) were used as a conductive base to supply electrons from the external circuit for the ORR to occur. The SEM results indicated that CFS contain interwoven threads of carbon fiber bundles (Fig. 4c). Carbon fiber sheets (CFS) are widely used as a base in cathodes because they are highly conductive and have good chemical and electrochemical stability. Meanwhile, NM provides structural integrity in an AC and acts as a current collector.

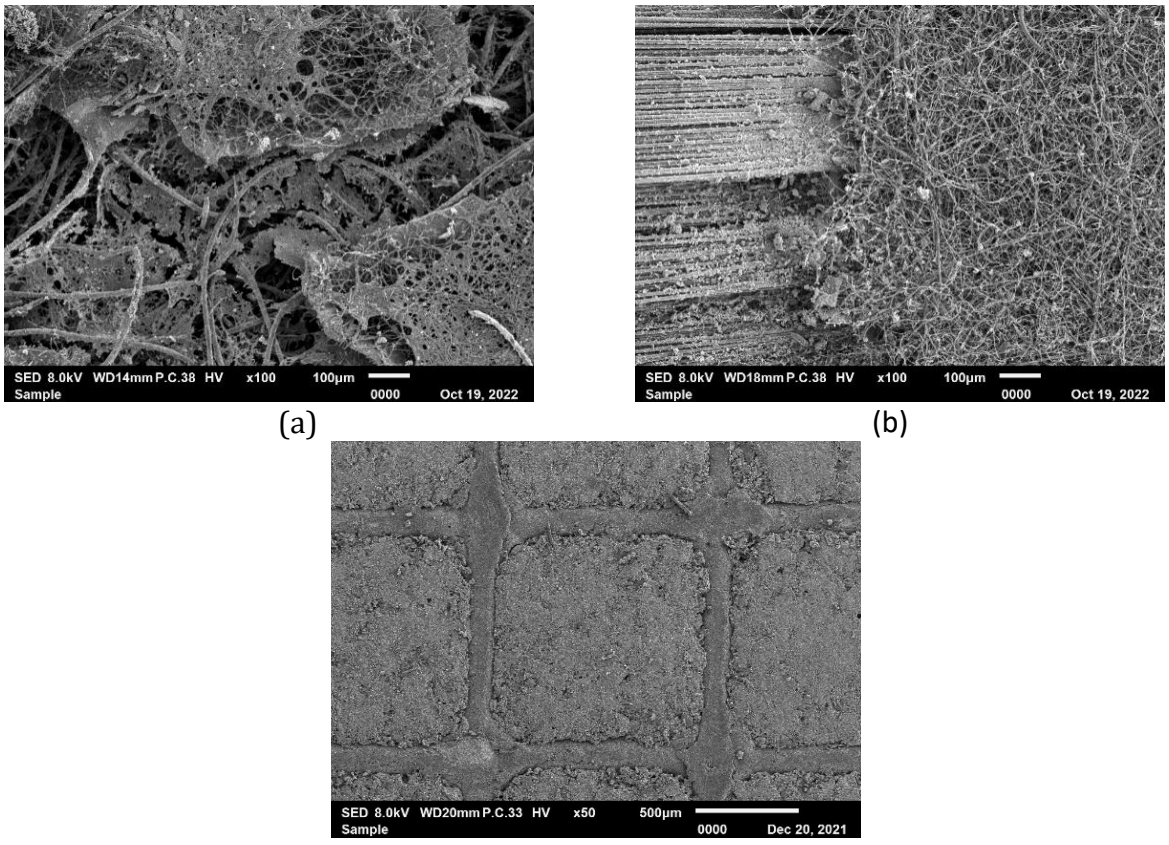


Fig. 5. (a) An SEM image depicting fungal colony growth on the CF, (b) an SEM image depicting the dense mycelia colony of the *P. chrysosporium* on the CFS, (c) an SEM image of E4 AC depicting a compact structure of the waterside with no trace of fungal growth.

When benchmarked against an E4 AC, the discharge duration and capacity of the Zn-air MFC containing the proposed AC were shorter and 48% less than that of the E4 AC. The E4 AC possesses two advantages: first, it is dispersed with an Mn-based catalyst within the AC structure. On the other hand, the proposed AC entirely relied on the freely suspended biocatalyst excreted by the fungal microbes. Second, the electrolyte side of the E4 electrode is dense and prevents the penetration of fungal mycelia colonies. When the proposed AC was configured bipolarly, which is a stack design, it discharged more energy than the E4 AC for the first 31 days. Its discharge duration, however, was still comparable to that of a single AC configuration. Scanning electron microscopy (SEM) was used to examine the proposed

enhanced AC once it had ceased operation. As seen in Fig 5a and 5b, dense *P. chrysosporium* mycelia colonies had formed on the surfaces of both the CF and CFS of the proposed enhanced AC. Conversely, as the surfaces of the carbon components of the E4 AC are dense, they prevented the penetration and proliferation of fungal mycelia colonies (Fig. 5c). The airside of an E4 AC strip is encased in a Teflon™ membrane, which has comparable hydrophobicity to CF. It is possible that the proliferation of fungal mycelia colonies between the CF and CFS layers of the proposed enhanced AC prevented it from functioning effectively and sustaining the desired current output. Therefore, the CFS, or waterside, of BFCs that utilize fungal microbes must be protected or coated with a dense layer to prolong its service life. It is also noteworthy that, although the proposed AC does not contain any catalytic materials, its discharge performance could be tailored simply by increasing its surface area.

4. CONCLUSION

This present study demonstrated a simple method of fabricating an effective, submerged carbon-based AC for use in BFCs. Commercially available CF and CFS were, respectively, chosen as the GDL and conductive base of the proposed AC based on their as-purchased properties and used without any modifications. Therefore, the O_2-reducing performance and efficacy of the proposed AC could be further enhanced by pre-treating the CF and CFS layers using multiple methods. The proposed AC was tested as a fully submerged AC in a Zn-air MFC and found to be able to discharge a comparatively high constant current over a one-month period. Therefore, the proposed AC is viable and effective for use under submerged conditions, whether for non-air-circulated or air-circulated BFCs.

ACKNOWLEDGEMENT

This work was funded by the research grant from the Ministry of Science, Technology and Innovation Malaysia (Research Grant IF0219E1059). The authors gratefully acknowledge the financial support.

REFERENCES

[1] Vishnevskaya M, Gazizova D, Victorenko A, & Konova I. (2019). Membraneless microbial biofuel cell for municipal waste water treatment. IOP Conference Series: Earth and Environmental Science, 337(1):2–7. https://doi.org/10.1088/1755-1315/337/1/012002.

[2] Pal M, & Sharma RK. (2020). Development of wheat straw based catholyte for power generation in microbial fuel cell. Biomass and Bioenergy, 138(January):105591. https://doi.org/10.1016/j.biombioe.2020.1055916.

[3] Li WW, Yu HQ, & He Z. (2014). Towards sustainable wastewater treatment by using microbial fuel cells-centered technologies. Energy and Environmental Science, 7(3):911–924. https://doi.org/10.1039/c3ee43106a.

[4] Rozendal RA, Hamelers HVM, Rabaey K, Keller J, & Buisman, CJN. (2008). Towards practical implementation of bioelectrochemical wastewater treatment. Trends in Biotechnology, 26(8):450–459. https://doi.org/10.1016/j.tibtech.2008.04.008.

[5] Reinhard S, Sitorus I, Hani M, Bakar A, & Majlan EH. (2020). Review on High-Performance Air Cathode Microbial Fuel Cell for Power Generation and COD Reduction. Jurnal Kejuruteraan, 32(4):569–577.

[6] Khilari S, Pandit S, Das D, & Pradhan D. (2014). Manganese cobaltite/polypyrrole nanocomposite-based air-cathode for sustainable power generation in the single-chambered microbial fuel cells. Biosensors and Bioelectronics, 54:534–540. https://doi.org/10.1016/j.bios.2013.11.044.

[7] Choi YJ, Mohamed HO, Park SG, Al Mayyahi RB, Al-Dhaifallah M, Rezk H, Ren X, Yu H, & Chae KJ. (2019). Electrophoretically fabricated nickel/nickel oxides as cost effective nanocatalysts

for the oxygen reduction reaction in air-cathode microbial fuel cell. International Journal of Hydrogen Energy, 45(10):5960–5970. https://doi.org/10.1016/j.ijhydene.2019.05.091.

[8] Ahmed J, Yuan Y, Zhou L, & Kim S. (2012). Carbon supported cobalt oxide nanoparticles-iron phthalocyanine as alternative cathode catalyst for oxygen reduction in microbial fuel cells. Journal of Power Sources, 208:170–175. https://doi.org/10.1016/j.jpowsour.2012.02.005.

[9] Das S, Chakraborty I, Rajesh PP, & Ghangrekar MM. (2020). Performance Evaluation of Microbial Fuel Cell Operated with Pd or MnO_2 as Cathode Catalyst and Chaetoceros Pretreated Anodic Inoculum. Journal of Hazardous, Toxic, and Radioactive Waste, 24(3), 1–7. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000501.

[10] Wang Y, & Leung DYC. (2017). Optimization of Cathode Flooding in Scaled-up Microfluidic Fuel Cells. Energy Procedia, 105:1454–1460. https://doi.org/10.1016/j.egypro.2017.03.429.

[11] Gajda I, Greenman J, Melhuish C, Santoro C, Li B, Cristiani P, & Ieropoulos I. (2014). Water formation at the cathode and sodium recovery using Microbial Fuel Cells (MFCs). Sustainable Energy Technologies and Assessments, 7:187–194. https://doi.org/10.1016/j.seta.2014.05.001.

[12] Sukri A, Othman R, Abd-Wahab F, & Noor NM. (2021). Self-sustaining bioelectrochemical cell from fungal degradation of lignin-rich agrowaste. Energies, 14(8):2098 https://doi.org/10.3390/en14082098.

[13] Pedersen O, Colmer TD, & Sand-Jensen K. (2013). Underwater photosynthesis of submerged plants - Recent advances and methods. Frontiers in Plant Science, 4: 1–20. https://doi.org/10.3389/fpls.2013.00140.

[14] Verboven P, Pedersen O, Ho QT, Nicolai BM, & Colmer TD. (2014). The mechanism of improved aeration due to gas films on leaves of submerged rice. Plant Cell and Environment, 37(10):2433–2452. https://doi.org/10.1111/pce.12300.

[15] Winkel A, Visser EJW, Colmer TD, Brodersen KP, Voesenek LACJ, Sand-Jensen K, & Pedersen O. (2016). Leaf gas films, underwater photosynthesis and plant species distributions in a flood gradient. Plant, Cell and Environment, 39(7):1537–1548. https://doi.org/10.1111/pce.12717.

[16] González-González P, Gómez-Manzo S, Tomasini A, Martínez Y Pérez JL, García Nieto E, Anaya-Hernández A, Ortiz Ortiz E, Castillo Rodríguez RA, Marcial-Quino J, & Montiel-González AM. (2023). Laccase Production from Agrocybe pediades: Purification and Functional Characterization of a Consistent Laccase Isoenzyme in Liquid Culture. Microorganisms, 11(3):568. https://doi.org/10.3390/microorganisms11030568.

[17] Cheng S, & Logan BE. (2011). Increasing power generation for scaling up single-chamber air cathode microbial fuel cells. Bioresource Technology, 102(6):4468–4473. https://doi.org/10.1016/j.biortech.2010.12.104.

[18] Canadian Council of Ministers of the Environment. (1999). Canadian Water Quality Guidelines for the Protection of Aquatic Life - Dissolved Oxygen (Freshwater). In Canadian environmental quality guidelines.

[19] Khaled F, Ondel O, & Allard B. (2016). Microbial fuel cells as power supply of a low-power temperature sensor. Journal of Power Sources, 306:354–360. https://doi.org/10.1016/j.jpowsour.2015.12.040.

[20] Tomboc GM, Yu P, Kwon T, Lee K, & Li J. (2020). Ideal design of air electrode-A step closer toward robust rechargeable Zn-air battery. APL Materials, 8(5):050905 https://doi.org/10.1063/5.0005137.

[21] Yang W, Kim KY, & Logan BE. (2015). Development of carbon free diffusion layer for activated carbon air cathode of microbial fuel cells. Bioresource Technology, 197:318–322. https://doi.org/10.1016/j.biortech.2015.08.119.

[22] Arif M, Cheung SCP, & Andrews J. (2020). Influence of Hydrophobicity and Porosity of the Gas Diffusion Layer on Mass Transport Losses in PEM Fuel Cells: A Simulation Study Supported by Experiments. Energy and Fuels, 34(10):13010–13022. https://doi.org/10.1021/acs.energyfuels.0c02596.

[23] Chen S, Patil SA, & Schröder U. (2018). A high-performance rotating graphite fiber brush air-cathode for microbial fuel cells. Applied Energy, 211:1089–1094. https://doi.org/10.1016/j.apenergy.2017.12.013.

[24] Saba B, Christy AD, Yu Z, Co AC, Islam R, & Tuovinen OH. (2017). Characterization and performance of anodic mixed culture biofilms in submersed microbial fuel cells. Bioelectrochemistry, 113:79–84. https://doi.org/10.1016/j.bioelechem.2016.10.003.

https://doi.org/10.31436/iiumej.v25i2.2978

EVALUATION AND KINETICS OF TOFU WASTEWATER BIOREACTOR WITH ADDITION OF WATER HYACINTH

HERAWATI BUDIASTUTI[1*], LAILY ISNA RAMADHANI[1], PRATAP PULLAMMANAPPALLIL[2], RANGGI OCTAVIANI PRATIWI[3], YULIANA NUR AMANAH[4]

[1]Chemical Engineering Department, Politeknik Negeri Bandung, Bandung 40559, Indonesia
[2]Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611, USA
[3]PT Kimia Farma, Jakarta 13920, Indonesia
[4]Center for Standardization and Services of Textile Industry, Ministry of Industry, Bandung 40272, Indonesia

**Corresponding author: herabudi@polban.ac.id*

(Received: 29 August 2023; Accepted: 29 April 2024; Published online: 15 July 2024)

ABSTRACT: Tofu wastewater is a nitrogen-rich wastewater type with a high concentration of organic contents. To enhance the digestion rate of tofu wastewater in an anaerobic bioreactor, water hyacinth was tested as an additional substrate. Water hyacinth is a dangerous weed with a high carbon-nitrogen ratio (C/N) of about 30-35. Substrate combination was intentionally used in this study to raise the C/N of wastewater. This study aims to evaluate the digestion rate of a bioreactor qualitatively based on experimental data to determine the biokinetic constants of the anaerobic bioreactor quantitatively based on microbial growth data, substrate degradation, and methane (CH_4) production. MATLAB was used as software to run mathematical modeling. The anaerobic bioreactor was designed and equipped with a circulation pump to maintain the homogeneity of the substrate and was completed with a biogas collector. Seeding and acclimatization were carried out before the main experiment started. Using a combination substrate of tofu wastewater and water hyacinth with a 5:3 volume ratio, the bioreactor was run at 20 days of hydraulic retention time (HRT) at room temperature. The bioreactor was able to remove 92.8% of chemical oxygen demand (COD) concentration and produced biogas with the highest CH_4 concentration of 56.9%. The obtained kinetic constants indicate that, in comparison with similar studies of anaerobic digestion of the low C/N wastewater, the addition of water hyacinth resulted in better performance of the bioreactor with the correlation of microbial growth rate (μ_m), substrate degradation ($Y_{X/CCOD}$) and CH_4 production ($Y_{CCH4/X}$) with the consecutive values of 0.65/day, 0.64 mg cells/mg COD, and 0.62 mg CH_4/mg cells. An appropriate ratio of water hyacinth as the high carbon source and nitrogen-rich tofu wastewater is recommended to obtain the optimum ratio of carbon to nitrogen and result in a higher percentage of methane formation.

ABSTRAK: Air sisa tauhu ialah air sisa yang kaya dengan nitrogen dengan kepekatan kandungan organik yang tinggi. Bagi meningkatkan kadar penghadaman air sisa tauhu dalam bioreaktor anaerobik, keladi bunting diuji sebagai substrat tambahan. Keladi bunting merupakan rumpai berbahaya dengan nisbah karbon nitrogen (C/N) yang tinggi iaitu kira-kira 30-35. Gabungan substrat sengaja digunakan dalam kajian ini bagi menaikkan C/N air sisa. Kajian ini bertujuan bagi menilai kadar pencernaan bioreaktor secara kualitatif berdasarkan data eksperimen dan menentukan pemalar biokinetik bioreaktor anaerobik secara kuantitatif berdasarkan data pertumbuhan mikrob, degradasi substrat, dan pengeluaran metana (CH_4). MATLAB digunakan sebagai perisian pemodelan matematik. Bioreaktor anaerobik direka bentuk dan dilengkapi dengan pam edaran bagi mengekalkan kehomogenan substrat dan

https://doi.org/10.31436/iiumej.v25i2.2978

dilengkapkan dengan pengumpul biogas. Penyemaian dan penyesuaiikliman telah dijalankan sebelum eksperimen utama bermula. Substrat gabungan air sisa tauhu dan keladi bunting digunakan dengan nisbah isipadu 5:3. Bioreaktor dijalankan selama 20 hari iaitu masa pengekalan hidraulik (HRT) pada suhu bilik. Bioreaktor tersebut mampu mengasingkan 92.8% kepekatan permintaan oksigen kimia (COD) dan menghasilkan biogas dengan kepekatan CH_4 tertinggi sebanyak 56.9%. Pemalar kinetik yang diperoleh menunjukkan bahawa, berbanding dengan kajian serupa tentang pencernaan anaerobik air sisa C/N rendah, penambahan keladi bunting menghasilkan prestasi bioreaktor yang lebih baik dengan korelasi kadar pertumbuhan mikrob (μ_m), degradasi substrat ($Y_{X/CCOD}$) dan penghasilan CH_4 ($Y_{CCH4/X}$) dengan nilai masing-masing 0.65/hari, 0.64 mg sel/mg COD, dan 0.62 mg CH_4/mg sel. Nisbah keladi bunting yang sesuai sebagai sumber karbon tinggi dan air sisa tauhu yang kaya dengan nitrogen disyorkan agar mendapatkan nisbah optimum karbon kepada nitrogen dan ini menghasilkan peratusan pembentukan metana yang lebih tinggi.

KEYWORDS: *Anaerobic process, kinetic constants, C/N ratio, tofu wastewater, water hyacinth*

1. INTRODUCTION

Tofu wastewater is one of the wastes produced in tofu production with a high content of organic substances. The organic content of tofu wastewater measured as COD is fairly high, which is in the range of 10,000 to 16,000 mg/L [1, 2]. The existence of this wastewater has the potential to endanger the environment if it is directly flowed into water bodies, as is sometimes happening today. Referring to Ministerial Regulation of Indonesia No. 5 of 2014 concerning Wastewater quality Standards, the COD concentration of tofu wastewater allowed is 275 mg/L [3].

The tofu industry in Indonesia is dominated by the home-scale industry, which uses conventional waste management. An anaerobic process with the help of microorganisms is one of the applicable options due to its ability both to degrade organic compounds and to form methane. The anaerobic digestion consists of mixed culture microorganisms, such as acidogenic bacteria that are responsible for the acidogenesis process and methanogenic bacteria for the methanogenesis process. Several factors greatly influence this process. According to the Authors of [4], the ratio of carbon to nitrogen (C/N) contained in the substrate is a factor of concern that must be considered by researchers. Tofu wastewater has a low ratio of C/N, which was 5.90 [1], while the Authors of [5] suggested that the optimum ratio of C/N in the formation of methane gas is 25-30. The low ratio of C/N contained in the substrate would result in the high formation of total ammonia nitrogen (TAN) and volatile fatty acids (VFAs) [5]. The concentrations of TAN and VFA accumulation inside the bioreactor decrease pH levels in acidic conditions; this condition potentially inhibits the growth of methanogenic bacteria and their activity in producing methane. The very high ratio of C/N is also not an ideal condition for the methanogenic process and impacts the low methane yield due to the lack of nitrogen concentration. Nitrogen is needed for the bacteria's growth, especially for methanogenic bacteria.

The alternative solution to increase the C/N ratio in this study was to use a substrate combination. The combination of substrates optimizes the C/N ratio inside the bioreactor to ensure the availability of nutrients needed by anaerobic microorganisms, both for acidogenic and methanogenic bacteria. Several studies reported information on increasing biogas formation in wastewater treatment processes through anaerobic treatment using combined substrates. Combining substrates increased biogas productivity compared to single substrates only [6,7]. The researchers of [7] presented that a substrate combination of organic kitchen

https://doi.org/10.31436/iiumej.v25i2.2978

waste and cattle manure with a ratio of 3:1 can increase the biogas yield by 47.13%, and a substrate combination of food waste and cattle manure with a ratio of 2:1 can increase the methane yield by 41.1%. Water hyacinth is a dangerous weed that can grow very quickly and deplete the nutrients and oxygen contained in water bodies. This weed is quite harmful to the surrounding ecosystem [8]. On the other hand, water hyacinth is a plant with a high C/N ratio which is 30-35 [9]. The optimum C/N ratio in anaerobic digestion can be achieved by mixing organic materials that have a low C/N ratio with materials that have a higher C/N ratio [10].

This study aims to add water hyacinth extract to the bioreactor to increase the C/N ratio of the substrate. The effects of this treatment are evaluated qualitatively through experimental data. The kinetic constants obtained from mathematical modeling complete the quantitative evaluation of this process, both in the process of organic compound degradation and CH_4 production.

2. METHODS

2.1. Raw Materials

The raw materials used in this study were tofu wastewater, water hyacinth extract and cow rumen. The materials were collected from a tofu home industry, a fish pond, and an animal farm in the Bandung area, Indonesia, respectively. Water hyacinth extract was prepared by blending and smoothing the water hyacinth leaves and water with a 2:13 weight ratio. Tofu wastewater and water hyacinth extract were mixed with a volume ratio of 5:3. Cow rumen as the starter bacteria was diluted using water with a ratio of 1:2 [1]. The characteristics of all materials were tested before the experiment started. The analysis of total nitrogen, carbon, and phosphate contents was carried out at the Center for Material and Technical Products in Bandung, Indonesia. COD, pH, volume, and gas concentrations were analyzed in the Water and Wastewater Treatment Laboratory in Politeknik Negeri Bandung, Indonesia.

2.2. Operational of Bioreactor

The anaerobic bioreactor used was specially designed for this research to increase biogas production through an anaerobic process, as shown in Fig. 1. The reactor was equipped with a circulation pump to maintain the homogeneity of the substrate concentration inside the bioreactor. Substrate circulation was carried out every day for 30 minutes before sampling and after feeding the wastewater. The circulation was conducted by closing valves 1 and 5 and opening valves 2, 3, and 4. The circulation pipe contained in the bioreactor was installed ¾ of the length of the feed pipe, which aims to prevent air from entering the bioreactor. The total capacity of this bioreactor (B) is 25 L with 20 L of active volume. The total capacity of the feed tank (A) is 28 L. The operating condition of this bioreactor was set at room temperature in neutral condition (pH 7), which was the optimum pH for the methanogenesis process. The bioreactor was operated at 20 days of hydraulic retention time (HRT) to prevent shock-loading problems for the microorganisms [1].

The experiment began by running the seeding stage and continued with the acclimatization process before the main process was carried out. The seeding process, as the preliminary step, was aimed to support the growth of microorganisms by adding 40 grams of glucose (2,000 mg/L) into the bioreactor. The use of an ideal substrate was proposed to shorten the lag phase and optimize the exponential phase of the microorganism's growth. The seeding process was stopped when MLVSS concentration reached more than 2,000 mg/L.

The acclimatization process was carried out by gradually replacing the ideal substrate (glucose) with tofu wastewater. One liter per day of mixing substrate was fed into the bioreactor

https://doi.org/10.31436/iiumej.v25i2.2978

with a concentration of 6,000 mg/L. The composition of tofu wastewater in the mixing substrate was gradually increased, starting from 5-10% per 3 days until all substrate was fully replaced by tofu wastewater. Acclimatization was ended when the fluctuation of decreasing COD was under 10% [11]. Appropriate acclimatization will result in the ability of microorganisms in the bioreactor to adapt efficiently to new substrates in the next process, namely the main process. The main research on the anaerobic bioreactor was conducted by feeding 1 L per day of tofu wastewater while keeping its concentration around 6,000 mg/L. Gas volume and pH were measured every day, while COD and MLVSS were tested every 3 days. Methane in the biogas was analyzed every 5 days to detect the methane gas formed.

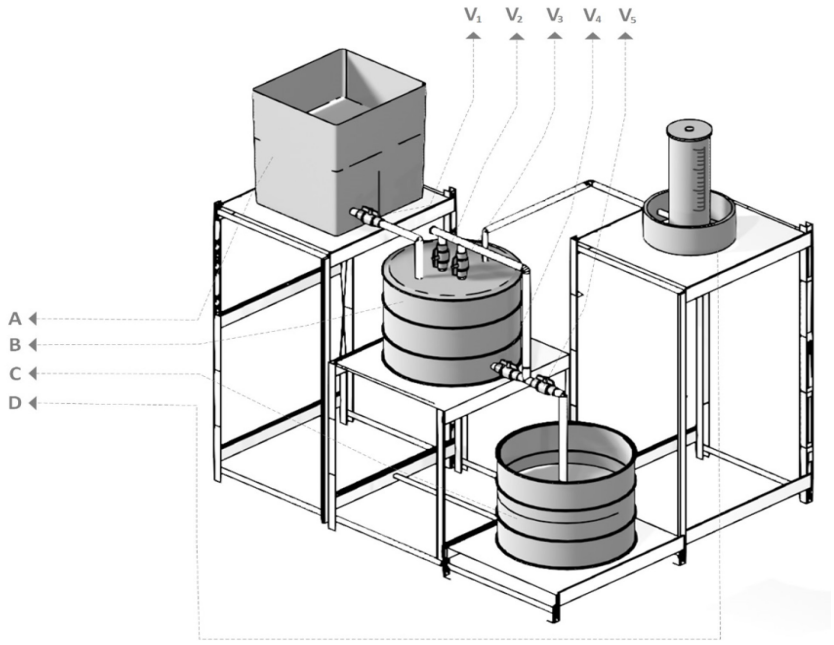


Fig. 1. Set up of bioreactor.

Annotation

a	: Feeding tank
b	: Anaerobic bioreactor
c	: Effluent tank
d	: Biogas collector
V_1	: Inlet valve
V_2, V_3, V_4	: Circulation valve
V_5	: Outlet valve

2.3. Analysis of COD, MLVSS, and Biogas

Chemical Oxygen Demand (COD) is a research parameter that represents the organic concentration, while Mixed Liquor Volatile Suspended Solid (MLVSS) represents the concentration of microorganisms inside the bioreactor. COD analysis was carried out using the closed reflux titrimetric method, referring to SNI 6989.73:2009, and MLVSS was tested using the gravimetric method, referring to SNI 06-6989.3-2004. All parameters were measured in duplicate. The volume of biogas formed was measured in a closed cylindrical gas device where the middle to bottom part of the cylindrical gas device containing water was immersed in an open-air container containing water (D in Fig. 1). The volume of biogas produced was

https://doi.org/10.31436/iiumej.v25i2.2978

calculated from measuring changes in water level in the cylindrical gas column. This method adopted the biogas volume measurement method outlined by [12].

2.4. Mathematical Modeling

The growth rate of anaerobic microorganisms (μ) was arranged based on cell mass balance and the specific growth rate following the Monod model. The Monod equation is defined by Eq. (1) where S is the substrate concentration (mg/L). In this study, it is measured as COD concentration (C_{COD}).

$$\mu = \frac{\mu_m . S}{K_{SX} + S} \tag{1}$$

The Monod equation for the growth rate of anaerobic bacteria is described in equation (2).

$$\frac{dX}{dt} = \left(\frac{\mu_m . C_{COD}}{K_{SX} + C_{COD}}\right) . X \tag{2}$$

The mathematical model for an anaerobic bioreactor for tofu wastewater for the batch system was defined from the mass balance on the substrate (COD) and product (CH_4). This mathematical model is adapted to the model used by [11] and combined with the model developed by [13].

The substrate (COD) mass balance is written as follows:

[rate of accumulation] = [rate of mass input] - [rate of mass output] + [rate of formation] - [rate of consumption]

$$V \frac{d(C_{COD})}{dt} = 0 - 0 + 0 - V . \frac{1}{Y_{X/CCOD}} \frac{dX}{dt} \tag{3}$$

Divided by the volume (V) and substituting the equation (2), the equation obtained is as follows:

$$\frac{d(C_{COD})}{dt} = -\frac{1}{Y_{X/CCOD}} \left(\frac{\mu_m . C_{COD}}{K_{SX} + C_{COD}}\right) . X \tag{4}$$

The product (CH_4) mass balance is written as:

[rate of accumulation] = [rate of mass input] - [rate of mass output] + [rate of formation] - [rate of consumption]

$$V \frac{d(C_{CH4})}{dt} = 0 - 0 + V . \frac{1}{Y_{X/CCH4}} \frac{dX}{dt} - 0 \tag{5}$$

Divided by the volume (V) and substituting the equation (2), the equation obtained is as follows:

$$\frac{d(C_{CH4})}{dt} = Y_{CCH4/X} \left(\frac{\mu_m . C_{COD}}{K_{SX} + C_{COD}}\right) . X \tag{6}$$

Nomenclature

C_{COD}	Chemical oxygen demand	mg/L
C_{CH4}	Methane concentration	mg/L
X	Cell concentration	mg/L
μ_m	Maximum specific growth rate	1/day
K_{SX}	Half-saturation constant	mg COD/mg cell
$Y_{X/CCOD}$	Yield of cell formation per mg COD reduction	mg cell/mg COD
$Y_{CCH4/X}$	Yield of CH_4 formation per mg cell	mg CH_4/mg cell

https://doi.org/10.31436/iiumej.v25i2.2978

The simultaneous ordinary differential equations were solved by fitting the experimental data using MATLAB software with the use of ODE15S and FMINCON solvers resulting in the kinetic constants of the process. The effect of water hyacinth addition to the anaerobic bioreactor of tofu wastewater was shown on the experimental data and on the kinetic constants obtained from the model.

3. RESULTS AND DISCUSSION

The initial characteristics of all materials used in this study are shown in Table 1. The C/N ratio of tofu wastewater was very low, at 5.90, far below the optimum C/N ratio range for anaerobic processes (25-30) [5,7]. In addition, the optimum C/N ratio depends on the substrate characteristics and process operating conditions.

Table 1. Characteristics of raw materials

Parameters	(1)	(2)	(3)	(4)
T (°C)	27	35	25	25
pH	6.80	5.45	6.25	5.75
COD (mg/L)	8,538.3	10,267.6	8,106.4	8,646.4
C/N	2.40	5.90	13.27	10.49

Annotation:
(1): Cow rumen, (2): Tofu wastewater, (3): Water hyacinth extract, (4): The mixture of tofu wastewater and water hyacinth extract

The mixture of tofu wastewater and water hyacinth extract was proposed to increase the C/N ratio of the bioreactor substrates. Tofu wastewater and water hyacinth extract were mixed with a volume ratio of 5:3, according to the research outlined in [1]. However, based on Table 1, the actual C/N ratio of the mixed substrate was 10.49. This might be caused by the water hyacinth extract used in this study, which contained a low ratio of C/N, which was only 13.27.

A low C/N ratio increases the pH value inside of the bioreactor. This is caused by the release of ammonia due to the large amount of nitrogen contained in the substrate. Apart from that, alkaline conditions higher than 8.5 have the potential to poison anaerobic bacteria, especially acidogenic bacteria. The pH value should be in the range of 7.0 to 7.5 to increase methane formation. Low pH values (<7.0) inhibit methane yield because alkalinity is not sufficient to buffer the production of volatile fatty acids that inhibit methanogenic activity [14].

3.1. Seeding Process

The seeding process, the first stage of this study, was carried out for 26 days. The main focus of this process was to cultivate anaerobic bacteria in the bioreactor. This stage is very important to determine the optimum process for the entire anaerobic treatment. Glucose as an ideal substrate was used because it has a high carbon content and is easily degraded by anaerobic bacteria. Organic matter measured as COD in the anaerobic bioreactor was maintained in the range of 6,000-8,500 mg/L. Very low concentrations of organic matter were avoided to ensure the availability of nutrients required for the growth of anaerobic bacteria. The COD concentration was set to a moderate value to prevent organic shock loads of the bacteria. In addition to the ideal media, operating conditions were also set to ideal conditions at this stage. During the seeding process, the pH conditions of the bioreactor were maintained in the range of 6.28-7.03 by adding sodium bicarbonate (Na_2CO_3) as a buffer solution. The growth of anaerobic bacteria is represented by the concentration of Mixed Liquor Volatile Suspended Solid (MLVSS), as shown in Fig.2.

https://doi.org/10.31436/iiumej.v25i2.2978

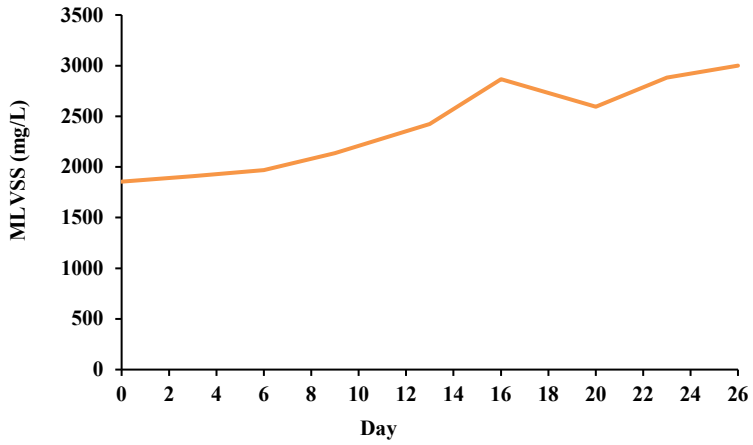


Fig. 2. MLVSS concentration versus time during the seeding process.

As seen in Fig. 2, the MLVSS concentration rose from 1,855 mg/L at the beginning of the seeding process to 3,000 mg/L on day 26. This means that the concentration of anaerobic bacteria met the minimum concentration required in the anaerobic process for wastewater treatment. The seeding process was stopped on the 26th day with a COD value of 6,755 mg/L and the MLVSS at 3,000 mg/L.

3.2. Acclimatization Process

The acclimatization process is referred to as the adaptation phase of microorganisms to process real organic matter contained in wastewater [15]. The addition of glucose prevented the risk of organic shock loads for microorganisms due to changes in the ideal substrate to the wastewater mixture [1]. Mixed substrates are more complex and difficult to degrade than glucose. After gradually replacing glucose with tofu wastewater, starting on day 16, the acclimatization process only used a mixture of tofu wastewater and water hyacinth until the end of the process on day 20, as shown in Fig. 3. Overall observations were made on the parameters of COD, MLVSS, and the volume of biogas formed.

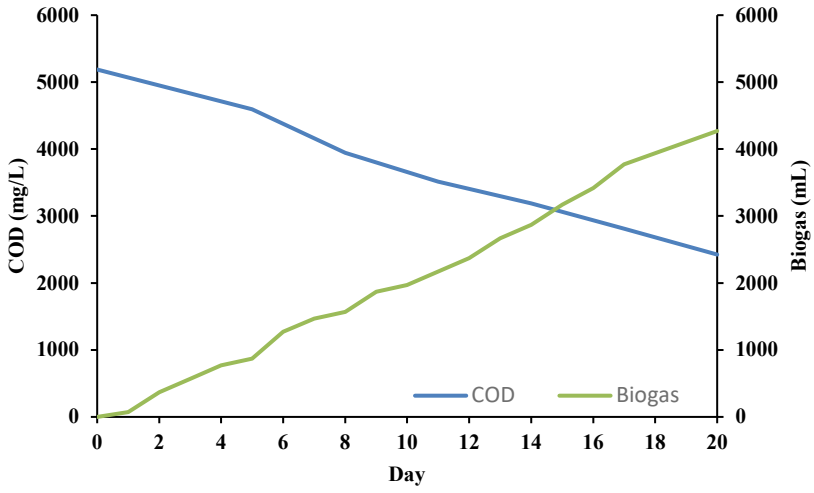


Fig. 3. COD concentration and biogas production during the acclimatization process

Fig. 3 shows that the COD concentration decreased progressively since the beginning of the acclimatization process from 5.187 mg/L to 2.424 mg/L at the end of the process. This

https://doi.org/10.31436/iiumej.v25i2.2978

shows that the anaerobic bacteria that were cultured in the previous stage were able to properly degrade complex organic compounds contained in the liquid waste. Reducing COD concentration is one of the determining factors for the success of the acclimatization process as evidenced by the presence of biogas formed, which was measured starting on day 1 and continued to increase until day 20, namely 4,270 mL of cumulative biogas (Fig. 3). The decrease in COD concentration and the formation of biogas indicate that all stages of the anaerobic process with the main stages of acidogenesis and methanogenesis processes were successfully conducted [16].

3.3. Main Anaerobic Process

The tofu wastewater anaerobic bioreactor was operated at an HRT of 20 days with an organic loading rate of 6,000 mg COD/L/day of fresh tofu wastewater. This tofu wastewater was fed into the bioreactor every day. At this stage, several research parameters were measured to be observed qualitatively, including pH (Fig. 4), MLVSS (Fig. 5), COD (Fig. 6), and methane formation (Fig. 7). These parameter data were used to solve the mathematical model presented in Eq. (2), Eq. (4) and Eq. (6) using MATLAB software and to generate calculation data based on model simulations. In addition, the kinetic constants obtained from the mathematical model were evaluated quantitatively.

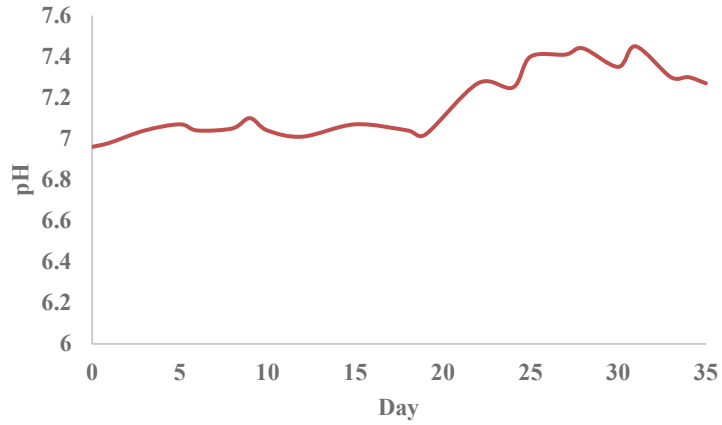


Fig. 4. pH measurements during anaerobic treatment

Fig. 4 provides bioreactor pH data during the main process. The bioreactor was run with an initial pH of 6.8, which is in the optimum pH range for anaerobic degradation. This stable condition was maintained until day 20. On day 31, the pH gradually increased to 7.5, indicating that the methanogenic process was more dominant than the acidogenic process.

Fig. 5 shows the measured microorganism concentrations obtained from research data (MLVSSdata) and also the predicted concentrations from mathematical calculations (MLVSScal). It can be seen that the concentration of microorganisms swiftly rose from 1,067 mg/L to 2,649 mg/L. This shows that the addition of water hyacinth to the mixed substrate played a positive role in supplying additional nutrients needed by microorganisms in the anaerobic digestion process, especially in processing waste with low C/N content. This is reinforced by the growth rate constant (μ_m) obtained, which was equal to 0.65/day, as shown in Table 2. This growth rate constant was much higher and comparable to the growth rate constant for several anaerobic processes in general, namely 0.17 – 0.25/day [17] and 0.21 – 1.54/day [18].

https://doi.org/10.31436/iiumej.v25i2.2978

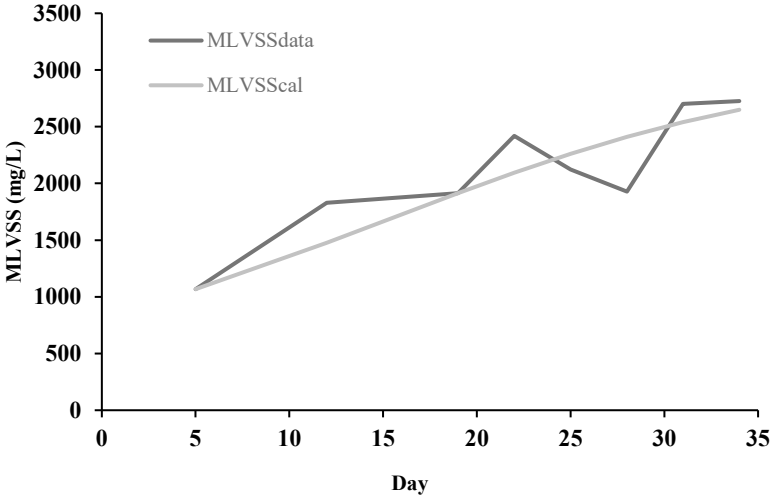


Fig. 5. Comparison of data and calculation of MLVSS concentrations

Table 2. Kinetic constants of anaerobic bioreactor

Constants	Value	Unit
μ_m	0.65	1/day
K_{SX}	9.62	mg COD/mg cells
$Y_{X/CCOD}$	0.64	mg cells/mg COD
$Y_{CCH4/X}$	0.62	mg CH_4/mg cells

An increase in MLVSS concentration was followed by a significant decrease in pollutant content in wastewater (Fig. 6). By maintaining the substrate concentration at 6,000 mg/L, anaerobic digestion with the addition of water hyacinth was able to reduce the COD concentration by 92.82% on day 35 with a final concentration of 431 mg/L. The kinetic constant associated with this decrease is $Y_{X/CCOD}$ which is equivalent to 0.64 mg cells/mg COD (Table 2).

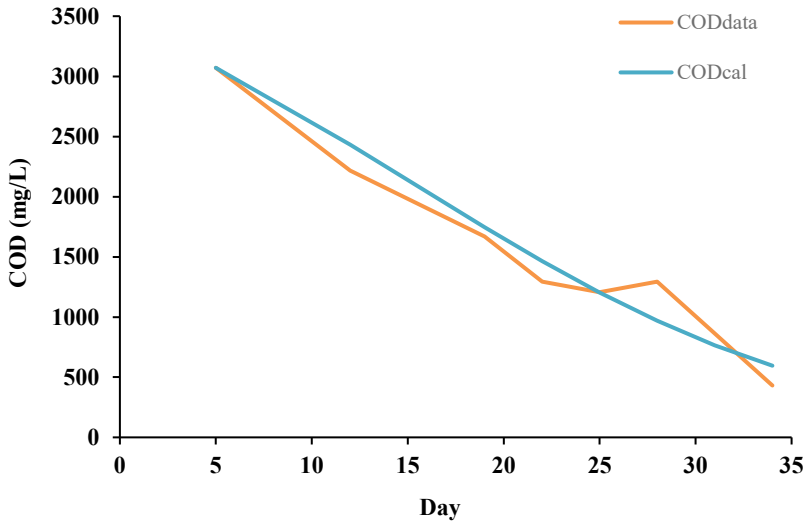


Fig. 6. Comparison of data and calculation of COD concentrations

https://doi.org/10.31436/iiumej.v25i2.2978

This constant ($Y_{X/CCOD}$) represents the percentage of COD consumed for cell formation. Based on Eq. (4), a decrease in $Y_{X/CCOD}$ resulted in a higher percentage reduction in COD. A study on the anaerobic process to treat wastewater with low C/N (Piyungan sanitary landfill leachate) was completed by [10] and resulted in a COD reduction of 62% with $Y_{X/CCOD}$ of 0.783-0.91 mg cells/mg COD. The total COD removal resulting from this research is higher than the result obtained by [10]. This shows that water hyacinth augmentation has contributed to providing a favorable atmosphere for the growth of microorganisms by providing the essential nutrients needed. Conducive anaerobic conditions were also supported by the pH of the bioreactor, which can be maintained in the range of 6.8-7.5. These factors have an impact on the high conversion of COD into fatty acids resulting from the acidogenic process, compared to the use of organic materials for the survival of microorganisms. Therefore, an aerobic system can be one of the advanced processing recommendations for completing anaerobic processing when the main concern is to meet waste standards before being discharged into the environment [1].

A comparison of the methane gas formed based on research data and prediction data is shown in Fig. 7. This figure presents the methane gas formed as the main product of the methanogenic process. The model predicts a constant increase in methane formation. These results are in accordance with experimental data up to day 27 of the experiment. Methane rose rapidly to 1,431 mL/g COD on day 30.

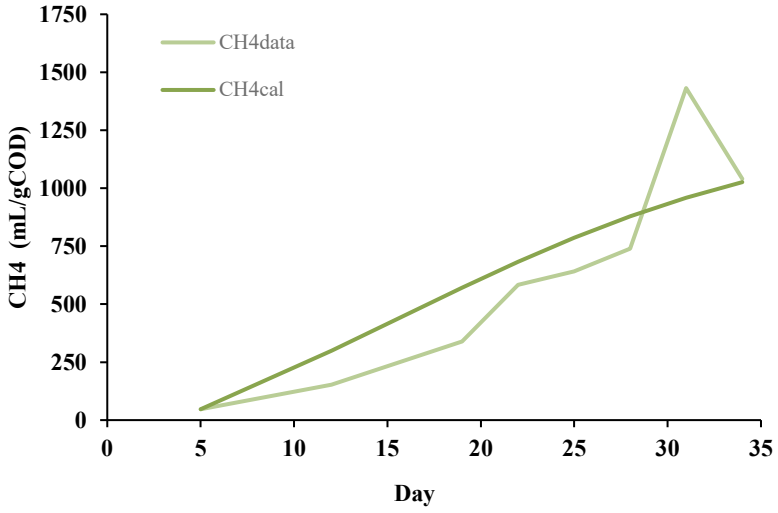


Fig. 7. Comparison of data and calculations of methane formation

The kinetic constant associated with this formation is $Y_{CCH4/X}$ (mg CH_4/mg cell), which is 0.62 mg CH4/mg cells, as illustrated in Table 2. This constant controls the amount of CH_4 produced in the bioreactor [10]. From Eq. (6), it can be concluded that increasing $Y_{CCH4/X}$ has an impact on increasing the percentage of methane formation. The highest methane concentration achieved was 56.9% on day 21. The $Y_{CCH4/X}$ resulting from this research was much higher than the $Y_{CCH4/X}$ for several anaerobic processes with the carbon-poor wastewater, namely 0.14 mg CH_4/mg cells [17], 0.27 mg CH_4/mg cells [19] and 0.35 mg CH_4/mg cells [20]. The appropriate ratio of water hyacinth as a high carbon source and tofu wastewater, which is rich in nitrogen, is recommended in subsequent research to obtain an optimal carbon-to-nitrogen ratio, resulting in a higher percentage of methane formation.

4. CONCLUSION

The use of water hyacinth as an additional substrate has great potential in improving bioreactor performance through qualitative evaluation based on research data with a COD reduction percentage of 92.8% and a methane concentration percentage of 56.9%. These results are supported by the kinetic constants resulting from this research which show better performance compared to previous similar anaerobic digestion using low C/N wastewater, namely microbial growth rate (μ_m), substrate degradation ($Y_{X/CCOD}$) and CH_4 production ($Y_{CCH4/X}$) with respective values of 0.65/day, 0.64 mg cells/mg COD and 0.62 mg CH_4/mg cells. These results complete the evaluation quantitatively.

ACKNOWLEDGEMENT

Financial support of this study and equipment provided by Politeknik Negeri Bandung through the Applied Research scheme no. B/186.55/PL1.R7/PG.00.03/2020 is gratefully acknowledged.

REFERENCES

[1] Budiastuti H, Amanah YN, Pratiwi RO, Muhari EH, Ramadhani LI. (2020) Water hyacinth extract addition towards tofu wastewater degradation by anaerobic treatment. In Proceedings of the International Conference on Innovation in Science and Technology (ICIST), Semarang, Indonesia, 2020; pp 341-345. doi: 10.2991/aer.k.211129.072.

[2] Budiastuti H, Rismawati R, Nurfauziah L, Ramadhani LI. (2021) Tofu wastewater treatment by using sequencing batch reactor (SBR) with variation of feeding rates. Rekayasa Hijau: Jurnal Teknologi Ramah Lingkungan, 5(3):197-206. doi: 10.26760/jrh.V5i3.197-206.

[3] Peraturan Menteri Lingkungan Hidup Republik Indonesia No.5 Tahun 2014 tentang Baku Mutu Air Limbah (2014). Regulation of the Minister of Environment of the Republic of Indonesia No. 5 of 2014 concerning Wastewater Quality Standards (2014). https://lhk.acehtamiangkab.go.id/downloads/permen-lh-05-tahun-2014-tentang-baku-mutu-air-limbah.

[4] Carotenuto C, Guarino G, D'Amelia LI, Morrone B, Minale M. (2019) The peculiar role of C/N and initial pH in anaerobic digestion of lactating and non-lactating water buffalo manure. Waste Manag., 103:12–21. doi: 10.1016/j.wasman.2019.12.008.

[5] Tanimu MI, Ghazi TIM, Harun RM, Idris A. (2014) Effect of carbon to nitrogen ratio of food waste on biogas methane production in a batch mesophilic anaerobic digester. Int. J. of Innovation, Manag. and Techno., 5:116-119. doi: 10.7763/IJIMT.2014.V5.497.

[6] Sutaryo S, Sempana AN, Mulya RM, Sulistyaningrum D, Ali MS, Damarjati RI, Purbowati E, Adiwinarti R, Purnomoadi A. (2022) Methane production of pistia stratiotes as a single substrate and as a co-substrate with dairy cow manure. Fermentation, 8(736): 1-9. doi: 10.3390/fermentation8120736.

[7] Omondi EA, Gikuma-Njuru P, Ndiba PK. (2019) Anaerobic co-digestion of water hyacinth (e. crassipes) with ruminal slaughterhouse waste for biogas production. Int. J. of Renew. Energy Dev. 8(3): 253-25. doi: 10.14710/ijred.8.3.253-259.

[8] Hadiyanto H, Octafalahanda FM, Nabila J, Jati AK, Christwardana M, Kusmiyati K, Khoironi A. (2023) Preliminary observation of biogas production from a mixture of cattle manure and bagasse residue in different composition variations. Int. J. of Renew. Energy Dev., 12(2): 390-395. doi: 10.14710/ijred.2023.52446.

[9] Munkar G, Syafrudin, Nugraha WD. (2017) Pengaruh C/N ratio pada produksi biogas dari daun enceng gondok dengan metode liquid Anaerobic digestion (L-AD). Effect of C/N ratio on biogas production from water hyacinth leaves using the liquid anaerobic digestion (L-AD) Method. Jurnal Teknik Lingkungan, 6(3): 1-8.

[10] Sudibyo H, Shabrina ZL, Halim L, Budhijanto W. (2017) Mathematical modelling and statistical approach to assess the performance of anaerobic fixed bed reactor for biogas production from

piyungan sanitary landfill leachate. Energy Procedia, 105: 256-262, doi: 10.1016/j.egypro.2017.03.311.
[11] Ramadhani LI, Damayanti SI, Sudibyo H, Budhijanto W. (2018) Kinetics of anaerobic digestion of palm oil mill effluent (POME) in double-stage batch bioreactor with recirculation and fluidization of microbial immobilization media. IOP Conference Ser., Mater. Sci. and Eng., 316, 012071; pp 1-9. doi: 10.1088/1757-899X/316/1/012071.
[12] Walker M, Zhang Y, Heaven S, Banks C. (2019) Potential errors in the quantitative evaluation of biogas production in anaerobic digestion processes. Bioresour. Technol., 100: 6339-6346. doi: 10.1016/j.biortech.2009.07.018.
[13] Pramashinta A, Abdullah A. (2014) Kinetika fermentasi limbah kulit nanas dan produktivitas etanol. Kinetics of pineapple peel waste fermentation and ethanol productivity. Metana, 10(1): 12-17. doi: 10.14710/metana.v10i01.9772.
[14] Jiang Y, Dennehy C, Lawlor PG, Hu Z, McCabe M, Cormican P, Zhan X. (2018) Inhibition of volatile fatty acids on methane production kinetics during dry co-digestion of food waste and pig manure. Waste Manag., 79: 302-311. doi:10.1016/j.wasman.2018.07.049.
[15] Budiastuti H, Ramadhani LI, Abdulloh SH, Maharani AD, Melina M, Sudarman R. (2023) Seeding and acclimatization for aerobic processing of restaurant wastewater with sequencing batch reactor. Fluida 16(1): 11-17. doi: 10.35313/fluida.v16i1.4521.
[16] Zuo X, Yuan H, Wachemo AC, Wang X, Zhang L, Li J, Wen H, Wang J, Li X. (2020) The relationships among sCOD, VFAs, microbial community, and biogas production during anaerobic digestion of rice straw pretreated with ammonia. Chinese J. of Chem. Eng., 28(1): 286-292. doi: 10.1016/j.cjche.2019.07.015.
[17] Sudibyo H, Shabrina ZL, Wondah HR, Hastuti RT, Purnomo CW, Budhijanto W. (2018) Anaerobic digestion of landfill leachate with natural zeolite and sugarcane bagasse fly ash as the microbial immobilization media in packed bed reactor. Acta Polytechnica, 58(1): 57-68. doi: 10.14311/AP.2018.58.0057.
[18] Maleki E, Bokhary A, Liao BQ. (2018) A review of anaerobic digestion bio-kinetics. Rev. Environ. Sci. and Biotechnol., Suppl:1-15. doi: 10.1007/s11157-018-9484-z.
[19] Fedailaine M, Moussi K, Khitous M, Abada S, Saber M, Tirichine N. (2015) Modeling of the anaerobic digestion of organic waste for biogas production. Procedia Computer Sci., 52(1): 730-737. doi: 10.1016/j.procs.2015.05.086.
[20] Fernandez N, Montalvo S, Borja R, Guerrero L, Hernandez EPS, Cortes I, Colmenarejo MF, Travieso L, Raposo F. (2008) Performance evaluation of an anaerobic fluidized bed reactor with natural zeolite as support material when treating high-strength distillery. Renew. Energy, 33: 2458-2466. doi: 10.1016/j.renene.2008.02.002.

https://doi.org/10.31436/iiumej.v25i2.2024

FUNCTIONAL METAGENOMICS APPROACH FOR DISCOVERY OF NOVEL COLD-ACTIVE PROTEASE FROM ANTARCTIC REGION

MUHAMMAD ASYRAF ABD LATIP[1,2], NOOR FAIZUL HADRY NORDIN[3*], SITI AISYAH ALIAS[4,5], JERZY SMYKLA[6], FARIDAH YUSOF[2], MOHD AZRUL NAIM MOHAMAD[7]

[1]Langkawi Mariculture Research Center, Kompleks Perikanan Bukit Malut, 07000 Langkawi, Kedah, Malaysia
[2]Department of Chemical Engineering & Sustainability, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur
[3]International Institute for Halal Research and Training (INHART), Block A, Level 3, KICT Building, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur
[4]Institute of Ocean and Earth and Sciences, C308, Level 3, Block C, Institute for Advanced Studies Building, 50603 Kuala Lumpur, Malaysia
[5]National Antarctic Research Centre, B303, Level 3, Block C, Institute for Advanced Studies Building, 50603 Kuala Lumpur, Malaysia
[6]Institute of Nature Conservation, Polish Academy of Sciences, al. A. Mickiewicza 33 PL-31-120 Krakow, Poland
[7]Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia

**Corresponding author: faizul@iium.edu.my*

(Received: 5 November 2023; Accepted: 12 March 2024; Published online: 15 July 2024)

ABSTRACT: The structural complexity of bacterial life makes most of it impossible to culture. Functional metagenomics approaches overcome the limitations of a culture-based approach in exploring and assessing the genetic materials of uncultured microbes. The objective of this study was to identify clones with cold-active proteases through functional metagenomics. In this work, the environmental DNA (eDNA) isolated directly from Antarctic soils was ligated into the pCC1FOS fosmid vector, transformed into EPI300-T1R *E. coli* host cells, and screened for proteolytic enzymes. Positive protease-producing clones were identified and isolated on skim milk agar supplemented with chloramphenicol and arabinose. This clone harbored a fosmid, pCC1FOS, which has a 48.5 kb insert that has been completely sequenced in both directions. Further analysis of the insert showed 70 NODEs. The NODE_42 encoded hypothetical protein of 297 amino acids showed a significant match to Peptidase M23 and PG-binding 1 proteins families. A three-dimensional model of the predicted protease was generated based on the known mesophilic protease of *Neisseria meningitides* (PDB: 3SLU). The structural alignment showed 27.07 % similarity with RMSD value of 0.402 Å based on 58 aligned residues. The active site residues were identical, but major deletions were observed in the predicted proteases. This predicted protease showed higher activity at -20 °C and 20 °C than the positive control (protease from bovine pancreas). Functional metagenomics is a promising approach in the discovery of cold-active protease with low homology to the known sequences and expressed in the host cell that has the potential for bioprospecting in low-temperature applications.

ABSTRAK: Kesukaran struktur kehidupan bakteria menyebabkan kebanyakan bakteria sukar dikultur. Saringan fungsi metagenomik dapat mengatasi kekangan saringan berasaskan kultur dalam meneroka dan menilai bahan genetik mikrob tidak kultur. Objektif kajian ini adalah

bagi mengenal pasti klon protease aktif sejuk melalui saringan fungsi metagenomik. Kajian ini mencadangkan DNA alam sekitar (eDNA) yang diasingkan secara langsung dari tanah Antartika, dan dimasukkan ke dalam vektor fosmid pCC1FOS, ditransformasikan ke dalam sel hos EPI300-T1R E. coli, dan disaring untuk enzim proteolitik. Klon yang menghasilkan protease ini dikenal pasti dan diasingkan daripada agar susu skim. Agar telah ditambah dengan kloramfenikol dan arabinosa. Klon ini mengandungi fosmid pCC1FOS, di mana sisipan sebesar 48.5 kb telah disusun berturut-turut sepenuhnya dari kedua-dua arah. Analisis lanjut pada sisipan ini menunjukkan terdapat 70 NODE. NODE_42 mengandungi 297 asid amino protein hipotetis, iaitu sangat serasi dengan protein Peptidase M23 dan PG-binding 1. Model tiga dimensi protease yang diramalkan ini dihasilkan berdasarkan protease mesofilik Neisseria meningitides (PDB: 3SLU). Susunan struktur menunjukkan 27.07% persamaan nilai RMSD sebanyak 0.402 Å berdasarkan 58 sisa sejajar. Sisa tapak aktif kedua-duanya adalah sama, tetapi terdapat pemotongan besar telah berlaku dalam protease ramalan. Protease ini menunjukkan aktiviti enzim lebih tinggi pada suhu -20 °C dan 20 °C berbanding kawalan positif (protease dari pankreas lembu). Saringan fungsi metagenomik berhasil menemukan protease aktif sejuk berhomologi rendah terhadap susunan yang diketahui dan diekspresi dalam sel hos. Ia berpotensi bagi bioprospeksi dalam aplikasi bersuhu rendah.

KEYWORDS: *metagenomics; cold-active; enzyme; protease; Antarctic*

1. INTRODUCTION

Metagenomics study was encouraged with the escalation of next-generation sequencing (NGS) technologies such as 454 pyrosequencing, Illumina, Sequencing by Oligonucleotide Ligation and Detection (SOLiD), Helicos Genetic Analysis Platform, Ion Torrent, PacBio (SMRT) and Oxford Nanopore sequencing. First-generation sequencers, Sanger or Maxam-Gilbert sequencing, were not capable of sequencing this eDNA due to its huge size. NGS will generate massive post-sequencing data and require bioinformatic skills for processing and analyzing. This is an approach of a direct route for analyzing the DNA from the sample. The DNA or genetic materials are extracted directly from the environment sample, and this DNA is called environmental DNA (eDNA) [1]. This powerful study has the capability for the discovery of specific genes from both culturable and non-culturable bacteria [2,3]. The limitation of culture-based screening restricts the exploration of the potential enzymes or genes because less than 1% of the total bacterial community can be cultivated in standard media [4]. Most of the bacteria are uncultivable because of their natural complexity.

Functional metagenomics is an approach to analyzing the whole genome for novel bioactivity. The main objective of functional metagenomics sequencing is to screen and identify novel enzymes or genes for a specific reaction. The advantages of this technique are that the researchers do not need to have prior knowledge about the gene sequence and do not have to depend on the known gene sequence from previous studies to identify the targeted enzyme. Thus, the culture-based method is important in the process to minimize false-positive errors. Prior to this approach, a more novel gene that is responsible for functional enzymatic expression was discovered. Many novel enzymes have been discovered via functional metagenomics, such as beta-lactamase [5], β-galactosidases [6], serine protease [7], cellulases [8], endoglucanases [9], phosphatases and phytases [10]. The novel enzymes can be defined when the similarity of gene homology is lesser (< 70%) compared to the known sequence [11]. A new characteristic of the enzyme can be identified as to whether it has the valuable potential for industrial and biotechnological application.

In addition, there are limited functional metagenomics studies that have been performed on the Antarctic regions. Although these cold-active enzymes have been discovered previously

within other extreme cool regions or the Arctic, unique characteristics and behaviors of the enzyme under certain conditions are important to be analyzed and applied in a real-life situation. Based on the findings, only five novel enzymes have been discovered from the Antarctic region via the functional metagenomics approach: lipase and esterase [12], esterase [13], lipase [14], chitinase [15] and alkane monooxygenase [16]. There are still a lot more opportunities to be discovered in this region, such as new species and novel enzymes. The objective of this study is to discover cold-active protease from Antarctic soil samples through the functional metagenomics approach and the scope of this study focused on the computational and prediction level.

2. MATERIALS AND METHODS

2.1. Environmental DNA (eDNA) Extraction

Soil samples obtained from the previous study were collected along with the coastal environments of the Southern Victoria Land during the Antarctic summer 2011/2012 [17]. Five grams of soil sample were weighed and transferred into the 10 mL elution buffer (4 % cetyltrimethylammonium bromide (CTAB), 1.5 M sodium chloride (NaCl), 0.1 M tris, 25 mM ethylenediaminetetraacetic acid (EDTA) and 1 % polyvinylpyrrolidone (PVP)). Subsequently, 0.03 mL of β-mercaptoethanol was added, and the soil was incubated at 65 °C for 30 minutes while being inversely mixed a few times. The sample was centrifuged for 5 minutes at 5,000 x g to remove the foam. The clear supernatant was transferred into a new tube, and one volume of chloroform phenol: chloroform: isoamyl alcohol was added. The sample was centrifuged at 5,000 x g for 10 minutes after inversely mixing for 5 minutes. The aqueous layer, or upper layer, was transferred into the new tube, and another volume of phenol, chloroform, or isoamyl alcohol, was added. The solution was inversely mixed for 5 minutes and centrifuged for another 10 minutes at 5,000 x g. The aqueous layer was carefully transferred to avoid contamination from the bottom layer and entering the new tube. The half volume of 5 M NaCl was added and inversely mixed. Three volumes of cold 95 % ethanol were added and inversely mixed. The sample was incubated at – 20°C for one hour. The sample was then centrifuged at the highest speed for 10 minutes. The supernatant was discarded, and the pellet DNA was washed with 3 mL of 70 % ethanol. The sample was centrifuged at 5,000 x g for 15 minutes. After discarding the supernatant, pelleted DNA was air-dried at room temperature. The air-dried DNA was suspended with 0.1 mL of TE buffer and stored at – 20°C.

2.2. Screening of Functional Expression Library

The extracted eDNA was ligated to the fosmid pCC1FOS and transfected into the host cells EPI300-T1R via a phage system. The host cell *E. coli* EPI300-T1R cell comes with the commercial kit CopyControl Fosmid Library Production Kit with pCC1FOS Vector (Epicentre, Madison, WI, USA). According to the manufacturer's protocol, the host cell or *E. coli* EPI300-T1R cell provides a mutant *trfA* gene whose gene product is required for initiation of replication from *oriV*. The cells have been engineered so that the *trfA* gene is under tight, regulated control of an inducible promoter. For this process, the CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre) was utilized according to the manufacturer's instructions. The infected EPI300-T1R cells were screened for functional gene expression on the 5 % (w/v) skim milk agar plates supplemented with 12.5 µg/mL chloramphenicol and 0.01 % arabinose. Arabinose induces the expression of the *trfA* gene product and subsequent amplification of the clones to a high copy number for direct function-based screening assays of the clone library (Epicentre, Madison, WI, USA). The clones were cultured for a week at 20 °C. After incubation, the formation of a clearing halo zone formation indicated the activity of the protease. A clone with positive proteolytic activity was isolated,

and the fosmid extraction process proceeded. The isolated clone was stored in 25 % glycerol at -80 °C. Fosmid extraction was performed using the FosmidMAX™ DNA Purification Kit (Lucigen). The purified fosmid was analyzed by 1 % agarose gel electrophoresis. Then, 50 µL of the suspended DNA was sent for sequencing services through Apical Scientific Sdn Bhd.

2.3. Metagenomics Data Analysis

Paired-end Illumina sequences were first removed from sequence adaptors and reads with low-quality scores using bbduk of the BBTools Packages. All QC-ed reads were filtered against the fosmid sequence (pCC1FOS). Unmapped reads (both paired-end and unpaired reads) were then assembled using SPAdes 3.11.1 [18]. We employed a two-prong approach to identify potential contigs associated with the fosmid insert sequence. First, all contigs >5000 bp were subjected to Principal Component Analysis of their tetranucleotide frequency [19]. Next, all contigs were filtered using BLASTN (maximum E value of 10) against a database of *Escherichia coli* consisting of AP021894, CP025520, NC_000913. The contigs within NODE_42 were subsequently annotated using Rapid Annotations using Subsystems Technology (RAST) [20].

2.4. *In silico* Analysis

The percentage of amino acid proportion was calculated using the GPMAW Lite program (Alphalyse, Palo Alto, CA, USA). Domains and active sites of the enzymes were searched through the Pfam database and NCBI Conserved Domain Search [21]. The protein sequences were aligned with the program ENDscript, and the secondary structures were compared with those of the program ESPript [22]. PyMOL calculated the RMSD values of psychrophilic and mesophilic enzymes after performing structure superimposition [23]. The structural differences between both enzymes were analyzed when aligning both enzymes together.

2.5. Protease Assay of Crude Enzyme

For the enzymatic assay, 250 µL of the crude enzyme, 250 µL of 1% Azocasein and 250 µL of 0.5 M Tris were incubated at -20 °C, 20 °C and 60 °C. The clone was incubated in 10 mL of skim milk broth at 200 rpm until the media became clear. Then, the culture was centrifuged at 15,000 x g for 15 minutes and the supernatant was considered as the crude enzyme. Before adding the crude enzyme, the assay was pre-incubated at mentioned temperature for 10 minutes. For positive control, 0.2 units/mL of protease from the bovine pancreas was used. After 20 minutes, 500 µL of 25% trichloroacetic acid (TCA) was added to stop the reaction and centrifuged at 11,000 x g for 30 minutes. The absorbance of the supernatant was read at 340 nm.

3. RESULTS AND DISCUSSION

3.1. Clone with Positive Proteolytic Activity

In this experiment, only one clone showed the formation of a halo zone (Fig. 1 (A)). This clone was isolated, and the fosmid was extracted and purified. Fig. 1 (B) shows the gel electrophoresis results of the digested and the undigested fosmid with DNA insertion. After the fosmid was treated with the restriction enzyme, *EcoRI*, two separated bands appeared as the upper band was the insert DNA (± 40 kb) and the lower band was the fosmid DNA (8 kb). The undigested fosmid showed a single band with a high molecular weight of more than 48.5 kb.

https://doi.org/10.31436/iiumej.v25i2.2024

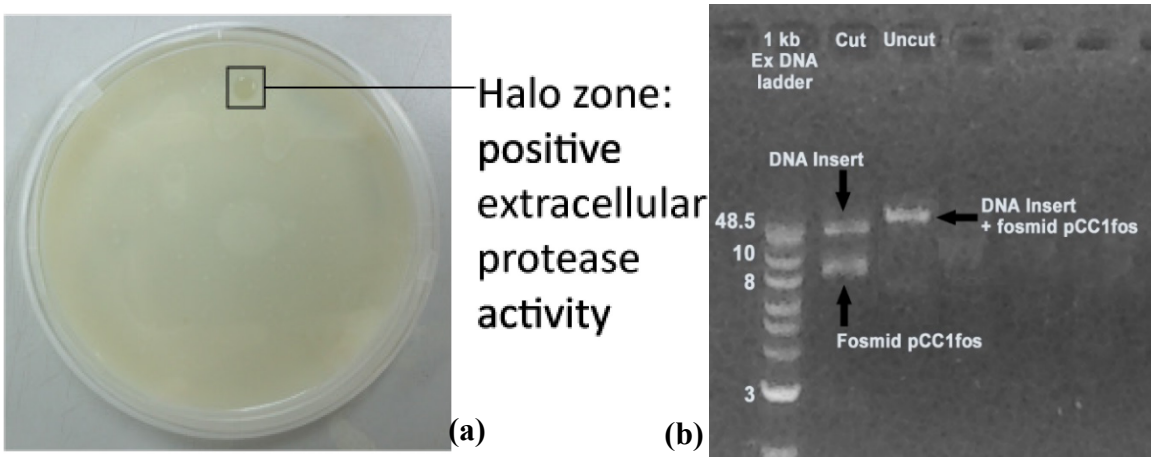


Figure 1. (a) A clone with halo zone formation on skim milk agar plate supplemented with chloramphenicol and arabinose. (b) The fosmid was analyzed by 1 % agarose gel electrophoresis. The undigested fosmid showed a single band in lane 1, while the digested fosmid with EcoR1 showed two separated bands in lane 2.

Other studies also showed that the hit rate for functional enzyme discovery through functional metagenomics was very low, as tabulated in Table 1[24]. This is because there were major bottlenecks for the unknown origin of the genes to be expressed in the host cell. Ekkers et al. (2012) have listed the causal factors for the lack of gene expression, including variations in the use of codons, inappropriate identification of promoters, lack of proper activation factors, ribosomal entry, improper folding of proteins, lack of important co-factors, rapid enzymatic degradation of the gene product, body development of incorporation, gene product toxicity, or the host's failure to secrete the product of gene expression. However, the advantage of this screening technique was that when the clone showed positive activity, it indirectly established a suitable host system for enzyme expression. Up to date, several studies successfully discovered protease through this approach, such as oxidant-stable serine protease from forest soil [7], putative proteases from slaughterhouse biofilm [25], alkaline protease from tannery-activated sludge [26] and mangrove-sediment [27]. However, some researchers hit zero rate frequency for protease activity through functional metagenomics approaches [28,29].

Table 1: Protease activity screening through functional metagenomics approach.

No. of positive clones	Vector	Total clones	Hit rate (%)	Sample	References
1	Plasmid	35,000	0.003	Forest soil	[7]
5	Plasmid	170,000	0.003	Slaughterhouse biofilm	[25]
1	Plasmid	10,000	0.010	Tannery sludge	[26]
4	Fosmid	1,824	0.219	Mangrove soil	[27]
0	Plasmid	32,000	0.000	Garden soils and compose	[28]
0	BAC	2,843	0.000	Ikaite columns	[29]

3.2. Metagenomics Data Analysis

From the shotgun metagenomic sequencing, the total number of reads for the raw data obtained for both forward and reverse sequences was 55,923,089 reads. After removing the sequence adaptors and low-quality reads, the total high-quality reads were reduced to 55,673,161 for both forward and reverse sequences. After assembling the reads using SPAdes, 70 NODEs were identified in the sample. NODE or scaffolds reconstructed the DNA sequences

based on the contigs and bridged via the gap [30]. Contigs were generated after the DNA reads (k-mers) were assembled and sequenced according to the adjacent overlapping of the DNA base. Initially, all the scaffolds were analyzed using the Principal Component of Analysis (PCoA) of their tetranucleotide frequency. Accumulation of the NODEs at the center of the graph meant that the NODEs have similarities in their tetranucleotide pattern between each other. These scaffolds were majorly subjected to the host cell, *E. coli*. The targeted scaffolds were outlying the eclipse area where these were potential candidates for fosmid with the insertion. From the graphs in Fig. 2, the selected scaffolds were NODE_42, NODE_55, NODE_64, NODE_67 and NODE_69.

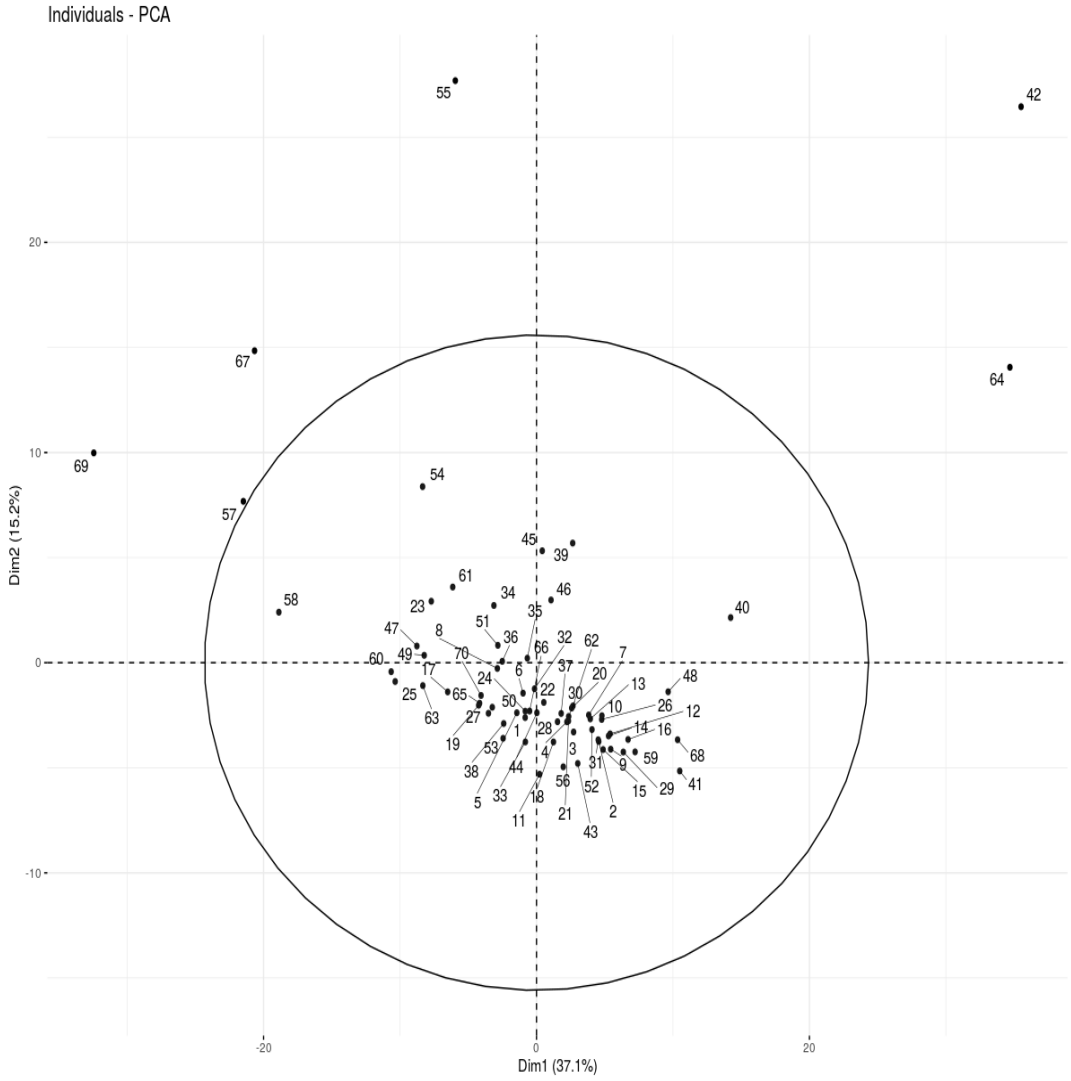


Fig. 2. The PCoA-tetranucleotide frequency of the assembled contigs. Contigs (NODE) outside of the eclipse were potential candidates for the fosmid insert.

Based on the BLASTN results, all the nodes showed high similarity to the *E. coli* except for NODE_42. Ambiguously, NODE_55, NODE_64, NODE_67 and NODE_69 have different tetranucleotide patterns to the *E. coli* genome based on the PCoA result because these scaffolds were part of mobile genetic elements (MGEs) and were expected to have differential patterns in their nucleotide composition as illustrated in Fig. 3. MGEs were genetic material that can move around the genome and also can be transferred among bacterial genomes. MEGs carried noncore genes and provided important contributions to bacterial adaptation to the environment,

such as antibiotic-resistant genes in *E. coli* [31]. The candidate NODE_42 showed similarity to the cloning vector pCC1FOS, but with less percentage query coverage, 1.04 %.

Figure 3. Alignment of NODE_67 to a fragment of the E. coli reference genome where mobile genetic elements (MGEs) were detected in this region. NODE_55, NODE_64 and NODE_6 also showed the same results.

3.3. *In silico* Analysis and Enzymatic Assay of Predicted Protease

Further analysis was focused on NODE_42 as this was the only candidate for DNA insertion. After trimming the fosmid sequence, the total length of the insertion was 34493 bp. As expected, the DNA insert size length should be between 20 kb to 40 kb as these were the range sizes of gDNA extracted from agarose gel. This scaffold NODE_42 contained 56 contigs of DNA with an average size of 572 bp with the lowest and the highest size was 104 bp and 2648 bp, respectively. From the RAST results, the majority of the contigs were hypothetical proteins. In the Pfam database, the function of the contigs was predicted through the conserved part of the protein sequence and tertiary structure called domain. In this experiment, contigs of NODE 42 (893 bp) showed significant match to Peptidase M23 and PG binding 1 protein families based on both Pfam and NCBI databases as illustrated in Fig. 4 (A) and (B), respectively. According to on Pfam database, the family members of peptidase M23 (PF01551) were generally categorized as endopeptidases. Specifically, zinc metallopeptidases with several specificities and Gly-Gly endopeptidases were included in this family. Based on the MEROPS, the activities of this peptidase family were cleaving the cross-linking peptide bonds either between or within the cell wall of peptidoglycan [32]. Besides, this family member has a propensity to cleave glycine-rich substrate, such as elastin.

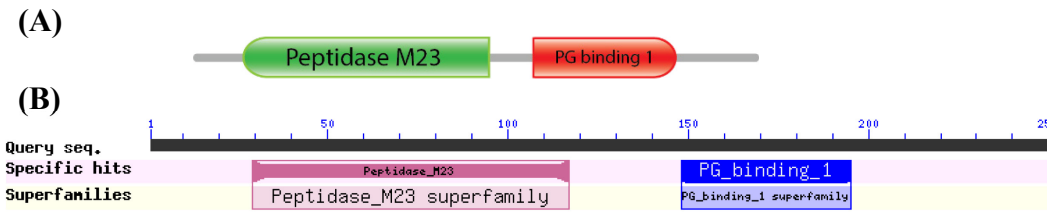


Figure 4: (A) Pfam result analysis showed two domain sections, Peptidase_M23 (green) and PG_binding_1 (red). (B) Analysis of the conserved domain through NCBI showed that the selected contig's protein sequence hit specific domain Peptidase M23 and PG binding 1 superfamily.

This contig with 893 bp was translated into a 297 amino acid sequence for further analysis. Based on this protein sequence, the Swiss-model tool constructed the structure of the predictive homology model based on the M23 peptidase domain template with PDB ID 3SLU. This template protein was characterized as a proteolytic enzyme located at the outer membrane of *Neisseria meningitidis* [33]. The percentage composition of an amino acid between the predictive model and the template was calculated in Table 2. In this experiment, the predictive model contained a lower arginine residue proportion, 6.06%, compared to the template enzyme, 7.82%. In contrast, the proportion of proline for the predictive enzyme was 4.71% higher than

https://doi.org/10.31436/iiumej.v25i2.2024

template enzymes, 3.77%, but the count numbers were the same, 14 residues. From this finding, it appeared that arginine residue proportion was more crucial than proline for cold adaptation in enhancing enzyme structure flexibility. In addition, there were also reports that proline residue appeared to be increased in some psychrophilic enzymes [34–36]. These psychrophilic enzymes showed thermostable characteristics, and they can work at lower to higher temperatures [37]. The flexibility of the enzyme structure could affect the stability and function of the enzyme. When the enzyme became more flexible, the conformation of the active site became easier to change and form. This allows the enzyme to bind tighter with various forms of substrates [38,39]. In this situation, the induced fit model was much more relevant compared to the lock and key model [40]. When the substrate was bound to the enzyme, the active site changed its conformation to align perfectly with the substrate's structure. This optimal orientation strengthens the binding of enzyme and substrate complex for catalytic activity.

Table 2: Amino acid compositions between predictive and template model.

Amino acids	3-Letter	1-Letter	Predictive model		Template model (3SLU)	
			Count	%	Count	%
Alanine	Ala	A	32	10.77	41	11.05
Cysteine	Cys	C	3	1.01	0	0
Aspartic Acid	Asp	D	26	8.75	20	5.39
Glutamic Acid	Glu	E	10	3.37	22	5.93
Phenylalanine	Phe	F	11	3.7	12	3.23
Glycine	Gly	G	31	10.44	50	13.48
Histidine	His	H	6	2.02	11	2.96
Isoleucine	Ile	I	9	3.03	15	4.04
Lysine	Lys	K	7	2.36	14	3.77
Leucine	Leu	L	19	6.4	28	7.55
Methionine	Met	M	3	1.01	4	1.08
Asparagine	Asn	N	9	3.03	8	2.16
Proline	Pro	P	14	4.71	14	3.77
Glutamine	Gln	Q	12	4.04	16	4.31
Arginine	Arg	R	18	6.06	29	7.82
Serine	Ser	S	18	6.06	23	6.2
Threonine	Thr	T	24	8.08	16	4.31
Valine	Val	V	26	8.75	31	8.36
Tryptophan	Trp	W	8	2.69	3	0.81
Tyrosine	Tyr	Y	11	3.7	14	3.77

In overall secondary structures in Fig. 5 (A), there were fewer differences that could be observed between the predicted protease and the template enzyme. The active site residue between the predicted protease and the template enzyme was identical. The catalytic residue for the predicted protease and the template enzyme was His111 and His373, respectively. Protein members in this M23 family were characterized by the conserved histidines that occurred at the HXH motif, as shown in Fig. 5 (a) [41,42]. Based on the structural alignment in Fig. 5 (b), the predicted structure showed 27.07% similarity with the template enzyme. These two structures were aligned according to rigid body superstition. Based on the superposed analysis, the predicted enzyme showed a root mean square deviation (RMSD) value of 0.402 Å based on aligned 58 residues. A lower RMSD value meant higher structure similarity between proteins, and the model was considered very close to the template when the RMSD value was less than 3 Å [43].

https://doi.org/10.31436/iiumej.v25i2.2024

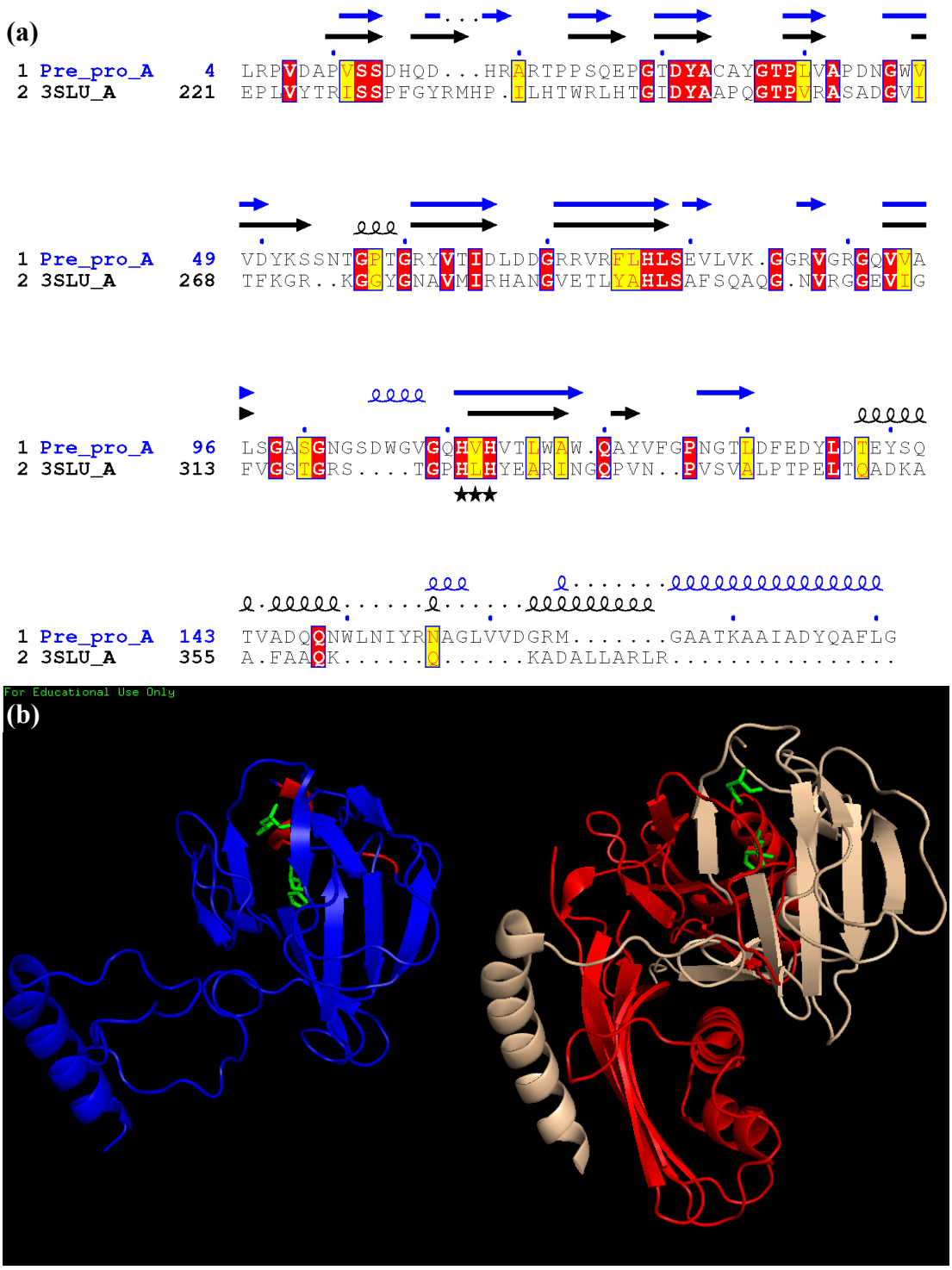


Figure 5. (a) Sequence alignment of predicted protease (Pre_pro) and the template enzyme (3SLU). There was not much difference that can be observed in the overall secondary structure of both enzymes (*** is conserved HXH motif). (b) Predicted protease homology structure (blue) and the template enzyme (3SLU) (brown). Large deletion (red) was observed when both enzymes aligned together.

Based on this computational analysis, the predicted enzyme sequence was shorter and the size was smaller than the template enzyme. The most noticeable variance was the deletion of a large section at the beginning of the predicted enzyme. In the majority, deletion of this section

was assumed to provide the more accessible active site for substrate binding of psychrophilic enzymes [44,45]. Besides, the function of the loop in assisting the enzyme-substrate binding required energy [46,47]. As psychrophiles were surrounded by a lower-temperature environment and exerted lower kinetic energy, the loop formation appeared to be functionless. To compensate for these deleted loops, the psychrophilic enzyme was adapted with a higher flexibility structure, as mentioned earlier. Higher flexibility and more accessible active site of enzyme reduced the substrate specificity and minimized the energy for substrate approachability [48]. Due to the scope of the study at the computational and predictive level, the determination of enzyme size will proceed later through SDS-PAGE.

Moreover, these structural features of the cold-adapted enzymes fitted their thermo-liability characteristic. When the temperature was raised, heat disrupted hydrogen bonds and non-polar hydrophobic interaction formation within the secondary and tertiary structure of the enzyme [49]. As previously discussed, the flexible structure of the cold-active enzyme was increased when these bond formations were decreased by lowering the proportion of arginine and proline residues. Therefore, the fewer bond formations within the enzyme, the easier the enzyme could be denatured by heat. In addition, loop formation also provided extra support for the enzyme to be structurally maintained in high temperatures [50]. When these loops were deleted in a cold-adapted enzyme, the enzyme could not withstand high temperatures, and it unfolded or denatured.

3.4. Enzymatic Assay

For the enzymatic assay, the positive control protease (mesophilic enzyme from bovine pancreases) is used to compare the activity of the predicted protease to determine whether the enzyme is categorized as a cold-active enzyme or not.

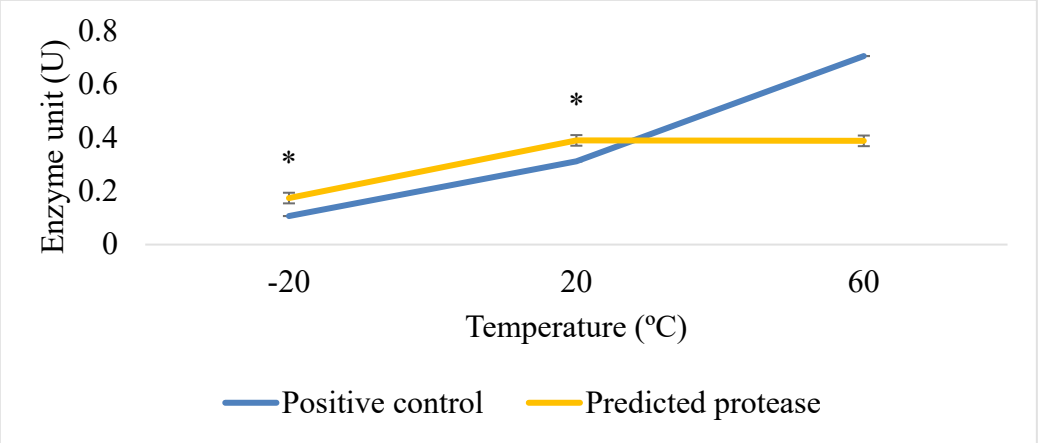


Figure 6. The crude enzyme activity of the predicted protease showed significantly higher at -20 and 20 ºC than the positive control (protease from bovine pancreases). Results were expressed as mean ± SEM. * $p < 0.05$.

Based on the graph in Fig. 6, the predicted protease showed significantly higher proteolytic activity than the positive control at -20 and 20 °C. Regarding cold-active protease, this enzyme has been extracted from psychrophiles inhabiting extreme cold environments, and the majority showed optimal activity at a temperature range between 30 – 40 ºC [51–59]. Researchers' interest in cold-active enzymes has led to the discovery of many novel cold-active proteases and reviewed their bioprospecting potential. Generally, cold-active enzymes are eco-friendly and economical due to their saving energy and cost processing. The use of cold-active protease

in the detergent is beneficial because washing with cold water protects the color and structure of the fabric [60]. In the leather industry, the peeling process with cold-active protease can be performed using tap water only [61]. In the textile industry, cold-active protease could be used in the treatment processes on synthetic fibers that are sensitive to higher temperatures to provide a unique finishing surface [62]. Furthermore, in the food industry, processing at a low temperature preserves the nutritional value and taste in fish and soy sauce production [63]. Additionally, cold-active protease can be used as a meat tenderizer while stored in the freezer [64].

4. CONCLUSION

This preliminary research is vital to unearth all potential enzymes for bioprospecting purposes. The discovery of cold-active proteases in this current research has the potential to be applied in the various industrial field and biotechnological applications. In future, purification of this enzyme is essential to inspect precisely the optimum conditions for enzyme activity.

ACKNOWLEDGEMENT

The research leading to these results received funding from Yayasan Penyelidikan Antartika Sultan Mizan (2015) and International Islamic University Malaysia under grant number RIGS 16-332-0496. The field survey and sampling were made possible with funding to J. Smykla from the Polish National Science Center within grant number NN305376438

REFERENCES

[1] Bálint M, Pfenninger M, Grossart HP, Taberlet P, Vellend M, Leibold MA, et al. (2018) Environmental DNA time series in ecology. Trends Ecol Evol., 33(12):945–57.

[2] Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, et al. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microbiol., 72(5):3291–301.

[3] Pearce DA, Newsham KK, Thorne MAS, Calvo-Bado L, Krsek M, Laskaris P, et al. (2012) Metagenomic analysis of a southern maritime Antarctic soil. Front Microbiol., 3(DEC):1–13.

[4] Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, et al. (2020) Challenges of unculturable bacteria: environmental perspectives. Reviews in Environmental Science and Biotechnology.

[5] Marathe NP, Janzon A, Kotsakis SD, Flach CF, Razavi M, Berglund F, et al. (2018) Functional metagenomics reveals a novel carbapenem-hydrolyzing mobile beta-lactamase from Indian river sediments contaminated with antibiotic production waste. Environ Int.

[6] Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD, et al. (2017) Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. PLoS One.

[7] Biver S, Portetelle D, Vandenbol M. (2013) Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. Springerplus, 2(1):1–10.

[8] Maruthamuthu M, Jiménez DJ, Stevens P, van Elsas JD. (2016) A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveilsnovel thermoalkaliphilic enzymes. BMC Genomics.

[9] Pabbathi NPP, Velidandi A, Tavarna T, Gupta S, Raj RS, Gandam PK, et al. (2021) Role of metagenomics in prospecting novel endoglucanases, accentuating functional metagenomics approach in second-generation biofuel production: a review. Biomass Conversion and Biorefinery.

[10] Castillo Villamizar GA, Nacke H, Boehning M, Herz K, Daniel R. (2019) Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. MBio.

[11] Ngara TR, Zhang H. (2018) Recent advances in function-based metagenomic screening. Genomics, Proteomics Bioinforma., 16(6):405–15.

[12] Berlemont R, Pipers D, Delsaute M, Angiono F, Feller G, Galleni M, et al. (2011) Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements. Rev Argent Microbiol., 43(2):94–103.

[13] Heath C, Xiao PH, Cary SC, Cowan D. (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. Appl Environ Microbiol., 75(13):4657–9.

[14] Cieśliński H, Białkowska A, Tkaczuk K, Długołęcka A, Kur J, Turkiewicz M. (2009) Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. Polish J Microbiol., 58(3):199–204.

[15] Xiao X, Yin X, Lin J, Sun L, You Z, Wang P, et al. (2005) Chitinase genes in lake sediments of Ardley Island, Antarctica. Appl Environ Microbiol., 71(12):7904–9.

[16] Kuhn E, Bellicanta GS, Pellizari VH. (2009) New alk genes detected in Antarctic marine sediments. Environ Microbiol., 11(3):669–73.

[17] Latip MAA, Alias SA, Smykla J, Yusof F, Mohamad MAN, Nordin NFH. (2020) Discovery of cold-active protease from psychrophilic bacteria isolated from Antarctic region for bio-prospecting. Malays Appl Biol., 49:55–60.

[18] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol., 19(5):455–77.

[19] Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, et al. (2009) Community-wide analysis of microbial genome sequence signatures. Genome Biol., 10(8).

[20] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res., 42(D1).

[21] Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. (2017) CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res., 45(D1):D200–3.

[22] Robert X, Gouet P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res., 42(W1).

[23] De Lano WL. (2020) The PyMOL molecular graphics system, version 2.3. Schrödinger LLC.

[24] Ekkers DM, Cretoiu MS, Kielak AM, Van Elsas JD. (2012) The great screen anomaly-a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol., 93(3):1005–20.

[25] Thies S, Rausch SC, Kovacic F, Schmidt-Thaler A, Wilhelm S, Rosenau F, et al. (2016) Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community. Sci Rep., 6.

[26] Devi SG, Fathima AA, Sanitha M, Iyappan S, Curtis WR, Ramya M. (2016) Expression and characterization of alkaline protease from the metagenomic library of tannery activated sludge. J Biosci Bioeng., 122(6):694–700.

[27] Pessoa TBA, Rezende RP, Marques E de LS, Pirovani CP, dos Santos TF, dos Santos Gonçalves AC, et al. (2017) Metagenomic alkaline protease from mangrove sediment. J Basic Microbiol., 57(11):962–73.

[28] Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, et al. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol., 127(4):575–92.

[29] Vester JK, Glaring MA, Stougaard P. (2014) Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb Cell Fact., 13(1):72.

[30] Douglass A, O'Brien C, Offei B, Coughlan A, Ortiz-Merino R, Butler G, et al. (2018) Coverage-versus-Length plots, a simple quality control step for de novo yeast genome sequence assemblies. bioRxiv.

[31] Jarocki VM, Reid CJ, Chapman TA, Djordjevic SP. (2020) Escherichia coli ST302: Genomic analysis of virulence potential and antimicrobial resistance mediated by mobile genetic elements. Front Microbiol.

[32] Rawlings ND, Salvesen G. (2013) Handbook of proteolytic enzymes. In: Handbook of Proteolytic Enzymes. Academic Press.

[33] Wang X, Yang X, Yang C, Wu Z, Xu H, Shen Y. (2011) Crystal structure of outer membrane protein NMB0315 from Neisseria meningitidis. PLoS One. 6(10).

[34] Kumar V, Sharma N, Bhalla TC. (2014) In silico analysis of β -galactosidases primary and secondary structure in relation to temperature adaptation. J Amino Acids.

[35] Goldstein RA. (2007) Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: Insights from the quasi-chemical approximation. Protein Sci., 16(9):1887–95.

[36] Metpally RPR, Reddy BVB. (2009) Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. BMC Genomics. 10.

[37] Óskarsson KR, Sævarsson AF, Kristjánsson MM. (2020) Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops. Sci Rep., 10(1).

[38] Huber R, Bennett WS. (2013) Functional significance of flexibility in proteins. Pure Appl Chem., 54(12):2489–500.

[39] Janin J, Sternberg MJE. (2013) Protein flexibility, not disorder, is intrinsic to molecular recognition. F1000 Biol Rep., 5(1).

[40] Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, et al. (2014) Induced-fit mechanism in class I terpene cyclases. Angew Chemie - Int Ed., 53(29):7652–6.

[41] Grabowska M, Jagielska E, Czapinska H, Bochtler M, Sabala I. (2015) High resolution structure of an M23 peptidase with a substrate analogue. Sci Rep., 5.

[42] Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M. (2004) Latent LytM at 1.3 Å resolution. J Mol Biol., 335(3):775–85.

[43] Reva BA, Finkelstein A V., Skolnick J. (1998) What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? Fold Des., 3(2):141–7.

[44] Russell RJM, Gerike U, Danson MJ, Hough DW, Taylor GL. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure. 6(3):351–61.

[45] Aghajari N, Van Petegem F, Villeret V, Chessa JP, Gerday C, Haser R, et al. (2003) Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. Proteins Struct Funct Genet., 50(4):636–47.

[46] Yüksel KÜ, Sun AQ, Gracy RW, Schnackerz KD. (1994) The hinged lid of yeast triose-phosphate isomerase: Determination of the energy barrier between the two conformations. J Biol Chem., 269(7):5005–8.

[47] Chesters C, Wilding M, Goodall M, Micklefield J. (2012) Thermal bifunctionality of bacterial phenylalanine aminomutase and ammonia lyase enzymes. Angew Chemie - Int Ed., 51(18):4344–8.

[48] Siddiqui KS, Poljak A, Guilhaus M, De Francisci D, Curmi PMG, Feller G, et al. (2006) Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted α-amylase from Pseudoalteramonas haloplanktis. Proteins Struct Funct Genet., 64(2):486–501.

[49] Povarova OI, Kuznetsova IM, Turoverov KK. (2007) Different disturbances - one pathway of protein unfolding. Actin folding-unfolding and misfolding. Cell Biol Int., 31(4 SPEC. ISS.):405–12.

[50] Vemparala S, Mehrotra S, Balaram H. (2011) Role of loop dynamics in thermal stability of mesophilic and thermophilic adenylosuccinate synthetase: A molecular dynamics and normal mode analysis study. Biochim Biophys Acta - Proteins Proteomics. 1814(5):630–7.

[51] Furhan J, Salaria N, Jabeen M, Qadri J. (2019) Partial purification and characterisation of cold-active metalloprotease by Bacillus sp. Ap1 from Apharwat Peak, Kashmir. Pakistan J Biotechnol., 16(1):47–54.

[52] Leng W, Gao R, Wu X, Zhou J, Sun Q, Yuan L. (2021) Genome sequencing of cold-adapted Planococcus bacterium isolated from traditional shrimp paste and protease identification. J Sci Food Agric., 101(8):3225–36.

[53] Park HJ, Han SJ, Yim JH, Kim D. (2018) Characterization of an Antarctic alkaline protease, a cold-active enzyme for laundry detergents. Korean J Microbiol., 54(1):60–8.

[54] Yang XS, Chen XL, Xu XZ, Zeng RY. (2011) Cold-adaptive alkaline protease from the psychrophilic Planomicrobium sp . 547: enzyme characterization and gene cloning. Adv Polar Sci., 22(1):49–54.

[55] Perfumo A, Freiherr von Sass GJ, Nordmann EL, Budisa N, Wagner D. (2020) Discovery and characterization of a new cold-active protease from an extremophilic bacterium via comparative genome analysis and in vitro expression. Front Microbiol., 11.

[56] Ji XL, Taj MK, Lu XB, Lin LB, Zhang Q, Wei YL. (2014) Purification and characterization of an extracellular cold-active protease produced by the psychrotrophic bacterium Serratia Sp. WJ39. In: Applied Mechanics and Materials. 330–4.

[57] Tariq AL. (2011) Purification and characterization of 56 kDa cold active protease from Serratia marcescens. African J Microbiol Res., 5(32).

[58] Nor NM, Mohamed MS, Loh TC, Foo HL, Rahim RA, Tan JS, et al. (2017) Comparative analyses on medium optimization using one-factor-at-a-time, response surface methodology, and artificial neural network for lysine–methionine biosynthesis by Pediococcus pentosaceus RF-1. Biotechnol Biotechnol Equip., 31(5):935–47.

[59] Mohammed K, Pramod WR. (2011) Production optimization of an extracellular cold-active alkaline protease from Stenotrophomonas maltophilia MTCC 7528 and its application in detergent industry. African J Microbiol Res., 5(7):809–16.

[60] Salwan R, Sharma V, Kasana RC, Gulati A. (2020) Bioprospecting psychrotrophic bacteria for serine-type proteases from the cold areas of western Himalayas. Curr Microbiol., 77(5):795–806.

[61] Paul T, Jana A, Mandal AK, Mandal A, Das Mohpatra PK, Mondal KC. (2016) Bacterial keratinolytic protease, imminent starter for NextGen leather and detergent industries. Sustain Chem Pharm., 3:8–22.

[62] Kanelli M, Vasilakos S, Ladas S, Symianakis E, Christakopoulos P, Topakas E. (2017) Surface modification of polyamide 6.6 fibers by enzymatic hydrolysis. Process Biochem., 59:97–103.

[63] Kaewkrod A, Niamsiri N, Likitwattanasade T, Lertsiri S. (2018) Activities of macerating enzymes are useful for selection of soy sauce koji. LWT - Food Sci Technol., 89:735–9.

[64] Zhao GY, Zhou MY, Zhao HL, Chen XL, Xie B Bin, Zhang XY, et al. (2012) Tenderization effect of cold-adapted collagenolytic protease MCP-01 on beef meat at low temperature and its mechanism. Food Chem., 134(4):1738–44.

https://doi.org/10.31436/iiumej.v25i2.3098

CONTINUOUS FLOW BEAD-MILLING IMPACT ON SULFUR CURING FOR ADVANCED ELASTOMERIC RUBBER COMPOSITES

MOHAMAD FIRDAUS OMAR[1,2], FATHILAH ALI[1*], MOHAMMED SAEDI JAMI[1], AZLIN SUHAIDA AZMI[1], FARAH AHMAD[1], MOHD ZAHID MARZUKI[3], AND JAMAROSLIZA JAMALUDDIN[4]

[1]*Department of Chemical Engineering and Sustainability, Kulliyyah of Engineering, International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur, Malaysia.*

[2]*SC Johnson Asia Pacific Sdn. Bhd., UOA Corporate Tower, The Vertical, 8, Jalan Kerinchi, Bangsar South, 59200 Kuala Lumpur, Malaysia.*

[3]*Faculty of Chemical Engineering, Universiti Teknologi MARA (UiTM) 40450 Shah Alam, Selangor, Malaysia.*

[4]*Department of Chemical Engineering, School of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.*

**Corresponding author: fathilah@iium.edu.my*

(Received: 7 November 2023; Accepted:6 October 2024; Published on-line: 15 July 2024)

ABSTRACT: As an eco-friendly and technically feasible method for physical modification of materials, bead-milling has been extensively used in many industrial applications ranging from chemicals, nanomaterials, foods, and pharmaceuticals with impacts on particle size, surface morphology, stability, and overall products' performance. Apparently, there have been limited studies conducted on sulfur curative dispersion using this technology, necessitating a thorough investigation of its performance. The objectives of the present study were to explore the influence of the bead-milling process parameters, particularly rotational speed and flow rate, on the sulfur curative dispersion characteristics and to analyze its behavior within the rubber elastomer matrix. Taguchi's L9 orthogonal array experimental design was employed to identify the optimal rotational speed and flow rate of a 60-L bead-milling machine on the sulfur curative dispersion. The stability and morphology of the resulting sulfur curative dispersion were characterized, along with its mechanical properties in rubber elastomers. It was found that higher rotational speed (800 rpm) and lower flow rate (350 L/h) of the bead-milling process resulted in smaller sulfur particle sizes, leading to improved tensile strength of the rubber elastomer. This research may provide valuable insights to determine the ideal bead-milling process for sulfur curative, enhancing the mechanical properties and overall performance of elastomeric rubber composites as well as across various fields.

ABSTRAK: Sebagai sebahagian kaedah mesra alam dan secara teknikal dapat mengubah suai bentuk fizikal bahan, mesin pengisaran manik telah digunakan secara meluas dalam kebanyakan aplikasi industri termasuk kimia, nanomaterial, makanan, dan farmaseutikal dengan kesan pada saiz zarah, morfologi permukaan, kestabilan, dan prestasi keseluruhan produk. Namun, terdapat kurang kajian terhadap taburan kuratif sulfur menggunakan teknologi ini, menyebabkan penyelidikan menyeluruh diperlukan bagi menilai prestasinya. Objektif kajian ini adalah bagi mengkaji pengaruh parameter proses pengisaran manik, terutamanya pada kelajuan putaran dan kadar aliran, iaitu terhadap ciri-ciri taburan kuratif sulfur dan mengkaji tindak balasnya pada matriks elastomer getah. Reka bentuk eksperimen Taguchi L9 bersusunan ortogonal telah digunakan bagi mengenal pasti kelajuan putaran dan

kadar aliran optimal mesin pengisaran manik 60 liter terhadap taburan kuratif sulfur. Kestabilan dan morfologi taburan kuratif sulfur yang terhasil telah dicirikan, bersama sifat mekanikal dalam elastomer getah. Dapatan mendapati bahawa proses pengisaran manik pada kelajuan putaran lebih tinggi (800 rpm) dan kadar aliran lebih rendah (350 L/j) menyebabkan saiz zarah sulfur lebih kecil, membawa kepada peningkatan kekuatan regangan elastomer getah. Kajian ini mungkin memberikan pengetahuan penting bagi menentukan proses pengisaran manik yang ideal untuk kuratif sulfur, meningkatkan sifat mekanikal dan prestasi keseluruhan komposit getah elastomerik serta pelbagai bidang lain.

KEYWORDS: *bead-milling, sulfur, rubber.*

1. INTRODUCTION

The bead-milling process has been recognized as an eco-friendly and technically feasible method for the physical modification of materials that can give a mechanochemical effect to the characteristics of a material through friction, collision, impact, and shear from the milling beads and the mill's chamber or container wall. This physical modification is not only able to produce variations of outcome in terms of the particle size, surface properties, and morphology – where the chemical reactions and alteration may impact the structural, physical, and chemical properties – but also in terms of solubility index as well as functional properties such as water absorption, swelling capacity, pasting, and gelation of the milled materials [1–3].

Many researchers reported the use of bead-milling processes as one of the approaches to physically modify their products or materials to achieve the desired characteristics. A study on the effects of multi-wall carbon nanotubes (MWCNTs) and the bead-milling process parameters on aluminum-magnesium (Al-Mg) alloy powder structure proved that increasing time and speed of the bead-milling process leads to the formation of finer particles, which consequently intensifies the plastic deformation and results in a smaller crystallite size. The results indicated that the crystal size was reduced by almost 83%, which consequently improved the distribution of the MWCNTs as well as enhanced their diffusion into the Al-Mg matrix [4]. Other than that, a study on wheat gluten protein's structural and functional properties using bead-mill processing also proved significant outcomes. Those outcomes include breaking of disulfide bonds in gluten, increasing the free sulfhydryl groups, improving the surface hydrophobicity from 940.97 to 1197.50, enhancing the foaming capacity from 8.7 to 31 cm^3, reducing the particle size from 85.9 to 32.3 μm; and increasing in whiteness from 49.51 to 65.59, indicating the advantage of bead-milling technology in improving the physicochemical properties of wheat gluten that diversifies its application potential [5]. Besides that, another study conducted to produce an ultrafine drug nanocrystal suspension that can pass through 0.22 μm sterilization filters found that finer milling beads, as in 0.1 mm yttrium-stabilized zirconia beads, were the most optimal beads size to mill the AZD4854 drug compound finely at the milling conditions of 2 hours milling time and 700 rpm rotation speed. Plus, the ultrafine drug suspensions were stable, and they did not sediment after several months of refrigeration – making them advantageous in ensuring a homogenous dose in clinical settings without the need for sonication or stirring [6]. Furthermore, the bead-milling method enhances the stability and performance of cesium lead bromide perovskite quantum dots/polymethyl methacrylate ($CsPbBr_3$ PQDs/PMMA) composites, where the use of the 1.5% $CsPbBr_3$ PQDs/PMMA composite prepared by the said method not only demonstrated excellent performance with a photoluminescence quantum yield (PLQY) of approximately 78% and high stability, but also exhibited high luminous efficiency in a white light-emitting diode, and its luminescence played a significant role in Rhodamine B (RhB) degradation under visible light [7]. It was highlighted in the paper that the benefits of the bead-milling method include its straightforwardness, being

free from solvent and ligand, being able to operate at low synthesis temperatures, and being easily scalable. These examples of studies show that bead-milling has been used extensively in studies for many potential industrial applications ranging from chemicals, nanomaterials, food processing as well as pharmaceutical drug manufacturing with impacts on particle size, surface morphology, stability, as well as overall performance of the products studied.

Separately, a curing agent, also known as a curative, is a chemical agent used to initiate a chemical reaction, particularly in cross-linking elastomer molecules. When used alongside an accelerator(s) and activator(s) at a significant temperature, the curative forms stable covalent carbon-carbon double bonding. Sulfur is a commonly used curative for rubber elastomers, and it should have a purity of at least 99.5% with very minimal ash content to achieve a maximized curing performance. It can be prepared in a colloid or dispersion state through mixing and/or milling processes, which in turn produces fine sulfur particles that remain suspended (stable) and are suitable for latex compounding in rubber elastomer manufacturing. Remarkably, the solubility of sulfur inside the rubber elastomer system is another critical aspect to look at, where typically, the use of insoluble sulfur is preferred over soluble sulfur to address potential "sulfur blooms" issues in uncured rubber. These blooms can hinder the adhesion of rubber to adjacent layers or substrates, affecting the performance of molded rubber elastomer products. While complete dissolution is usually achieved upon mixing the compound ingredients, unexpected blooms can still occur during actual industrial processing due to the reversion of insoluble sulfur to orthorhombic soluble sulfur before the rubber is cured. Consequently, the sulfur supersaturates the rubber over time, leading to the formation of crystals within the bulk rubber and initiating growth toward the surface [8, 9]. As a result, extensive research has explored methods to enhance the characteristics of sulfur curatives while minimizing blooming events in rubber compounds, which encompass optimizing process parameters such as the use of ball-milling to improve sulfur curative dispersion [10], chemical modification of sulfur for better compatibility with rubber elastomers [11], as well as incorporation of nanoscale additives to enhance mechanical properties [12]. These investigations aim to enhance sulfur's properties and dispersion within rubber, ultimately leading to the development of high-performance elastomeric materials.

Although a lot of studies have been done on different milling parameters over the products' behavior, fewer studies have been done, particularly on sulfur dispersion, and a comprehensive investigation is needed to assess any changes in sulfur properties after bead-milling. The primary objectives of this research are to investigate the influence of continuous flow bead-milling process parameters on the behavior of sulfur curative and its subsequent impact on the performance of elastomeric rubber composites. Specifically, the study aims to explore how varying milling parameters, such as rotational speed and flow rate, affect the dispersion and stability of sulfur curative within the rubber matrix. By understanding these interactions, the research aims to optimize the bead-milling process for sulfur curative, leading to the production of elastomeric articles with enhanced overall performance via characterizations on stability, visual, morphology, and tensile properties by using a Stability Analyzer, a Light Emitting Diode (LED) magnifier lens, Scanning Electron Microscopy / Energy Dispersive X-Ray Analyzer (SEM/EDX), and a Universal Testing Machine (UTM) respectively. Furthermore, the article seeks to contribute valuable insights that can be applied to other industrial applications, promoting the use of bead-milling as an eco-friendly and technically feasible method for enhancing material characteristics in various fields.

https://doi.org/10.31436/iiumej.v25i2.3098

2. MATERIALS AND METHODS

2.1. Chemicals

The main materials used in this project include Hi-Pure Double-Refined sulfur powder supplied by Taiko Group, a surfactant of sodium salt of naphthalene-sulfonic acid from Behn Meyer Group, and deionized water. For the preparation of the rubber elastomer, the high-ammonia natural rubber latex used was obtained from Getahindus Sdn. Bhd. and the compounding ingredients were all industrial grades from Behn Meyer Group, including the high-ammonia natural rubber latex, calcium carbonate, potassium hydroxide, zinc dibutyl-dithiocarbamate, zinc oxide, antioxidant, and coagulant premix with minimum 94% purity each.

2.2. Characterizations

To determine the particle size in terms of D_{90} value of the sulfur curative bead-milled dispersion, a HORIBA Particle Size Analyzer of LA-960V2 model was utilized. Other than that, Total Solids Content (TSC) was calculated to ensure proper control of water loss or excess input. TSC, measured in percentage, was measured as the remaining weight of sample after drying with the aid of an oven at 100 ºC, expressed in terms of the original weight of the wet sample as in Eq. 1.

$$TSC\ \% = \left[\frac{(\text{weight of dry sample+container})-(\text{weight of empty container})}{(\text{weight wet sample+container})-(\text{weight empty container})}\right] \times 100\% \quad (1)$$

The pH value of the sulfur curative dispersion was crucial for its compatibility with rubber as a curing agent, and thus, it should be carefully maintained within the alkaline range. A Brookfield viscometer was utilized together with its RV04 spindle type to assess the dispersion's viscosity.

Furthermore, the stability of selected sulfur dispersion samples after the bead-milling process was analyzed using a multi-sample photo centrifuge – LUMiSizer LS650, providing insights into the dispersion's stability. The properties of the elastomeric rubber samples produced were analyzed in terms of mechanical properties and morphological behavior. A universal testing machine supplied by GT Instruments Sdn. Bhd. (Model: Al-3000) was used to test the tensile strength of the samples, and the samples' thicknesses were measured using a thickness gauge to ensure they were within 0.06 ± 0.005 mm. The morphological behavior of the rubber samples was observed using an LED magnifier lens and SEM/EDX machine, enabling detailed microphotographs to support the understanding of the sulfur curative's behavior within the rubber elastomer matrix.

2.3. Process Equipment

The general process started with transporting the raw materials for sulfur dispersion into a 4 MT mixing tank. The tank's mixing time and speed were optimized before the sulfur was continuously pumped into a CDS-60 60-litre continuous large-flow bead-milling machine. The bead-milling machine was equipped with approximately 140 kg and 0.9 mm-sized Yttrium-stabilized Zirconia beads. The parameters studied for this process were rotational speed (motor speed) and flow rate (pump rate).

2.4. Preparation of Sulfur Curative

The dispersion of sulfur curative was done by mixing the raw materials for sulfur curative dispersion using the formula provided in Table 1. The mixing speed and time were set at 800

rpm and 30 minutes, respectively. Subsequently, the sulfur curative dispersion was continuously pumped to flow into a CDS-60 60-litre continuous flow bead-milling machine for further processing. The parameters under study for this process were rotational speed and flow rate with respective values and units (Table 2), and the runs were selected using Taguchi's L9 orthogonal array in Table 3 to analyze the different process parameters' outcomes.

2.5. Rubber Elastomer Casting

For latex compounding, 4 significant samples of sulfur from the bead-milling process were chosen primarily based on their particle size, D_{90}. The chemical formulation for compounding was prepared in Table 4, with the dosage calculated in per hundred rubber (phr). The latex compounding process involved mixing the chemicals in a beaker before incorporating them into the raw latex. The compounded latex was then stirred and matured for 36 to 48 hours.

Table 1. Ingredients in preparing sulfur curative dispersion

Ingredients	Weight Percentage (%)
Deionized water	34.0
Surfactant sodium salt of naphthalene-sulfonic acid	6.0
Sulfur curative solids	60.0
TOTAL	100.00

Table 2. Control factors with their respective levels for sulfur curative bead-milling

Factors	Parameter (unit)	Levels		
		1	2	3
A	Rotational Speed (rpm)	800	700	600
B	Flow Rate (L/h)	750	550	350

Table 3. Taguchi's L9 orthogonal array experimental design on sulfur curative bead-milling

Run No.	A Rotational Speed	B Flow Rate
1	Level 1	Level 1
2	Level 1	Level 2
3	Level 1	Level 3
4	Level 1	Level 1
5	Level 2	Level 1
6	Level 2	Level 2
7	Level 3	Level 1
8	Level 3	Level 2
9	Level 3	Level 3

The rubber films were then cast using glaze-coated ceramic plates. The process was initiated by heating the ceramic plates in an oven at 70 °C. The formers were then dipped into a coagulant premix and left to dry in a 60 °C oven. After dipping the formers into the latex compound for 5 to 7 seconds, they were cured in oven at 120 °C for 20 minutes before they were left to cool prior stripping the cast rubber elastomers from the ceramic plates.

https://doi.org/10.31436/iiumej.v25i2.3098

Table 4. Formulation of natural rubber latex compound for rubber elastomer making

Compounding Ingredient	Dosage (phr)
Natural rubber latex	100
Calcium carbonate filler	30
Potassium hydroxide	1.4
Zinc dibutyl-dithiocarbamate	0.2
Sulfur curative	1.2
Zinc oxide	1.0
Antioxidant	0.5

3. RESULTS AND DISCUSSION

3.1. Curative Properties and Stability Analysis

Table 5 shows the dispersion characteristics of sulfur curatives achieved through varying bead-milling process parameters. The dispersion properties include pH, TSC, viscosity, and particle size. Triplicate samples of sulfur curatives were taken, and their properties recorded for each run and averaged out. The particle size values were assessed further to understand how this property is impacted by the changes in the bead-milling process parameters.

Table 5. Sulfur curative specifications results

Run No.	Rotational Speed (rpm)	Flow Rate (L/h)	pH	TSC (%)	Viscosity (cP)	Particle Size, D_{90} (µm)
1	800	750	9.5	61.31	472	7.750
2	800	550	9.5	62.99	610	6.718
3	800	350	9.6	61.63	598	6.154
4	700	750	9.7	62.05	470	8.650
5	700	550	9.9	61.84	646	7.760
6	700	350	9.5	62.31	616	6.820
7	600	750	9.7	62.43	370	9.998
8	600	550	9.7	62.69	432	8.444
9	600	350	9.5	61.89	518	7.214

From Table 5, it can be roughly observed that particle size and viscosity demonstrated variations across the different runs, while pH and TSC were observed to be constant with fewer variations. This variation in particle sizes may be able to influence the resulting dispersion's stability and performance within the rubber matrix. Thus, Fig. 1 was plotted, representing the graph of the effect of particle size on means according to the responses of Taguchi's L9 orthogonal array in Table 6. The calculation and response analysis for the means plot and Signal-to-Noise (S/N) ratio for the particle size in D90 was based on a smaller-is-better analysis for the effects' investigation, which means that the lower or smaller the particle size value, the better the response. Thus, the "mean of means" was plotted in Fig. 1 to signify the trend of the particle size based on the average particle sizes from Table 6.

It can be observed from both the table and figure that the particle size is inversely proportional to the bead mill's rotational speed and directly proportional to the flow rate. Particle size reduction is greatest at higher collision rate of rolling or rotating beads inside the milling chamber, where it is achievable at higher mixing speed and more frequent milling time [13]. Fig. 2, in turn, illustrates particles that are dispersed and then crushed between rolling beads [14].

https://doi.org/10.31436/iiumej.v25i2.3098

Intuitively, the bead mill's rotational speed can affect the rate of energy input into the system, having concomitant effects on the process efficiency – where at very low rotational speed, the system would be likely to have very poor mixing in the system, leading to sedimentation of the various materials in the milling chamber that limits total particles milling [15]. The behavior at very high speed (for example, more than 800 rpm in this study) may not be so clear, although it might be expected that the efficiency continues to decrease until there are practical limitations to the increases (such as the material leaving the bead-mill's chamber). Additionally, higher speeds are likely to cause a move away from a stable flow pattern. However, a lower speed is more efficient when considering milling energy. If the control was the total volume of material that achieves the required size reduction in a given period of time, then the effect result may favor higher rotational speeds. For this reason, full-scale plants from this chemical plant use higher rotational speed to enable greater plant throughput, which in turn can increase the overall plant efficiency [15].

In conjunction with this, upon being adsorbed on material surfaces, surfactant molecules provide electrostatic and steric forces to stabilize the particles in the milling media, minimizing the uncontrolled breakdown of particles. Relatively, the sulphonate surfactant used in this study can also aid in lowering the surface energy of the freshly formed fine particles of sulfur dispersion by forming a thin organic layer and introducing long-range capillary forces that lower the energy for crack propagation. The smaller particle size formed at a higher milling rotational speed and slower flow rate is due to the inhibition of the particles from agglomeration the agglomeration will substantially increase particle sizes in the high-energy milling operation employed [16].

Table 6. The particle size of sulfur curative dispersion after bead-milling

Run No.	Rotational Speed (rpm)	Flow Rate (L/h)	Mean D_{90} (µm)	Standard Deviation	S/N Ratio
1	800	750	7.750	0.2143	-17.7890
2	800	550	6.718	0.1057	-16.5455
3	800	350	6.154	0.0338	-15.7837
4	700	750	8.650	0.1363	-18.7414
5	700	550	7.760	0.1322	-17.7981
6	700	350	6.820	0.0740	-16.6769
7	600	750	9.998	0.1250	-19.9990
8	600	550	8.444	0.0473	-18.5311
9	600	350	7.214	0.0836	-17.1647

https://doi.org/10.31436/iiumej.v25i2.3098

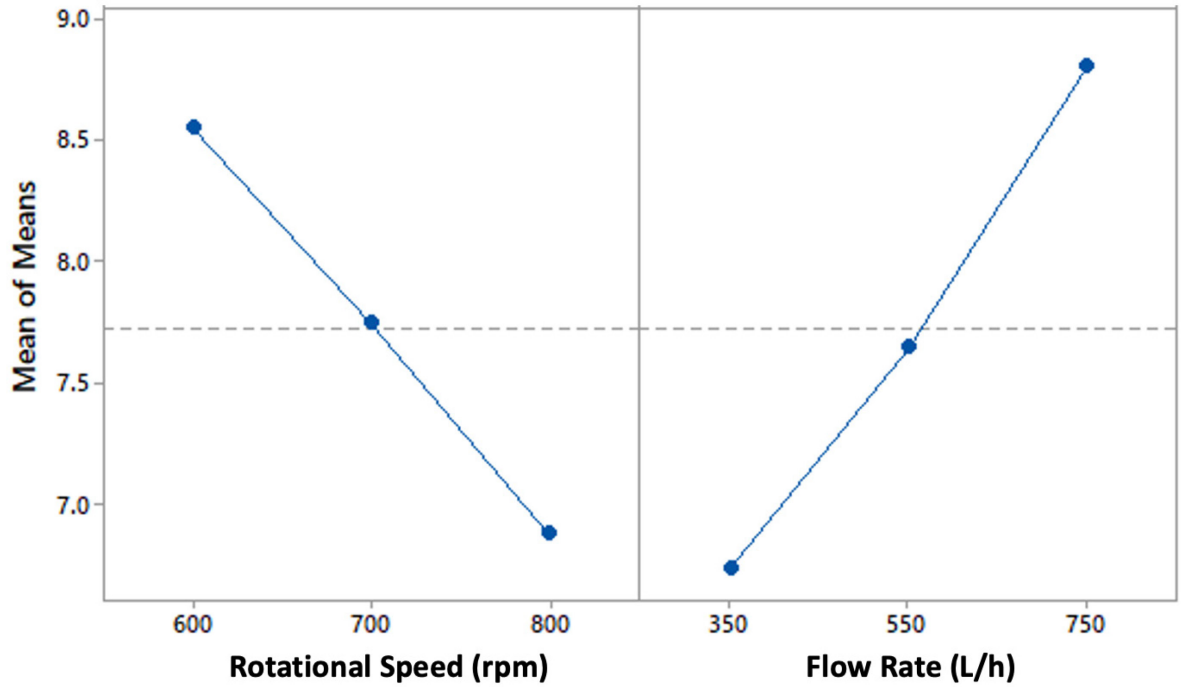


Figure 1. Effect of bead-milling parameters on particles size, D_{90} mean of means

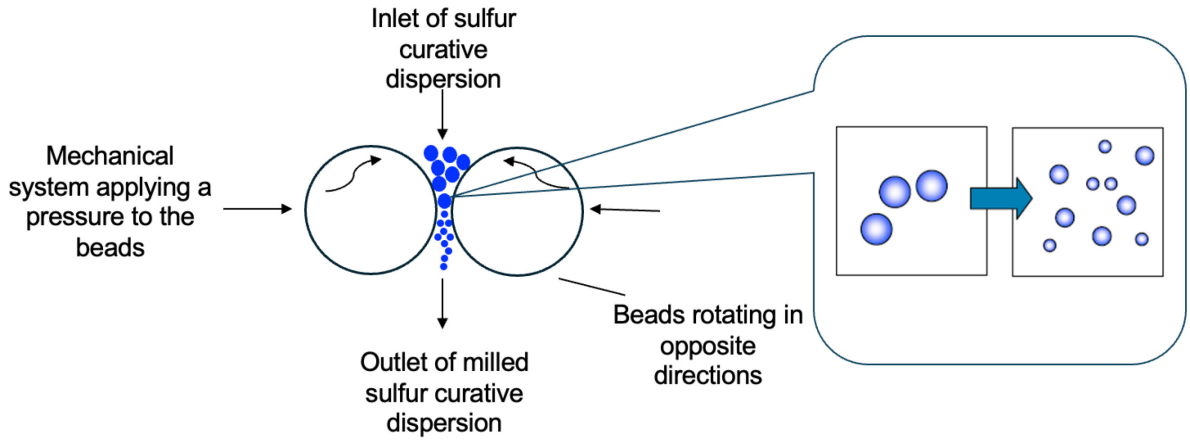


Figure 2. Illustration of particles in between rolling beads in solid particle dispersion by bead-milling

Table 7. ANOVA on sulfur curative dispersion bead-milling particle size

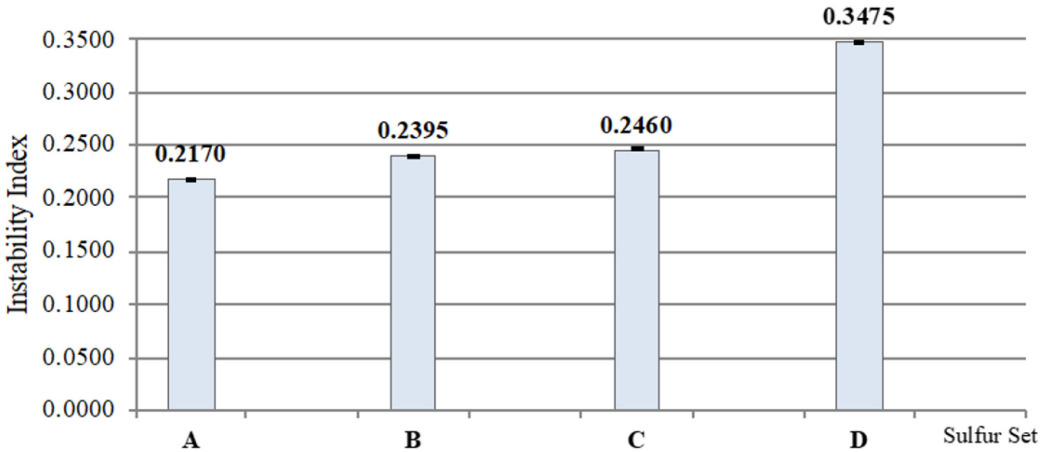
Source	DF	Adj SS	Adj MS	*F*-Value	*p*-Value	Contribution (%)	Remark
Rotational Speed	2	5.2043	2.60216	41.56	0.002	38.81	Significant
Flow Rate	2	7.9535	3.97675	63.52	0.001	59.32	Significant
Error	4	0.2504	0.06261			1.87	
Total	8	13.4083				R^2 = 98.13%	

On the other hand, Analysis of Variance (ANOVA) was performed to determine the percent contribution of the operational variables to the response, as in Table 7. The table shows that

https://doi.org/10.31436/iiumej.v25i2.3098

the results had been influenced considerably, with the flow rate employed at 59.32 % and rotational speed at 38.81 % influence, respectively. Besides that, the p-values are both less than 0.05 (0.002 and 0.001 for rotational speed and flow rate, respectively), which is close to the significance value (95 % confidence level). At a 95% confidence level, factors having a p-value less than 0.05 are considered significant [17,18], proving these parameters are significant to impact the sulfur's particle size.

The stability analysis for the selected four sulfur curative dispersion samples was conducted using centrifugal separation analysis. Fig. 3 shows the average instability indices of the four different sets of sulfur curative dispersion. Samples A, B, C, and D indicate sulfur curative dispersion after bead-milling, each with different set parameters, which were 800 rpm rotational speed with 350 L/h flow rate, 700 rpm rotational speed with 350 L/h flow rate; 700 rpm rotational speed with 550 L/h flow rate; and 600 rpm rotational speed with 750 L/h flow rate, respectively. The selection of the sulfur was based on the smallest particle size (optimized) and compared with the other three subsequent sulfur sets of various medium to largest size of sulfur. One of them is combined with the lowest viscosity obtained after the bead-milling process.



Rotational speed (rpm)	800		700		700		600	
Flow Rate (L/h)	350		350		550		750	

Figure 3. Instability index of different sulfur dispersion bead-milling parameters

It can be observed that the sulfur curative dispersion is most stable for sample A, followed by sample B, C, and D. Briefly, the instability index quantifies the clarification in transmission based on particle size and separation at a given time in the presence of accelerated gravitational force, divided by the maximum clarification possible. The clarification is due to the increase in transmission or decrease in droplet concentration due to the movement of nanodroplets towards the cream layer. It is a dimensionless number between 0 and 1, where 0 indicates no separation or no change in sulfur dispersion's transmission and hence highest stability, and 1 represents complete segregation of phases and hence lowest stability under the centrifugal field. Comparing instability indices of sulfur dispersion under an accelerated gravitational field ultimately aids in a quick comparison of their shelf-life rather than waiting for the separation to occur under normal gravity conditions [19]. Intuitively, more stable sulfur dispersion can be produced at a higher rotational speed that causes more frequent bead collisions in combination with the lower flow rate for sulfur introduction; the sulfur has more time exposed to more bead

https://doi.org/10.31436/iiumej.v25i2.3098

collisions, thus altering the sulfur dispersion's chemical properties. Unlike unstable suspension or dispersion, a stable suspension or dispersion will not form aggregates over time. In unstable suspension, particles tend to create particle-particle interactions back, forming aggregates at early stages. Upon late stages, networks of aggregates may form agglomerates that increase particle size and reduce surface area. Networks of agglomerates, in turn, form solid-like gels, which is unfavorable for most applications that use suspensions or dispersions.

3.2. Influence on the Resulting Rubber Elastomer

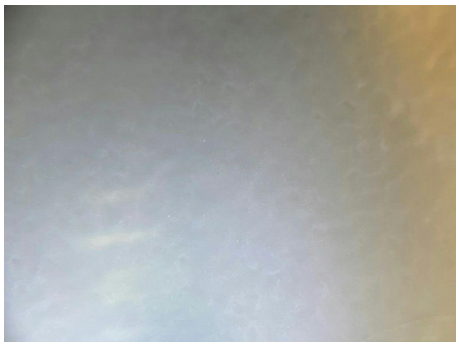
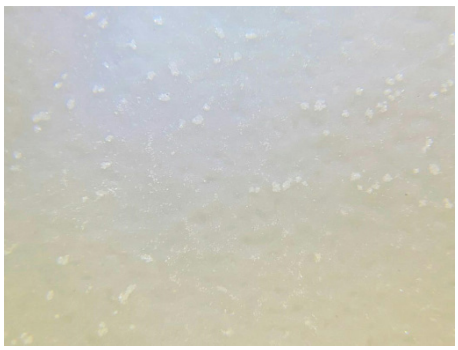
Table 8 shows the rubber elastomer film observation results as seen from the LED magnifier lens while Table 9 shows the microphotographs of the rubber elastomer films under SEM/EDX after they were stored at room temperature for 30 days. From Table 8, obviously, a lot of sulfur crystals/bloom were observed on the rubber film surface for a lower rotational speed of bead-milling of 600 rpm at a higher flow rate of 750 L/h. From the microphotographs and the following spectrum table analysis in Table 9, sulfur can be seen to be one of the dominant elements present on the rubber films (as in column of Spectrum "S"). However, the percentage of presence differs for each different rubber film with different sulfur dispersion produced at designated different bead-milling parameter settings. Set A, with higher rotational speed and lower flow rate of bead-milling, has significantly lower sulfur content on the surface, and Set D, with lower rotational speed and higher flow rate of bead-milling, has higher sulfur content dominance on the rubber film. This analysis supports the observation from the LED magnifier lens results in Table 8.

According to the writings by [20,21], insoluble sulfur is usually used rather than soluble sulfur to overcome the "sulfur blooms" issue in unvulcanised rubber. This blooming can minimize the adhesion of rubber to the adjacent layer of rubber or substrate and consequently hinder the performance of the resulting molded rubber articles. Complete dissolution is normally achievable upon mixing the compound ingredients, including insoluble sulfur, because the mixing temperature is high enough to melt sulfur and prevent crystallization. However, when subjected to actual processing in industrial practice, normally unexpected blooms still occur due to the reverting of insoluble sulfur to the orthorhombic soluble sulfur well before the rubber is cured. Eventually, the sulfur supersaturates the rubber when the stock is cooled for a long period of time, and thus, the soluble portion of the sulfur forms crystals within the bulk rubber and initiates growth towards the surface and crystallizes there due to the lack of interference. This might be due to the instability of the sulfur dispersion when processed at this bead-milling parameter setting. From this observation, choosing a less suitable parameter set for sulfur processing could potentially destabilize insoluble sulfur to convert into soluble sulfur that, in turn, migrate to the surface of the rubber to produce blooms.

Fig. 4 illustrates the tensile strength mechanical properties of the cast rubber elastomer films across the four different sulfur curative sets, comparing their respective particle sizes. The mechanical properties underwent testing to assess the mechanical behavior of the rubber elastomer samples according to the ASTM D3578 standard and remarkably passed the minimum requirement of the standard ($\geq$ 18 MPa) [22]. The values for each set remained closely aligned, with slight improvements in tensile strength evident for Set A, which featured the smallest particle size, showing the highest tensile strength value, which apparently contributed to this positive impact.

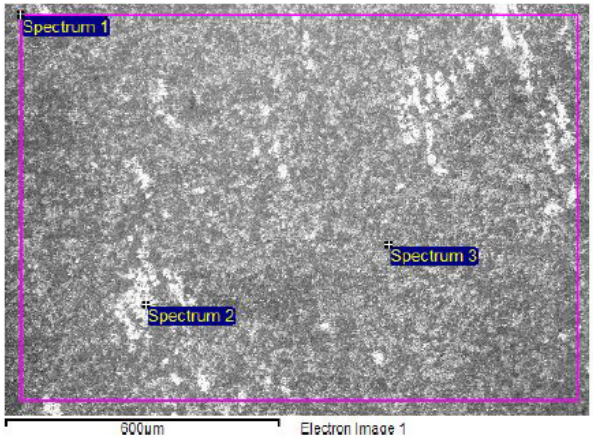
https://doi.org/10.31436/iiumej.v25i2.3098

Table 8. Visual observation of rubber film with different sulfur dispersion bead-milling parameters using an LED magnifier lens

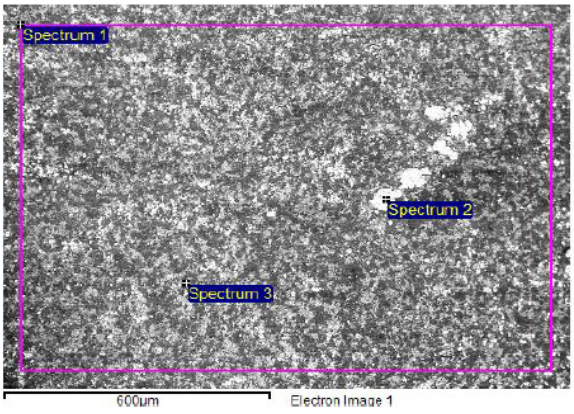
Sulfur Bead-milling Set	Pictorial observation; 3.0 x magnification	Visual Observation
Set A Rotational Speed: 800 rpm Flow Rate: 350 L/h		Less to zero sulfur crystals / bloom observed on the rubber film
Set B Rotational Speed: 700 rpm Flow Rate: 350 L/h		Less to zero sulfur crystals / bloom observed on the rubber film
Set C Rotational Speed: 700 rpm Flow Rate: 550 L/h		Less tiny particles of sulfur crystals / bloom observed on the rubber film
Set D Rotational Speed: 600 rpm Flow Rate: 750 L/h		A lot of sulfur crystals / bloom observed on the rubber film

https://doi.org/10.31436/iiumej.v25i2.3098

Table 9. SEM/EDX elemental analysis on rubber film of different sulfur dispersion bead-milling parameters

Sulfur Bead-Milling Set	**Sulfur-dispersed rubber film with area of sulfur crystals/bloom (Magnification: 200X)**
Set A Rotational Speed: 800 rpm Flow Rate: 350 L/h	

Spectrum	C	N	O	Mg	Al	**S**	Cl	Ca
Spectrum 1	62.71	0.00	23.76	0.17	0.24	**3.08**	0.34	9.71
Spectrum 2	58.96	0.00	23.27	0.19	0.26	**7.25**	0.30	9.77
Spectrum 3	59.24	0.00	24.33	0.20	0.29	**3.04**	0.40	12.50
Standard Deviation	1.71	0.00	0.43	0.01	0.02	**1.98**	0.04	1.30

Set B Rotational Speed: 700 rpm Flow Rate: 350 L/h	

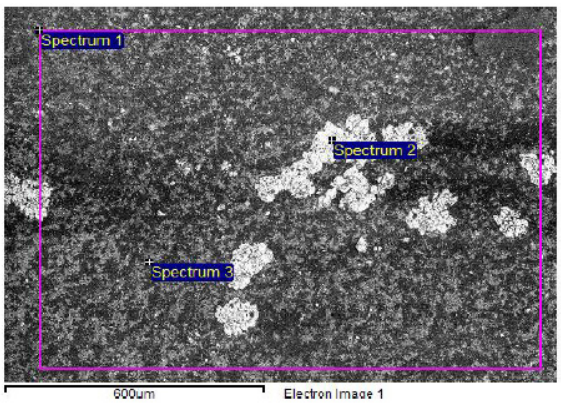
Spectrum	C	N	O	Mg	Al	**S**	Cl	Ca
Spectrum 1	65.61	0.00	21.95	0.15	0.19	**4.10**	0.21	7.79
Spectrum 2	54.17	0.00	27.58	0.18	0.24	**7.96**	0.34	9.22
Spectrum 3	62.47	0.00	22.71	0.14	0.20	**3.15**	0.25	11.09
Standard Deviation	4.83	0.00	2.49	0.02	0.02	**2.08**	0.05	1.35

https://doi.org/10.31436/iiumej.v25i2.3098

Set C

Rotational Speed: 700 rpm

Flow Rate: 550 L/h

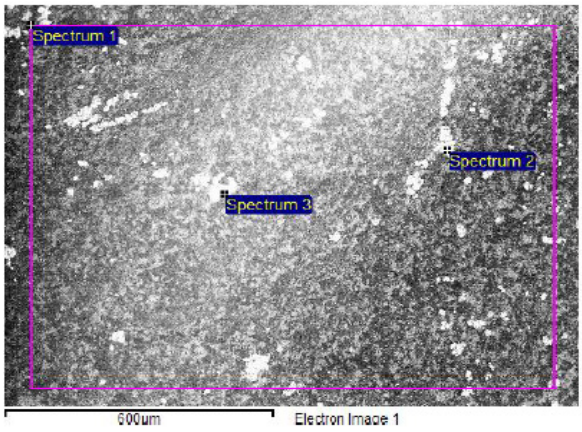


Spectrum	C	N	O	Mg	Al	**S**	Cl	Ca
Spectrum 1	64.67	0.00	21.90	0.14	0.19	**4.76**	0.36	7.98
Spectrum 2	56.84	0.00	6.68	0.07	0.00	**33.78**	0.14	2.49
Spectrum 3	58.37	0.00	29.32	0.25	0.03	**1.64**	0.28	10.10
Standard Deviation	3.39	0.00	9.42	0.07	0.08	**14.47**	0.09	3.21

Set D

Rotational Speed: 600 rpm

Flow Rate: 750 L/h



Spectrum	C	N	O	Mg	Al	**S**	Cl	Ca
Spectrum 1	62.09	0.00	23.45	0.24	0.07	**3.36**	0.33	10.46
Spectrum 2	58.83	0.00	10.38	0.09	0.04	**26.80**	0.00	3.85
Spectrum 3	58.16	0.00	7.01	0.03	0.10	**31.74**	0.15	2.81
Standard Deviation	1.72	0.00	7.09	0.09	0.02	**12.38**	0.13	3.39

https://doi.org/10.31436/iiumej.v25i2.3098

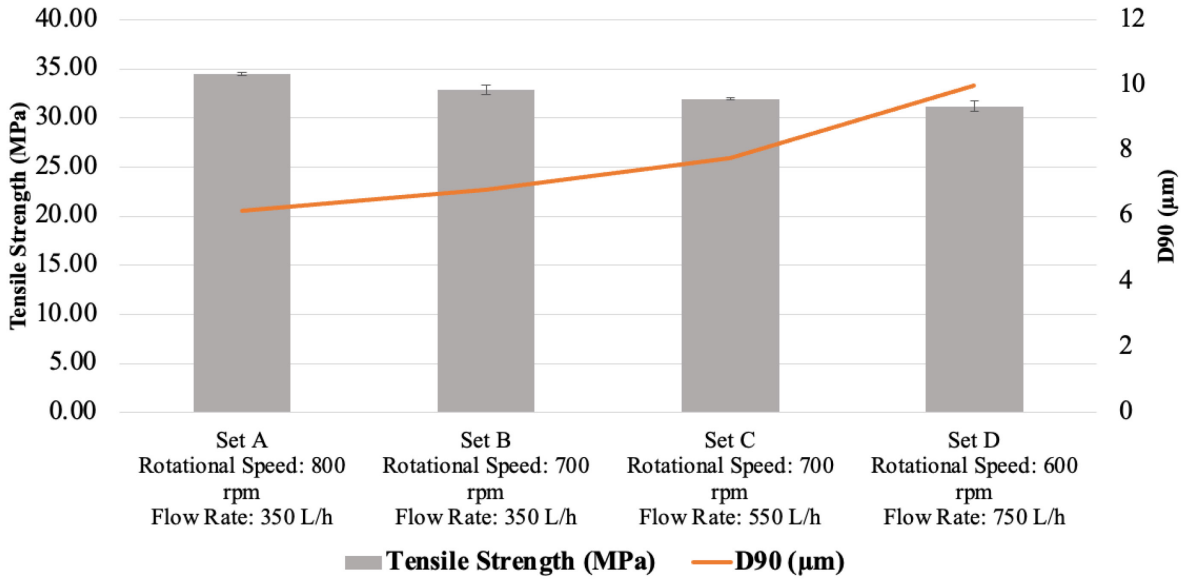


Figure 4. Tensile strength of cast rubber elastomer films with different bead-milled sulfur curative.

4. CONCLUSION

In conclusion, this study investigated the impact of bead-milling process parameters, namely rotational speed and flow rate, on sulfur curative dispersion and its subsequent effects on rubber elastomer properties. The findings revealed that higher rotational speed (800 rpm) and lower flow rate (350 L/h) resulted in smaller sulfur dispersion particle sizes, which slightly improved the tensile strength of rubber elastomers and significantly reduced the formation of sulfur crystals. This bead-milling process parameter setting also improved sulfur dispersion stability from the lowest instability index result. The visual observations and morphology analysis confirmed the influence of bead-milling parameters on the presence of sulfur crystals on rubber film surfaces. Overall, the research highlights the potential for eco-friendly processes such as bead-milling to optimize overall properties and material characteristics of dispersion products in various industrial applications, particularly in elastomeric rubber composites.

ACKNOWLEDGEMENT

Mohamad Firdaus Omar would like to extend his appreciation to the Department of Chemical Engineering and Sustainability in IIUM that made this work possible. The authors gratefully acknowledge the financial support of research funding from Universiti Teknologi Malaysia, UTM Fundamental Research vote no. 22H56.

REFERENCES

[1] Tan, X, Zhang, B, Chen, L, Li, X, Li, L, Xie, F. (2015) Effect of Planetary Ball-Milling on Multi-Scale Structures and Pasting Properties of Waxy and High-Amylose Cornstarches. Innovative Food Science & Emerging Technologies, 30: 198–207. https://doi.org/10.1016/j.ifset.2015.03.013.

[2] Páramo-Calderón, DE, Vázquez-León, LA, Palma-Rodríguez, HM, Utrilla-Coello, RG, Vargas-Torres, A, Meza-Nieto, MA, Romero-Cortes, T, Aparicio-Saguilán, A. (2023) Effect of High-Energy Mechanical Milling on the Physicochemical and Rheological Properties of Chayotextle (Sechium Edule Sw.) Starch. Food Chemistry, 427: 136720. https://doi.org/10.1016/j.foodchem.2023.136720.

[3] Bangar, SP, Singh, A, Ashogbon, AO, Bobade, H. (2023) Ball-Milling: A Sustainable and Green Approach for Starch Modification. International Journal of Biological Macromolecules, 237: 124069. https://doi.org/10.1016/j.ijbiomac.2023.124069.

[4] Ahmadian, H, Sallakhniknezhad, R, Zhou, T, Kiahosseini, SR. (2022) Mechanical Properties of Al-Mg/MWCNT Nanocomposite Powder Produced under Different Parameters of Ball Milling Process. Diamond and Related Materials, 121: 108755. https://doi.org/10.1016/j.diamond.2021.108755.

[5] Liu, Z, Zheng, Z, Zhu, G, Luo, S, Zhang, D, Liu, F, Shen, Y. (2021) Modification of the Structural and Functional Properties of Wheat Gluten Protein Using a Planetary Ball Mill. Food Chemistry, 363: 130251. https://doi.org/10.1016/j.foodchem.2021.130251.

[6] Carling, C-J, Pekkari, A. (2023) Bead Milled Drug Nanocrystal Suspensions FIne Enough to Pass through 0.22 Mm Sterilization FIlters. https://doi.org/https:/example.net.

[7] Liu, W, Xie, H, Guo, X, Wang, K, Yang, C, Wang, N, Ge, C. (2023) Luminescence, Stability, and Applications of CsPbBr3 Quantum Dot/Polymethyl Methacrylate Composites Prepared by a Solvent- and Ligand-Free Ball Milling Method. Optical Materials, 136: 113398. https://doi.org/10.1016/j.optmat.2022.113398.

[8] Motavalizadehkakhky, A, Shahrampour, H. (2018) Determination of the Most Efficient Form of Sulfur for Use as a Natural Rubber Curing Agent by Comparison of Physical and Thermal Attributes of Cured Rubber. Petroleum Chemistry, 58: 89–93. https://doi.org/10.1134/S0965544118010048.

[9] Shahrampour, H, Motavalizadehkakhky, A. (2017) The Effects of Sulfur Curing Systems (Insoluble-Rhombic) on Physical and Thermal Properties of the Matrix Polymeric of Styrene Butadiene Rubber. Petroleum Chemistry, 57: 700–704. https://doi.org/10.1134/S0965544117080138.

[10] Pangamol, P. (2016) Use of By-Product Sulfur from Petroleum Refinery as Vulcanizing Agent in Natural Rubber. Chiang Mai J. Sci., 43: 569–576. http://epg.science.cmu.ac.th/ejournal/.

[11] Zheng, L, Jerrams, S, Xu, Z, Zhang, L, Liu, L, Wen, S. (2020) Enhanced Gas Barrier Properties of Graphene Oxide/Rubber Composites with Strong Interfaces Constructed by Graphene Oxide and Sulfur. Chemical Engineering Journal, 383: 123100. https://doi.org/10.1016/j.cej.2019.123100.

[12] Duan, X, Tao, R, Chen, Y, Zhang, Z, Zhao, G, Liu, Y, Cheng, S. (2022) Improved Mechanical, Thermal Conductivity and Low Heat Build-up Properties of Natural Rubber Composites with Nano-Sulfur Modified Graphene Oxide/Silicon Carbide. Ceramics International, 48: 22053–22063. https://doi.org/10.1016/j.ceramint.2022.04.196.

[13] Sopicka-Lizer, M. (2010) High-Energy Ball Milling. Woodhead Publishing Limited. https://doi.org/10.1533/9781845699444.

[14] Flach, F, Konnerth, C, Peppersack, C, Schmidt, J, Damm, C, Breitung-Faes, S, Peukert, W, Kwade, A. (2016) Impact of Formulation and Operating Parameters on Particle Size and Grinding Media Wear in Wet Media Milling of Organic Compounds – A Case Study for Pyrene. Advanced Powder Technology, 27: 2507–2519. https://doi.org/10.1016/j.apt.2016.09.026.

[15] Tamblyn, RJ. (2009) Analysis of Energy Requirements in Stirred Media Mills.

[16] Ullah, M, Ali, Md, Hamid, S. (2014) Structure-Controlled Nanomaterial Synthesis Using Surfactant-Assisted Ball Milling- A Review. Current Nanoscience, 10: 344–354. https://doi.org/10.2174/15734137113096660114.

[17] Sudheer, M, Prabhu, R, Raju, K, Bhat, T. (2013) Modeling and Analysis for Wear Performance in Dry Sliding of Epoxy/Glass/PTW Composites Using Full Factorial Techniques. ISRN Tribology, 2013: 1–11. https://doi.org/10.5402/2013/624813.

[18] Achuthamenon Sylajakumari, P, Ramakrishnasamy, R, Palaniappan, G. (2018) Taguchi Grey Relational Analysis for Multi-Response Optimization of Wear in Co-Continuous Composite. Materials, 11: 1743. https://doi.org/10.3390/ma11091743.

[19] Yerramilli, M, Ghosh, S. (2017) Long-Term Stability of Sodium Caseinate-Stabilized Nanoemulsions. Journal of Food Science and Technology, 54: 82–92. https://doi.org/10.1007/s13197-016-2438-y.

[20] Aziz, Y, Hepburn, C. (1979) Problems of Bloom Experienced When Insoluble Sulphur Is Used in Natural Rubber. Journal of the Rubber Research Institute of Malaysia, 2: 57–67.

[21] Jurkowski, B, Jurkowska, B. (1998) On the Mechanism of Sulfur Behavior in Rubber Compounds. Journal of Macromolecular Science, Part B, 37: 135–142. https://doi.org/10.1080/00222349808220461.

[22] Ab Rahman, MF, Norfaizal, NS, Azura, AR. (2019) The Influence of Sago Starch Dispersion on Mechanical Properties of Biodegradable Natural Rubber Latex Films. Materials Today: Proceedings, 17: 1040–1046. https://doi.org/10.1016/j.matpr.2019.06.507.

https://doi.org/10.31436/iiumej.v25i2.3240

EFFECT OF INCORPORATING CELLULOSE NANOFIBERS AND LEMONGRASS ESSENTIAL OIL IN POLYVINYL ALCOHOL-POLYETHYLENE GLYCOL/GLYCERIN HYDROGEL FOR WOUND DRESSING

NUR HUDA SYAZWANI JAFRI[1], DZUN NORAINI JIMAT[1*], WAN MOHD FAZLI WAN NAWAWI[1], YUSILAWATI AHMAD NOR[1], AZURA AMID[2]

[1]Department of Chemical Engineering and Sustainability, Faculty of Engineering, International Islamic University Malaysia, P.O Box 10, 50728, Kuala Lumpur, Malaysia

[2]International Institute for Halal Research and Training, International Islamic University Malaysia, P. O Box 10, 50728, Kuala Lumpur, Malaysia

**Corresponding author: jnoraini@iium.edu.my*

(Received: 27 March 2024; Accepted: 27 June 2024; Published online: 15 July 2024)

ABSTRACT: Hydrogels attract increased interest as wound dressings due to their biomimetic properties, creating a moist environment conducive to natural wound healing. In this study, a PVA-PEG/gly-CNF-LG hydrogel incorporating cellulose nanofibers (CNF) and lemongrass essential oil (LG) into the polyvinyl alcohol-polyethylene glycol/glycerin (PVA-PEG/gly) hydrogel via the freeze-thaw method was developed. The addition of CNFs and LG aimed to improve the physicochemical and antibacterial aspects of the hydrogel. Optimal hydrogel composition, determined through response surface methodology (RSM) and central composite design (CCD), consisted of 3.5% (w/v) CNFs and 3% (v/v) LG concentrations, resulting in an optimal moisture retention capability (MRC) of 37.69 ± 0.54%. The optimized PVA-PEG/gly-CNF-LG demonstrated impressive characteristics: a swelling capacity of 176.89 ± 1.56%, a gel fraction of 78.89 ± 0.42%, and a porosity of 47.51 ± 0.53%. FESEM images revealed the relatively porous nature of PVA-PEG/gly-CNF-LG hydrogels. Furthermore, the hydrogel exhibited excellent resistance against S. aureus and B. subtilis bacteria, along with notable tensile properties of 1.44 MPa. These findings underscore the promising attributes of the PVA-PEG/gly-CNF-LG hydrogel, positioning it as a versatile and effective wound-healing dressing with significant antimicrobial properties.

ABSTRAK: Hidrogel mendapat perhatian ramai sebagai pembalut luka di sebabkan oleh ciri-ciri biomimik, di mana menghasilkan persekitaran lembab yang baik bagi penyembuhan luka secara semula jadi. Kajian ini, mencadangkan hidrogel PVA-PEG/gly-CNF-LG yang mengandungi selulosa nanofiber (CNF) dan minyak pati serai (LG) dalam hidrogel polivenil alkohol-polietilin glikol/gliserin (PVA-PEG/gly) melalui kaedah beku-cair. Penambahan CNFs dan LG diperlukan bagi memperbaiki aspek fisiokimia dan antibakterial hidrogel. Komposisi optimal hidrogel, dibentuk melalui kaedah respons permukaan (RSM) dan reka bentuk komposit pusat (CCD), mengandungi 3.5% (w/v) CNFs dan 3% (v/v) kepekatan LG, menghasilkan kemampuan retensi kelembapan optimal (MRC) sebanyak 37.69 ± 0.54%. Kadar optimum PVA-PEG/gly-CNF-LG menunjukkan ciri-ciri yang menarik: iaitu kapasiti pembengkakan sebanyak 176.89 ± 1.56%, pecahan gel sebanyak 78.89 ± 0.42%, dan keliangan 47.51 ± 0.53%. Imej FESEM menunjukkan sifat keliangan semula jadi hidrogel PVA-PEG/gly-CNF-LG. Tambahan, hidrogel memiliki rintangan tinggi terhadap bakteria S. aureus dan B. subtilis, sejajar dengan ciri-ciri ketara tegangan 1.44 MPa. Dapatan kajian ini penting bagi ciri-ciri hidrogel yang berpotensi besar seperti PVA-PEG/gly-CNF-LG,

menjadikannya serba guna dan berkesan sebagai balutan penyembuhan luka dengan ciri-ciri antimikrob yang ketara.

KEYWORDS: *Cellulose Nanofibers, Glycerin, Lemongrass essential oil, Polyethylene glycol, Polyvinyl alcohol*

1. INTRODUCTION

Polyvinyl alcohol (PVA), a synthetic polymer, shows significant promise for wound dressing applications due to its hydrophilic nature and high biocompatibility [1–3]. PVA-based hydrogel dressings demonstrate exceptional moisture retention ability by mimicking natural tissue conditions, thereby establishing and maintaining the moist environment necessary for faster healing processes [4]. Moreover, PVA hydrogels are easily synthesized by adjusting their structure, composition, and interactions to meet specific requirements. They can also be combined with bioactive agents such as antibiotics and growth factors to enhance therapeutic effectiveness. These combined attributes render PVA hydrogels versatile and valuable materials in the field of wound care and other medical applications [2, 5].

Despite their beneficial properties, PVA hydrogels have several drawbacks that limit their use as wound dressings. Their relatively weak mechanical properties make them susceptible to breakage and deformation, thus failing to adequately support and protect wounds [2, 5]. Reduced flexibility contributes to the challenge of PVA hydrogel dressings adhering properly to wounds and surrounding skin. Additionally, the rapid hydration of PVA hydrogels can lead to the production of stiff, dry dressings that hinder their ability to maintain a sufficiently hydrated environment, which is crucial for accelerating wound healing [6, 7].

To enhance wound dressing performance and provide effective wound care, it is essential to modify the hydrogel composition to mitigate the drawbacks of PVA hydrogel wound dressings [8, 9]. Various components can be blended together in a unified or single polymeric matrix to create materials with desired properties [10, 11]. Polyethylene glycol (PEG) is renowned for its remarkable hydrophilic properties, significantly enhancing the water retention and absorption capacity of hydrogel formulations when incorporated into the hydrogel matrix to promote wound healing. This feature is crucial for preventing the production of rigid and dry hydrogels [8, 12]. Additionally, PEG has low toxicity and is highly biocompatible, making it safe for use in contact with biological tissue. This makes it a common polymer choice in wound dressing applications [2].

The inclusion of glycerin in hydrogel formulation is advantageous due to its excellent humectant and moisturizing characteristics [6, 13]. Glycerin enhances the hydrating action of PEG when added to PVA hydrogel by attracting and retaining water molecules within the hydrogel matrix. This feature is vital to maintain the proper moisture content of the wound dressing and prevent it from drying out [13]. Glycerin can also improve the flexibility and elasticity of PVA hydrogel dressings, allowing them to better conform to the shape of wounds, ensuring proper adherence, and maximizing the healing process [13].

Cellulose nanofibers (CNFs) are fibrous materials with nanoscale dimensions. Besides being lightweight, CNFs exhibit outstanding mechanical properties, making them highly desirable for reinforcement in wound dressings [14, 15]. CNFs serve as reinforcing agents that help strengthen the hydrogel structure, improving the structural integrity of hydrogels. This feature is vital for forming durable hydrogel dressings resistant to degradation and deformation [15, 16]. CNFs can be derived from empty fruit bunches (EFB), the major biomass waste of

palm oil mills. Due to its high cellulose content, EFB can be utilized as a feedstock for large-scale nanocellulose production, providing a sustainable alternative [15].

Lemongrass essential oil (LG) exhibits potent antibacterial properties due to bioactive compounds such as citral and geraniol, which can combat a wide range of pathogens. Incorporating LG into the hydrogel formulation imparts antibacterial properties to the hydrogel dressings, reducing the risk of infection and accelerating wound healing [17, 18]. Moreover, LG possesses other therapeutic qualities that promote wound healing, such as anti-inflammatory and antioxidant actions, enhancing the effectiveness of hydrogel wound dressings [17, 19].

There are several methods for making hydrogels, including various crosslinking techniques such as physical, radiation, and chemical methods [20]. The most commonly employed technique is the freeze-thaw method due to its ease of synthesis, cost-effectiveness, and environmentally friendly characteristics. This method does not use chemicals, thereby avoiding toxicity and leaching problems caused by chemical crosslinking. It also enables the production of hydrogels with tailored properties [21].

This study aimed to investigate the properties of polyvinyl alcohol-polyethylene glycol/glycerin (PVA-PEG/gly) hydrogels incorporated with cellulose nanofibers (CNFs) and lemongrass essential oil (LG). This hydrogel, referred to as PVA-PEG/gly-CNF-LG, utilizes central composite design (CCD) within response surface methodology (RSM) to optimize CNF and LG concentrations to achieve maximum moisture retention capability (MRC) of the hydrogel. The optimized hydrogel compositions are evaluated specifically for their swelling properties, gel fraction, and porosity. Additionally, morphological analysis is performed using field emission scanning electron microscopy (FESEM), while Fourier transform infrared spectroscopy (FTIR) is used to analyze functional groups within the hydrogel structure. Furthermore, the antibacterial activity and mechanical strength of the hydrogel are determined via the disc diffusion method and tensile testing, respectively.

2. MATERIALS AND METHODS

2.1 Materials

Dimethyl sulfoxide (DMSO; ≥99.5%) and glycerin (≥99.5%) were provided by R&M Chemical Co., Ltd. Polyethylene glycol (PEG; MW 3350) and polyvinyl alcohol (PVA; MW 89,000-98,000; 99% hydrolyzed) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Luria-Bertani (LB) agar and broth (MILLER) were purchased from Sigma-Aldrich, and tetracycline hydrochloride was sourced from Fisher Scientific, UK. Lemongrass essential oil (LG) from the Soul brand was supplied by BF1 Cosmetic Co., Ltd (Kuala Lumpur, Malaysia).

Empty fruit bunches (EFB) from Kilang Sawit FELCRA Maran in Pahang, Malaysia, were processed for cellulose nanofiber (CNF) extraction. EFB underwent cutting, crushing into small particles using a mini grinder and sieving to produce fine powders (0.5 to 2 mm). Fibers were then heated in distilled water at 80 °C, stirred for 1 h, dried overnight at 40 °C until they reached a constant weight, and stored in airtight containers. The chemicals utilized for CNF extraction were laboratory-grade and sourced from Sigma-Aldrich, USA. These include oxalic acid dihydrate ($C_2H_2O_4$) (≥ 99%), choline chloride ($C_6H_{14}CINO$) (≥ 98%), 98% sodium hydroxide pellets (NaOH), and 35% hydrogen peroxide (H_2O_2). All chemicals used in this research were of analytical grade and underwent no further purification.

2.2. Extracting CNF from Empty Fruit Bunches

The extraction of CNFs from EFB followed the method described by Asem et al. with several modifications [22]. Approximately 30 g of EFB fibers were boiled in a 4% (w/v) NaOH solution at 85 °C with a solid-to-solvent ratio of 1:30 and continuously stirred for 2 hours. The fibers were then filtered, washed with deionized water until reaching a neutral pH, and subsequently bleached for 3 h at 95 °C using a 1:1 mixture of deionized water and H_2O_2. This bleaching process was repeated three times. After bleaching, the fibers underwent multiple rinses with distilled water until the wash pH reached neutrality. They were then filtered using a vacuum pump to acquire solid residues. Subsequently, these solid residues were dried at 50 °C until reaching a constant weight. The obtained solid residue is referred to as EFB cellulose [22].

Next, the EFB cellulose was dissolved in a DES mixture of choline chloride-oxalic acid dehydrates (ChCl-OAD) at a DES molar ratio of 1:2 with a fiber-to-solution ratio of 1:10 (w/v), and heated to 90 °C for 90 min with continuous stirring. The cellulose suspensions obtained were centrifuged repeatedly at 9,000 rpm for 15 min in deionized water until reaching a neutral pH. They were then subjected to ultrasonication (using FisherbrandTM Model 705 Sonic Dismembrator) for 2 h at 80% amplitude (280 W). The CNF suspension underwent a 15-minute centrifugation at 10,000 rpm to isolate the CNFs from the liquid. The collected CNFs were then stored in an airtight container in a refrigerator set at 4 °C.

2.3 Synthesis of PVA-PEG/gly-CNF-LG Hydrogel

The hydrogel was prepared following the method outlined by Altaf et al. with modifications [6]. Initially, a 20% (v/v) DMSO solution was prepared by mixing it with deionized water at 45 °C for 15 min. Subsequently, a 15% (w/v) PVA solution in DMSO was prepared and heated to 90 °C for 1 h with constant stirring until translucent. Once complete PVA dissolution was achieved, 6% (w/v) PEG was gradually added, and the resulting solution was agitated for 1 h at 90 °C to form a homogenized PVA-PEG polymeric solution. Glycerin (gly) at a concentration of 4% (w/v) was then incrementally added, heated to 95 °C, and vigorously agitated for 1 h.

CNFs and LG were blended into the polymeric solution at various compositions according to the central composite design (CCD) to optimize their concentrations. Similarly, CNFs were dissolved in deionized water at various concentrations and heated for 30 min at 95 °C with mechanical stirring, aligning with the optimization of the hydrogel composition. Then, the CNFs solution was mixed with the PVA-PEG/glycerin polymer solution in a 1:1 ratio and stirred at 95 °C for 1 h. Afterward, LG was added to the polymeric solution at different concentrations and stirred continuously for 30 min at a constant temperature, aligning with hydrogel composition optimization. The resulting solution was transferred into a petri dish and cooled to room temperature to solidify into a gel. Finally, the hydrogel underwent three cycles of freezing (18 h at -20 °C) and thawing (2 h at ambient temperature). The resulting hydrogel was designated as PVA-PEG/gly-CNF-LG hydrogel.

2.4 Experimental Design and Optimization

Experimental design, data processing, optimization, and modeling were conducted using Design-Expert® Version 12 software (Stat-Ease Inc., Minneapolis, MN). Response surface methodology (RSM) was employed to identify the optimal hydrogel composition for maximum moisture retention capability (MRC) through central composite design (CCD). CNF and LG concentrations were chosen as independent variables, while MRC served as the response or dependent variable. Table 1 provides the details of the CCD design.

https://doi.org/10.31436/iiumej.v25i2.3240

Table 1. CCD design of each variable with its corresponding ranges.

Variables	Levels		
	-1	0	+1
CNF concentration [A] (% w/v)	1	3.5	6
LG concentration [B] (% v/v)	1	3.0	5

Each independent variable was assessed at three levels (-1, 0, +1), resulting in 13 experimental runs with five replications, including center points, four axial points, and four cube points. The response, measured moisture retention capability (MRC), was fitted to a second-order polynomial model. Model validity was assessed through analysis of variance (ANOVA) at a confidence level of 95%. Three-dimensional surface plots were employed to illustrate the effects and interactions of variables on responses. In the validation study, experimental results falling within the 95% confidence interval (*CI*) were utilized to confirm the projected optimum hydrogel formulation by RSM.

2.5 Field Emission Scanning Electron Microscope

The surface morphology of the produced hydrogel was examined using a field emission scanning electron microscope (FESEM) (Model Supra 55VP, Carl Zeiss AG, Germany). The sample was coated with platinum prior to microscopic examination. The hydrogel's microstructure was observed under standard conditions at a magnitude of 50 kV.

2.6 Fourier Transform Infrared Spectroscopy

The functional groups and chemical bonds present in the produced hydrogels were analyzed using a Spectrum RX Fourier Transform Infrared (FTIR) spectrophotometer (PerkinElmer, Norwalk, USA). The FTIR spectra of the samples were recorded with a resolution of 4 cm^{-1} within the spectral transmittance region spanning from 400 to 4000 cm^{-1}.

2.7 Gel Fraction

The hydrogel samples were cut into uniform pieces (2 cm width × 2 cm length) with a thickness of 0.3 cm. Subsequently, they underwent drying for 24 hours in a 40°C oven until reaching a constant weight (W_o). Next, they were submerged in deionized water and allowed to swell for 48 h at room temperature until equilibrium swelling was achieved, effectively eliminating any soluble materials. Afterward, the sample was dried at 40 °C for 24 h to attain a stable weight (W_e). The percentage of gel fraction (GF) was evaluated using Equation 1.

$$\% \text{ GF} = (W_e/W_o) \times 100 \quad (1)$$

Here, W_o is the weight of the dried hydrogel, and W_e is the weight of the swollen hydrogels.

2.8 Swelling Behavior Measurement

The hydrogels, with dimensions of 2 cm length, 2 cm width, and 0.3 cm thickness, were prepared to assess their swelling properties. Initially, the initial weight (W_d) of the hydrogels was recorded before soaking them in deionized water for 24 h at room temperature. Afterwards, the swollen samples were retrieved, and excess water was removed by blotting with filter paper. They were then weighed again (W_s). Equation 2 was utilized to calculate the swelling ratio.

$$\text{Swelling ratio } (\%) = \frac{(W_s - W_d)}{W_d} \times 100 \quad (2)$$

https://doi.org/10.31436/iiumej.v25i2.3240

Here, W_d represents the weight of the dried hydrogels, and W_s represents the weight of the swollen hydrogels.

2.9 Moisture Retention Capability

The hydrogel samples, standardized to dimensions of 2 cm in width, 2 cm in length, and 0.3 cm in thickness, were initially weighed, placed in a 40 °C oven for 6 h, and then weighed again. The moisture retention capability (MRC) was calculated using Equation 3.

$$\text{MRC}\ (\%) = (W_f/W_i) \times 100 \quad (3)$$

where, W_i denotes the starting weight of hydrogel samples, whereas W_f signifies the weight after heating at 40 °C for 6 h.

2.10 Hydrogel Membrane Porosity

The hydrogel samples, standardized to dimensions of 2 cm in width, 2 cm in length, and 0.3 cm in thickness, were immersed in ethanol for 24 h until saturated at room temperature. Measurements were taken before and after ethanol absorption to determine the hydrogel membrane porosity (φ) using Equation 4.

$$\varphi\ (\%) = (W_f - W_i)/(\rho V_f - \rho V_i) \times 100 \quad (4)$$

where, W_i and W_f represent the mass of samples before and after ethanol absorption, respectively. Additionally, V_i and V_f are the ethanol volume before and subsequent absorption, respectively. ρ is the density of ethanol at room temperature.

2.11 Zeta Potential Measurement

The zeta potential of hydrogel samples was analyzed using the Zetasizer Nanoparticle Analyzer (Nano ZS, Model ZEN 3600, Malvern Instrument Ltd., UK). Dynamic light scattering (DLS) was employed to measure samples at 25 °C, utilizing Helium-Neon (He-Ne) lasers emitting a 633 nm wavelength and a 176° backscatter detector.

2.12 Antibacterial Activity Assessment

The antibacterial activity was assessed using a disc diffusion method inspired by Altaf et al. (2021) with modifications to evaluate the hydrogel's effectiveness against bacteria. Various strains, including *Bacillus subtilis (B. subtilis)*, *Staphylococcus aureus (S. aureus)*, *Escherichia coli (E. coli)*, and *Salmonella enteritidis (S. enteritidis)*, were employed. A loopful of bacterial cells from stock cultures was transferred into sterile Luria-Bertani (LB) broth and incubated at 37 °C for 24 h to produce an inoculum of 10^6 CFU/mL.

Hydrogel samples (1 cm × 1 cm, 0.3 cm thickness) were placed on LB agar plates containing bacteria, sterilized with UV light for 20 min, and then incubated for 24 h at 37 °C. The diameter of the inhibition zones measured bacterial growth inhibition, and Vernier calipers were used to quantify the zone of inhibition after removing the agar plates from the incubator. Tetracycline hydrochloride and DMSO served as positive and negative controls, respectively. All apparatus used in the antibacterial test were autoclaved to prevent contamination.

The hydrogel samples (measuring 1 cm × 1 cm, with a thickness of 0.3 cm) were placed on LB agar plates containing bacteria, sterilized with UV light for 20 min, and then incubated for 24 h at 37 °C. After removing the agar plates from the incubator, the diameter of the inhibition zones, which indicated bacterial growth inhibition, was measured using Vernier calipers. Tetracycline hydrochloride and DMSO served as positive and negative controls,

respectively. All apparatus used in the antibacterial test were autoclaved to prevent contamination.

2.12 Mechanical Testing

The initial gap between the sample and grips was adjusted to 40 mm. Tensile testing was conducted to assess the mechanical properties of hydrogels following ASTM D638 standards, employing the electronic universal testing machine (Instron® Model 3366, UK) equipped with a 10 kN load cell. Initially, the hydrogels were shaped into standard dumbbell forms with dimensions of 6 mm in height, 40 mm in diameter, and 3 mm in thickness. The tensile tests were performed with 90% compression at a strain rate of 10 mm min^{-1}. The initial gap between the sample and grips was set to 40 mm.

3. RESULTS AND DISCUSSION

3.1 Statistical Optimization of Hydrogel Composition

The study examined the synergistic impact of CNFs and LG concentrations on the MRC of the hydrogel utilizing RSM with a CCD, as detailed in Table 2. Thirteen experimental runs were carried out for hydrogel compositions to evaluate the influence of these two factors on the MRC of the hydrogel.

Table 2. CCD for hydrogel composition.

Run	CNF concentration [A] (% w/v)	LG concentration [B] (% v/v)	MRC (%)
1	3.5	5	31.03
2	1	1	27.20
3	6	3	32.25
4	3.5	3	37.03
5	6	5	25.46
6	3.5	3	37.68
7	3.5	1	32.46
8	6	1	30.89
9	3.5	3	36.36
10	1	3	32.74
11	3.5	3	37.36
12	1	5	30.00
13	3.5	3	37.69

Mean data values were employed to formulate a second-order polynomial equation, resulting in a regression equation to determine the optimal MRC within the experimental limits. The linear equation for MRC (Y) depicted the experimental data, as presented in Equation 5.

$$Y = +15.233 + 5.700A + 8.327B - 0.412AB - 0.651A^2 - 1.204B^2 \quad (5)$$

In this context, Y represents the MRC, while A and B denote the independent variables for CNFs and LG concentrations, respectively.

Positive coefficients indicated that the chosen independent variables, CNFs and LG concentrations, positively influenced the hydrogel's MRC. This suggested that within the specified experimental range, higher concentrations of CNFs and LG led to increased MRC of the hydrogel, highlighting a favorable correlation between these variables and the desired outcome [7, 23].

https://doi.org/10.31436/iiumej.v25i2.3240

The model's effectiveness and statistical significance were assessed through variance analysis (ANOVA), confirming its adequacy, as delineated in Table 3. ANOVA aided in validating the constructed model by maintaining the hierarchical structure and identifying and eliminating irrelevant interactions [6]. The model's probability value (p-value) had to be below 0.05 to be considered significant [7, 24].

Table 3. ANOVA of hydrogel formulation.

Source	Sum of squares	DF	Mean square	F-value	p-value	
Model MRC (Y)	196.60	5	39.32	108.76	<0.0001	Significant
A-CNF concentration	0.2993	1	0.2993	0.8278	0.3932	
B-LG concentration	2.75	1	2.75	7.60	0.0282	
AB	16.93	1	16.93	46.84	0.0002	
A^2	45.68	1	45.68	126.35	<0.0001	
B^2	64.08	1	64.08	177.25	<0.0001	
Residual	2.53	7	0.3615	-	-	
Lack of Fit	1.30	3	0.4344	1.42	0.3617	Not significant
Pure Error	1.23	4	0.3069	-	-	
Cor Total	199.13	12	-	-	-	
R^2	0.9873		Adequate Precision		28.988	
Adjusted R^2	0.9878		C.V. %		1.83	
Predicted R^2	0.9463		Standard Deviation %		0.6013	

Table 3 indicated that the model demonstrated significance with a low p-value ($p < 0.0001$) and an F-value of 108.76. The likelihood of noise resulting in such a high F-value was only 0.01%. All terms (B, AB, A^2, and B^2) exhibited lower p-values, designating them as significant model terms, except for term A (CNFs concentration). The regression coefficients R^2, adjusted R^2, and estimated R^2 aided in evaluating the model's fit to the experimental data, with values closer to 1 indicating a superior fit [25]. The model achieved a substantial R^2 of 0.9873, ensuring an accurate fit to the experimental data. The predicted R^2 (0.9463) closely matched the adjusted R^2 (0.9878), suggesting that the insignificant term had minimal impact on the model [3]. The insignificant lack of fit affirmed the model's adequacy, demonstrating a robust correlation between process variables and output response [25].

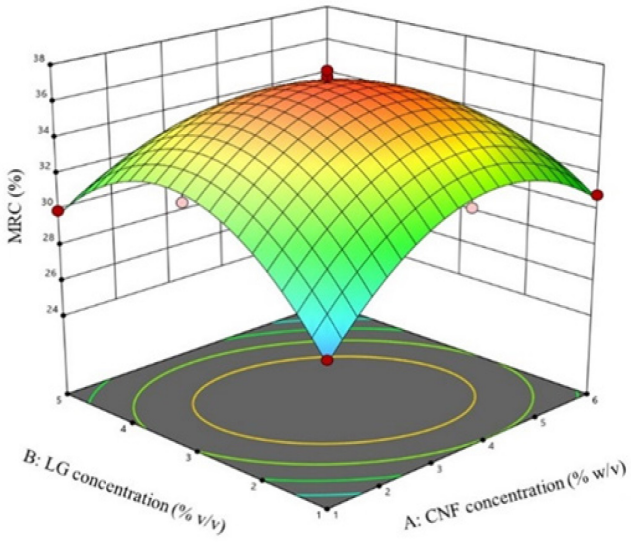


Figure 1. The response surface graph showed the impact of CNF and LG concentrations on the MRC of the hydrogel.

Fig. 1 illustrates the influence of CNFs and LG concentrations on the hydrogel's MRC through three-dimensional response surface graphs. Circular shapes represent insignificant interactions, whereas elliptical forms indicate more pronounced interactions [25, 26]. The downward-facing curve and symmetrical shapes confirm a robust correlation between CNF and LG concentrations and the hydrogel's MRC [26].

Through RSM optimization, the ideal hydrogel composition was determined to be 3.5% (w/v) CNF and 3% (v/v) LG concentrations, resulting in the highest MRC of 37.69%. Building upon this optimal formulation, a validation experiment yielded a hydrogel with an MRC of 37.82 ±0.54%, which closely aligns with the predicted MRC value of 37.06%. This validation underscores the efficacy of the RSM model for hydrogel formulation [7]. Hydrogels must exhibit substantial moisture retention capabilities, essential for creating and maintaining a moist environment that fosters optimal healing processes and ensures effective wound dressing [27, 28].

3.2 Swelling Behavior Assessment

The ability of the hydrogel to absorb a significant volume of wound exudates and biological fluids serves as a criterion for assessing its effectiveness as a wound dressing [29, 30]. As depicted in Fig. 2b, the swelling capacities of various hydrogels are presented. The optimized PVA-PEG/gly-CNF-LG hydrogel displayed the highest swelling capacity at 176.89 ± 1.56%, surpassing that of PVA and PVA-PEG/gly hydrogels, which measured 154.4 ± 0.46% and 143.24 ± 1.66%, respectively. A previous study on PVA/PEG hydrogels indicated a lower swelling capacity than the PVA-PEG/gly-CNF-LG hydrogel [2]. Moreover, a hydrogel with a swelling capacity exceeding 100% is typically considered a highly absorbent polymer [7].

PVA and PVA-PEG/gly-CNF-LG hydrogels displayed distinct characteristics attributed to differences in their compositions. For example, due to its flexible nature, the PVA-PEG/gly-CNF-LG hydrogel demonstrated the ability to bend or stretch without breaking. Conversely, PVA hydrogel typically exhibited a solid-like structure that was rigid and susceptible to deformation. The flexible attributes the PVA-PEG/gly-CNF-LG hydrogel exhibited were attributed to the incorporation of moisture-retaining components such as PEG and glycerin into the hydrogel matrix [8, 31].

3.3 Gel Fraction Measurement

Gel fraction (GF) is a critical parameter that can be used to quantify the degree of crosslinking among the constituents employed in producing hydrogels within their polymeric matrix [6]. The GFs of hydrogels produced in this study are depicted in Fig. 2a. The optimized PVA-PEG/gly-CNF-LG hydrogel exhibited the highest GF of 78.89 ± 0.42%. In comparison, the GFs for PVA and PVA-PEG/gly hydrogels were 36.99 ± 0.34% and 58.06 ± 1.65%, respectively. The PVA-PEG/gly-CNF-LG hydrogel demonstrated a higher gel fraction than an earlier study with the PVA/starch hydrogel containing clove oils, where the gel fraction was reported at 64% [6].

For the optimized PVA-PEG/gly-CNF-LG hydrogel, the heightened gel fraction validates successful cross-linking interactions among its constituents, including PVA, PEG, glycerin, CNFs, and LG. This intricate cross-linking resulted in a robust and cohesive three-dimensional polymer network within the hydrogel matrix, significantly enhancing its mechanical strength, structural integrity, and ability to maintain shape and functionality when used as a wound dressing [8, 32].

https://doi.org/10.31436/iiumej.v25i2.3240

PEG, being hydrophilic, exhibited a significant affinity for water molecules. Its hydrophilic nature enhanced hydrogels' moisture retention and absorption capabilities when incorporated into their structures [2]. Additionally, glycerin complemented PEG's hydrating effect by permitting substantial water molecules to enter and exit the hydrogel matrix, thus ensuring a moist environment conducive to rapid wound healing [13]. As glycerin and PEG facilitated the penetration of water molecules into the hydrogel matrix, the interaction between the water molecules and polymer chains became stronger due to increased physical entanglements and contact between them. This resulted in complex and interconnected hydrogel structure formation, consequently elevating the gel fraction [13]. Moreover, integrating CNFs as reinforcement to strengthen the hydrogel structure led to the creation of new pores and channels within the hydrogel matrix, thereby enhancing its structural integrity and moisture absorption capabilities [33, 34].

3.4 Hydrogel Porosity

The porosities of hydrogels produced in this study are depicted in Fig. 2. Optimal porosities were achieved by the PVA-PEG/gly-CNF-LG hydrogel at 47.51 ± 0.53%, surpassing those of the PVA-PEG/gly hydrogel at 35.02 ± 1.34% and the PVA hydrogel at 27.42 ± 0.65%. Maintaining a porosity between 30% and 40% for hydrogel dressings is recommended to balance preserving a moist wound environment and effectively absorbing fluids and exudates [6, 8]. This range ensures that the dressing can effectively manage wound moisture without drying out the wound or being overwhelmed by exudates.

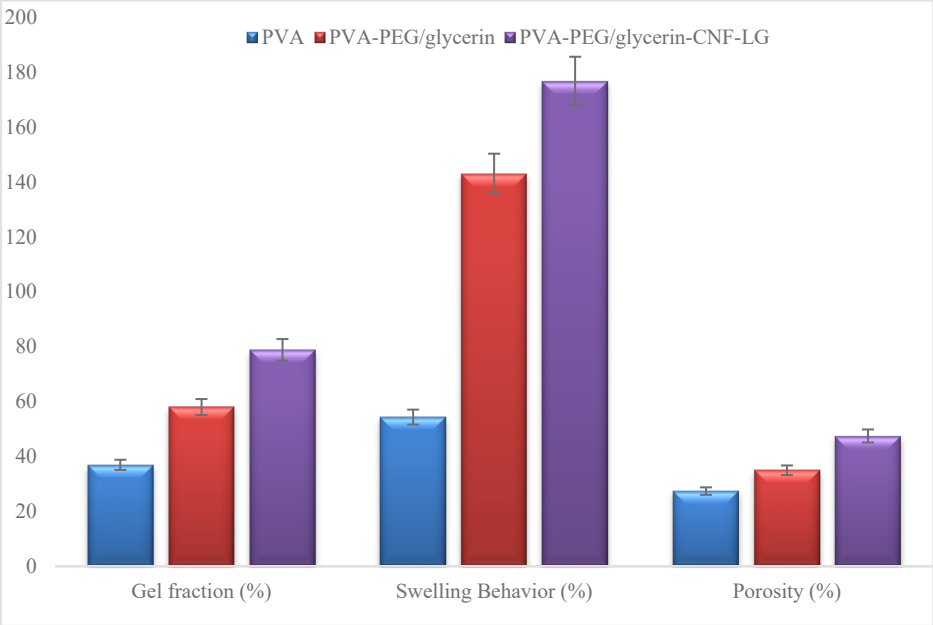


Figure 2. Physical properties, (a) gel fraction, (b) swelling behavior, and (c) porosity for PVA, PVA-PEG/gly, and PVA-PEG/gly-CNF-LG hydrogels.

However, a porosity level of 47.51% for the PVA-PEG/gly-CNF-LG hydrogel was deemed acceptable and beneficial. Higher porosity enabled greater fluid absorption, particularly advantageous for heavily exuding wounds, while also aiding in maintaining an optimal moist environment for healing. If the hydrogel maintained good structural integrity, it could still offer effective wound coverage and protection [6]. Conversely, a lack of porosity in hydrogel dressing hampered gaseous exchange and fluid absorption, potentially impeding the healing process. Additionally, excessive porosity leading to overabsorption of fluids could

result in issues such as wound drying and inadequate drainage, thereby causing discomfort to the patient [4]. Therefore, ensuring an appropriate porosity level in hydrogels was crucial to enhancing their performance as dressing materials [35].

3.5 Examination of Hydrogel Morphology

The microstructural properties of hydrogel wound dressings significantly influenced gaseous exchange, swelling capacity, and, consequently, the rate of wound healing [3]. Figure 4 presented the resulting FESEM images, revealing distinct morphologies between the PVA-PEG/gly hydrogel and the PVA-PEG/gly-CNF-LG hydrogel.

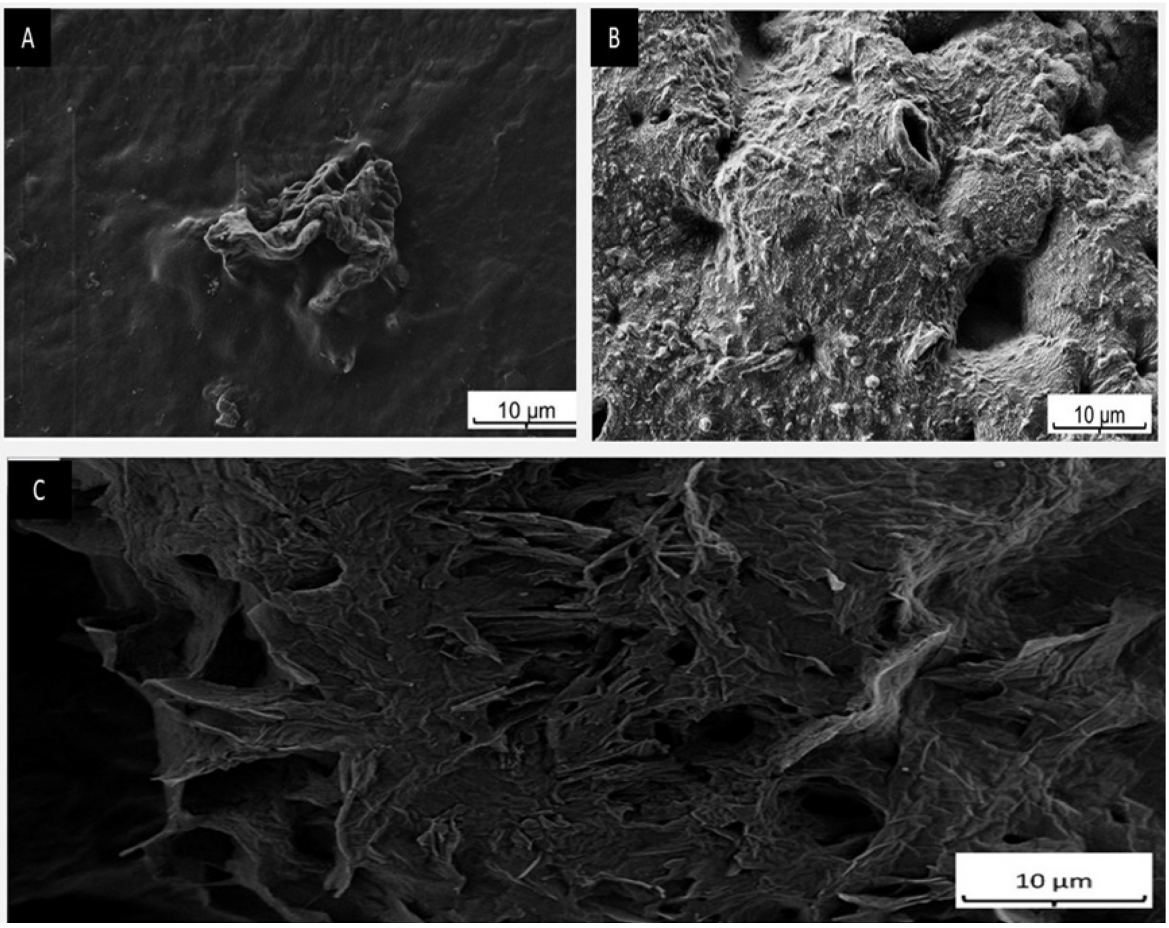


Figure 3. FESEM images (a) PVA; (b) PVA-PEG/gly; and (c) PVA-PEG/gly-CNF-LG hydrogels at 50 kV magnification.

In Fig. 3, the pure PVA hydrogel exhibited a rigid, non-porous bulk structure. In contrast, the PVA-PEG/gly-CNF-LG hydrogel displayed a highly porous structure. The fibrillar morphology of the PVA-PEG/gly hydrogel became more pronounced with incorporating CNF, consistent with previous findings [3, 7]. Surface ripples and bumps emerged on the PVA-PEG/gly hydrogel structure, attributed to changes in physical properties and surface tension resulting from including PEG and glycerin. Both compounds, being hydrophilic, enhanced the hydrogel's water-absorption capacity [8, 9]. Their introduction increased the hydrogel's moisture content, leading to expansion and swelling. However, this expansion could cause surface irregularities due to non-uniform water distribution and the formation of localized tension zones [6, 9]. These areas with varying tension resulted in ripples and bumps on the

https://doi.org/10.31436/iiumej.v25i2.3240

hydrogel's surface, visually representing the modified structure and increased moisture content [36]. Such characteristics were desirable in applications like wound dressings and drug delivery systems, emphasizing the hydrogel's ability to absorb and retain water [33, 37].

Adding CNFs, PEG, and glycerin to the PVA-PEG/gly-CNF hydrogel increased surface roughness and irregularities. CNFs created a fibrous network, enhancing mechanical strength and surface undulations. PEG and glycerin enhanced hydrophilicity, promoting greater water absorption and swelling [38]. This elevated moisture content resulted in uneven expansion and surface irregularities due to varying water distribution. These combined effects increased roughness, providing advantages in wound dressing materials by improving cell adhesion and tissue integration for enhanced wound healing. Additionally, the irregularities offered more sites for drug loading and release in drug delivery systems, making the hydrogel a valuable platform for controlled drug delivery [38, 39].

3.6 FTIR Analysis

Fig. 4 depicts Fourier transform infrared (FTIR) images of hydrogels, revealing the presence of hydroxyl (-OH) groups responsible for the hydrogel's water-binding ability. The peak strength observed in the hydrogel's FTIR findings directly correlated with its capacity to absorb water and wound exudates [7]. The hydrogen bonding interaction between the hydroxyl groups of PVA and PEG was evident in the -OH bond vibrations at around 1100-1450 cm^{-1}, with prominent peaks at 1438 cm^{-1} and 2916 cm^{-1} indicating stretching vibrations of -CH and CH2, respectively. These vibrations were primarily attributed to PVA [2, 3]. Additionally, the presence of C=O and C=C bonds in PVA, PEG, and LG was demonstrated by peaks at 1600 and 1800 cm^{-1} [6, 40].

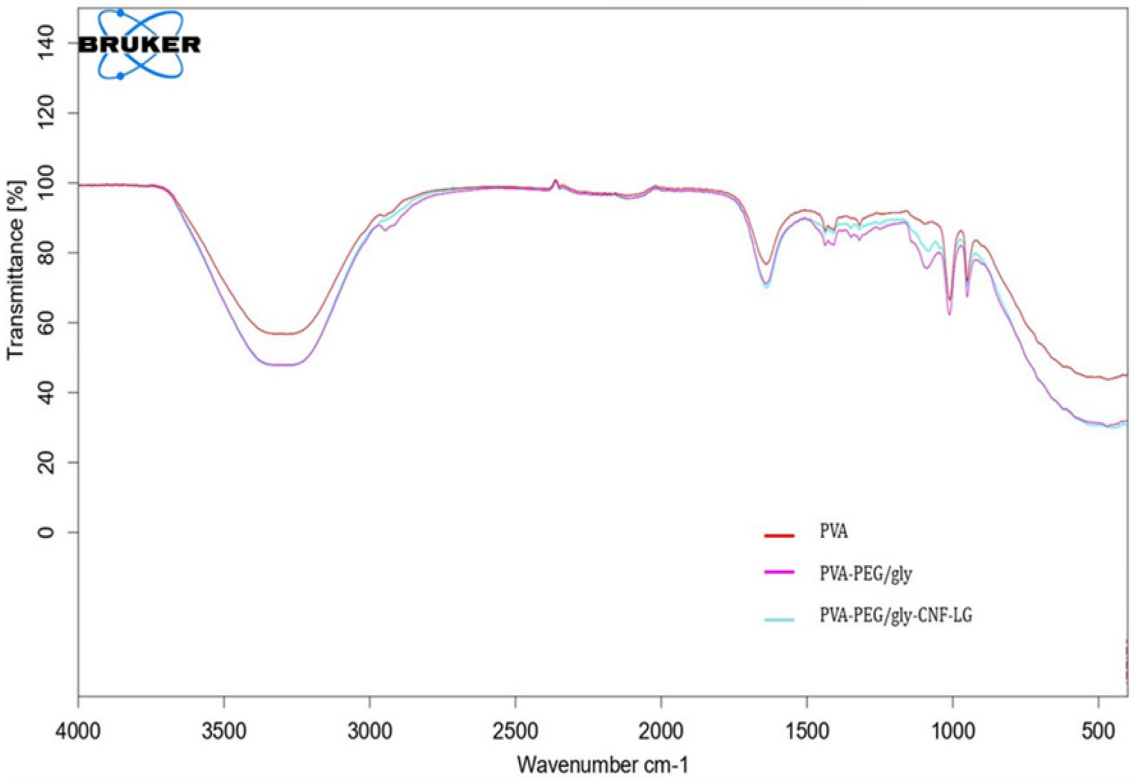


Figure 4. FTIR image of hydrogels. (a) PVA; (b) PVA-PEG/gly; and (c) PVA-PEG/gly-CNF-LG

The broadest band observed between 3200 and 3400 cm^{-1} in each spectrum indicated the stretching vibration of -OH bonds in PVA, PEG, and glycerin, a consistent feature across all hydrogel compositions [8, 13]. The presence of cellulose in the hydrogel was confirmed by a peak at 1044 cm^{-1}, signifying C-O and C-H stretching vibrations [41]. Noteworthy peaks at lower wavenumbers included the -CH2- group in the aliphatic PVA chain, resulting in C-H bonds with deformation vibrations in the 1430–1450 cm^{-1} region [42]. The stretching vibrations of C-H from PEG and glycerin caused a visible peak between 2916 and 2359 cm^{-1} in the PVA-PEG/glycerin spectra [3, 8]. Importantly, the FTIR spectra of the PVA hydrogel did not exhibit these peaks.

3.7 Zeta Potential Measurement

Zeta potential measurement served as a valuable method for evaluating the surface charge of nanoparticles within colloidal solutions [3]. In the case of the PVA-PEG/gly-CNF-LG hydrogel, a zeta potential of -15.9 mV was observed. This zeta potential implied a high degree of stability in the solution, reducing the possibility of coagulation or aggregation, which was vital for using this hydrogel as dressing material [6, 8]. A hydrogel with a negative zeta potential indicated a negatively charged surface, which repelled negatively charged particles or cells, lowering the potential risk of adhesion or fouling, a crucial property in biomedical applications [3].

In contrast, PVA and PVA-PEG/gly hydrogels recorded zeta potentials of -33.5 mV and -38.8 mV, respectively. The increasing pattern of absolute zeta potential values from PVA to PVA-PEG/gly hydrogels implied that the surface charge became greater after the inclusion of PEG and glycerin into the hydrogel structure, resulting in better stability by preventing adhesion or aggregation [6, 9]. The larger zeta potential was deemed beneficial in biomedical applications as it signified fewer interactions with proteins or cells, preventing fouling or undesirable interactions for forming stable colloids, which were essential to sustain consistent dispersion and functionality [12].

3.8 Antibacterial Testing

The disc diffusion method was applied to examine the antibacterial efficiency of the produced PVA-PEG/gly-CNF-LG hydrogel against several bacterial strains. The results showed that the PVA-PEG/gly-CNF-LG hydrogel was effective against gram-positive bacteria, as indicated by inhibition zones measuring 3.5 ± 0.2 cm for *B. subtilis* and 2.5 ± 0.3 cm for *S. aureus*. Citral, a key lemongrass (LG) component, is recognized for its efficacy against gram-positive bacteria [40]. However, no inhibition zone was observed against *E. coli* and *S. enteritidis*, indicating the absence of antibacterial effects of the PVA-PEG/gly-CNF-LG hydrogel against gram-negative bacteria. Detailed findings on the hydrogel's antibacterial activity were summarized in Table 4, alongside the control groups for comparison.

Table 4. Antibacterial effectiveness of hydrogels.

Compound	Inhibition zone diameter (cm)			
	E. coli	*B. subtilis*	*S. aureus*	*S. enteritidis*
PVA-PEG/gly-CNF-LG	NA	3.5 ± 0.2	2.5 ± 0.3	NA
Tetracycline hydrochloride (+ve control)	2.7± 0.4	3.2 ± 0.1	2.3 ± 0.1	2.8± 0.3
Dimethyl sulfoxide (- ve control)	NA	NA	NA	NA

The varying antibacterial outcomes observed in the PVA-PEG/gly-CNF-LG hydrogel against distinct bacterial strains—where it showed no efficacy against *E. coli* and *S. enteritidis* but demonstrated effectiveness against *B. subtilis* and *S. aureus*—may have stemmed from various factors. Initially, the inclusion of LG in the hydrogel formulation likely played a significant role. LG contained bioactive compounds like citral, which are known for their antimicrobial properties against specific bacteria. *B. subtilis* and *S. aureus* are gram-positive bacteria with a single-cell membrane, making them generally more susceptible to antimicrobial agents found in LG [43]. On the other hand, *E. coli* and *S. enteritidis* are gram-negative bacteria with a double cell membrane that acts as a barrier, rendering them less susceptible to the penetration of certain antimicrobial agents [18, 43].

3.9 Mechanical Characteristics of Hydrogel

Improving the tensile strength of hydrogels was crucial for ensuring their efficacy as wound dressings. Tensile strength assesses a material's capacity to withstand stretching or pulling forces without fracturing [44]. The tensile stress-strain graph of hydrogels, as depicted in Fig. 5 and considering comparable thickness, was utilized to evaluate their tensile characteristics.

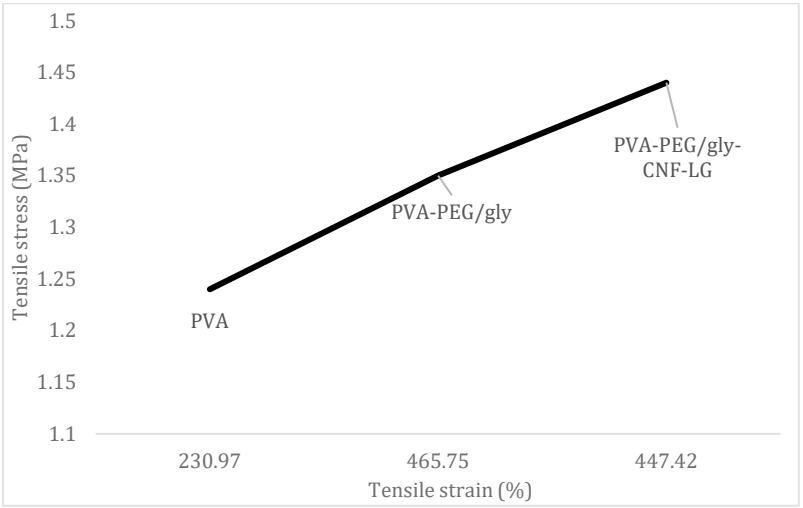


Figure 5. Tensile stress against strain graph for produced hydrogel.

The PVA hydrogel exhibited a tensile strength of 1.24 MPa. Meanwhile, the tensile strength increased to 1.35 MPa after incorporating 4% (w/v) glycerin and 6% (w/v) of PEG into the PVA hydrogel to yield the PVA-PEG/gly hydrogel. This enhancement was attributed to the increased number of crosslinking sites within the three-dimensional polymer network as a result of glycerin and PEG's hydrophilic nature, which created hydrogen bonds to fortify the hydrogel structure, resulting in the increased tensile properties of the hydrogels [3, 8]. Subsequently, the tensile strength further increased after incorporating 3.5% (w/v) CNF and 3% (v/v) LG, resulting in a tensile strength of 1.44 MPa for the PVA-PEG/gly-CNF-LG hydrogel.

4. CONCLUSION

Incorporating PEG, glycerin, CNF, and LG into the polymeric matrix of PVA was conducted via the freeze-thaw method to produce the PVA-PEG/gly-CNF-LG hydrogel for wound dressing application. The highest MRC of 37.69 ± 0.54% was attained for the PVA-PEG/gly-CNF-LG hydrogel at the optimum hydrogel composition with concentrations of 3.5%

(w/v) CNFs and 3% (v/v) LG. The optimized PVA-PEG/glycerin-CNF-LG exhibited impressive characteristics, including a 176.89 ± 1.56% swelling capacity, 78.89 ± 0.42% gel fraction, and 47.51 ± 0.53% porosity. FESEM analysis showed that the produced hydrogel was highly porous, with empty void spaces within its structure. The FTIR spectra showcased the effective cross-linking of the PVA-PEG/glycerin polymeric network with CNF and LG. Antibacterial investigations revealed the resilience of the PVA-PEG/gly-CNF-LG hydrogel against *S. aureus* and *B. subtilis*. Additionally, it demonstrated outstanding tensile characteristics, reaching 1.44 MPa. These findings collectively highlight the promising potential of the optimized PVA-PEG/glycerin-CNF-LG hydrogel as a wound dressing material.

ACKNOWLEDGEMENT

The author would like to thank the Biochemical Laboratory Department of Chemical Engineering & Sustainability, International Islamic University Malaysia, for the laboratory facilities.

REFERENCES

[1] Lin SP, Lo KY, Tseng TN, Liu JM, Shih TY, Cheng KC (2019) Evaluation of PVA/dextran/chitosan hydrogel for wound dressing. Cellular Polymers 38:15–30

[2] Ahmed AS, Mandal UK, Taher M, Susanti D, Jaffri JM (2018) PVA-PEG physically cross-linked hydrogel film as a wound dressing: experimental design and optimization. Pharm Dev Technol 23:751–760

[3] Li Y, Zhu C, Fan D, Fu R, Ma P, Duan Z, Li X, Lei H, Chi L (2019) A Bi-Layer PVA/CMC/PEG Hydrogel with Gradually Changing Pore Sizes for Wound Dressing. Macromol Biosci. https://doi.org/10.1002/mabi.201800424

[4] Baghaie S, Khorasani MT, Zarrabi A, Moshtaghian J (2017) Wound healing properties of PVA/starch/chitosan hydrogel membranes with nano Zinc oxide as antibacterial wound dressing material. J Biomater Sci Polym Ed 28:2220–2241

[5] Butylina S, Geng S, Laatikainen K, Oksman K (2020) Cellulose Nanocomposite Hydrogels: From Formulation to Material Properties. Front Chem. https://doi.org/10.3389/fchem.2020.00655

[6] Altaf F, Niazi MBK, Jahan Z, Ahmad T, Akram MA, safdar A, Butt MS, Noor T, Sher F (2021) Synthesis and Characterization of PVA/Starch Hydrogel Membranes Incorporating Essential Oils Aimed to be Used in Wound Dressing Applications. J Polym Environ 29:156–174

[7] Rahmi D, Paramadina S, Anjelika M, Widjajanti R (2020) Optimized swelling properties of hydrogels based on poly(vinyl alcohol)-carrageenan. AIP Conf Proc. https://doi.org/10.1063/5.0001098

[8] Cui L, Tong W, Zhou H, Yan C, Chen J, Xiong D (2021) PVA-BA/PEG hydrogel with bilayer structure for biomimetic articular cartilage and investigation of its biotribological and mechanical properties. J Mater Sci 56:3935–3946

[9] Chen K, Liu J, Yang X, Zhang D (2017) Preparation, optimization and property of PVA-HA/PAA composite hydrogel. Materials Science and Engineering C 78:520–529

[10] Carating CAM, Rosales RNM, Bongao HC, Magdaluyo ER (2019) Polyvinyl alcohol hydrogel reinforcement of cellulose and silica nanoparticles for wound healing application. In: Key Eng Mater. Trans Tech Publications Ltd, pp 23–28

[11] Butylina S, Geng S, Oksman K (2016) Properties of as-prepared and freeze-dried hydrogels made from poly(vinyl alcohol) and cellulose nanocrystals using freeze-thaw technique. Eur Polym J 81:386–396

[12] Abu Ghalia M, Dahman Y (2015) Radiation crosslinking polymerization of poly (vinyl alcohol) and poly (ethylene glycol) with controlled drug release. Journal of Polymer Research. https://doi.org/10.1007/s10965-015-0861-9

[13] Bialik-Was K, Pluta K, Malina D, Barczewski M, Malarz K, Mrozek-Wilczkiewicz A (2021) The effect of glycerin content in sodium alginate/poly(Vinyl alcohol)-based hydrogels for

wound dressing application. Int J Mol Sci. https://doi.org/10.3390/ijms222112022
[14] Ali MASS, Jimat DN, Nawawi WMFW, Sulaiman S (2022) Antibacterial, Mechanical and Thermal Properties of PVA/Starch Composite Film Reinforced with Cellulose Nanofiber of Sugarcane Bagasse. Arab J Sci Eng 47:5747–5754
[15] Gea S, Indra, Muis Y, Panindia N, Hutapea YA (2019) Preparation of polyvinyl alcohol/cellulose nano fiber nanocomposite isolated from empty oil palm fruit bunches. IOP Conf Ser Mater Sci Eng. https://doi.org/10.1088/1757-899X/553/1/012041
[16] Liu R, Dai L, Si C, Zeng Z (2018) Antibacterial and hemostatic hydrogel via nanocomposite from cellulose nanofibers. Carbohydr Polym 195:63–70
[17] Naik MI, Fomda BA, Jaykumar E, Bhat JA (2010) Antibacterial activity of lemongrass (Cymbopogon citratus) oil against some selected pathogenic bacterias. Asian Pac J Trop Med 3:535–538
[18] Fatunmibi OO, Njoku IS, Asekun OT, Ogah JO (2023) Chemical composition, antioxidant and antimicrobial Activity of the essential oil from the leaves of Cymbopogon citratus.
[19] Shah G, Shri R, Panchal V, Sharma N, Singh B, Mann AS (2011) Scientific basis for the therapeutic use of Cymbopogon citratus, stapf (Lemon grass). J Adv Pharm Technol Res 2:3–8
[20] Entezam M, Daneshian H, Nasirizadeh N, Khonakdar HA, Jafari SH (2017) Hybrid Hydrogels Based on Poly(vinyl alcohol) (PVA)/Agar/Poly(ethylene glycol) (PEG) Prepared by High Energy Electron Beam Irradiation: Investigation of Physico-Mechanical and Rheological Properties. Macromol Mater Eng. https://doi.org/10.1002/mame.201600397
[21] Figueroa-Pizano MD, Vélaz I, Martínez-Barbosa ME (2019) A freeze-thawing method to prepare chitosan-poly(Vinyl alcohol) hydrogels without crosslinking agents and diflunisal release studies. Journal of Visualized Experiments. https://doi.org/10.3791/59636
[22] Asem M, Noraini Jimat D, Huda Syazwani Jafri N, Mohd Fazli Wan Nawawi W, Fadhillah Mohamed Azmin N, Firdaus Abd Wahab M (2023) Entangled cellulose nanofibers produced from sugarcane bagasse via alkaline treatment, mild acid hydrolysis assisted with ultrasonication. Journal of King Saud University - Engineering Sciences 35:24–31
[23] Douard L, Bras J, Encinas T, Belgacem MN (2021) Natural acidic deep eutectic solvent to obtain cellulose nanocrystals using the design of experience approach. Carbohydr Polym. https://doi.org/10.1016/j.carbpol.2020.117136
[24] Rezvanian M, Ahmad N, Mohd Amin MCI, Ng SF (2017) Optimization, characterization, and in vitro assessment of alginate-pectin ionic cross-linked hydrogel film for wound dressing applications. Int J Biol Macromol 97:131–140
[25] Thakur M, Sharma A, Ahlawat V, Bhattacharya M, Goswami S (2020) Process optimization for the production of cellulose nanocrystals from rice straw derived α-cellulose. Mater Sci Energy Technol 3:328–334
[26] Bacha EG (2022) Response Surface Methodology Modeling, Experimental Validation, and Optimization of Acid Hydrolysis Process Parameters for Nanocellulose Extraction. S Afr J Chem Eng 40:176–185
[27] Alven S, Nqoro X, Aderibigbe BA (2020) Polymer-based materials loaded with curcumin for wound healing applications. Polymers (Basel) 12:1–25
[28] Shitole AA, Raut PW, Khandwekar A, Sharma N, Baruah M (2019) Design and engineering of polyvinyl alcohol based biomimetic hydrogels for wound healing and repair. Journal of Polymer Research. https://doi.org/10.1007/s10965-019-1874-6
[29] Singh BR, Singh V, Singh RK (2016) Antimicrobial activity of lemongrass (Cymbopogon citratus) oil against microbes of environmental, clinical and food originEpidemiology-of-emergence-of-antimicrobial-drugs-resistance-in-bacteria-of-veterinary-clinical-importance View project Epidemiology of emergence of antimicrobial drugs resistance in bacteria of veterinary clinical importance View project.
[30] Andreu V, Mendoza G, Arruebo M, Irusta S (2015) Smart dressings based on nanostructured fibers containing natural origin antimicrobial, anti-inflammatory, and regenerative compounds. Materials 8:5154–5193
[31] Bashir S, Hina M, Iqbal J, Rajpar AH, Mujtaba MA, Alghamdi NA, Wageh S, Ramesh K, Ramesh S (2020) Fundamental concepts of hydrogels: Synthesis, properties, and their

applications. Polymers (Basel) 12:1–60
[32] Naeimi M, Tajedin R, Farahmandfar F, Naeimi M, Monajjemi M (2020) Preparation and characterization of vancomycin-loaded chitosan/PVA/PEG hydrogels for wound dressing. Mater Res Express. https://doi.org/10.1088/2053-1591/abb154
[33] Shen R, Xue S, Xu Y, Liu Q, Feng Z, Ren H, Zhai H, Kong F (2020) Research progress and development demand of nanocellulose reinforced polymer composites. Polymers (Basel). https://doi.org/10.3390/POLYM12092113
[34] Nascimento DM, Nunes YL, Figueirêdo MCB, De Azeredo HMC, Aouada FA, Feitosa JPA, Rosa MF, Dufresne A (2018) Nanocellulose nanocomposite hydrogels: Technological and environmental issues. Green Chemistry 20:2428–2448
[35] Shahzadi MP (2017) Lemon Grass (Cymbopogon citratus). Grasses - Benefits, Diversities and Functional Roles. https://doi.org/10.5772/intechopen.69518
[36] Chamkouri H (2021) A Review of Hydrogels, Their Properties and Applications in Medicine. Am J Biomed Sci Res 11:485–493
[37] Kumar Soni R, M Kewatkar S, Jain V, Singh Saluja M (2023) A Review of Hydrogels, with its Properties and Applications in Medicine. Journal of Biomedical and Pharmaceutical Research 12:20–33
[38] Rolim WR, Pieretti JC, Renó DLS, Lima BA, Nascimento MHM, Ambrosio FN, Lombello CB, Brocchi M, De Souza ACS, Seabra AB (2019) Antimicrobial Activity and Cytotoxicity to Tumor Cells of Nitric Oxide Donor and Silver Nanoparticles Containing PVA/PEG Films for Topical Applications. ACS Appl Mater Interfaces. https://doi.org/10.1021/acsami.8b19021
[39] Rafieian S, Mirzadeh H, Mahdavi H, Masoumi ME (2019) A review on nanocomposite hydrogels and their biomedical applications. IEEE Journal of Selected Topics in Quantum Electronics 26:154–174
[40] Saleem M, Saeed MT (2020) Potential application of waste fruit peels (orange, yellow lemon and banana) as wide range natural antimicrobial agent. J King Saud Univ Sci 32:805–810
[41] Yang W, Xu F, Ma X, et al (2021) Highly-toughened PVA/nanocellulose hydrogels with anti-oxidative and antibacterial properties triggered by lignin-Ag nanoparticles. Materials Science and Engineering C. https://doi.org/10.1016/j.msec.2021.112385
[42] de Lima GG, Ferreira BD, Matos M, Pereira BL, Nugent MJD, Hansel FA, Magalhães WLE (2020) Effect of cellulose size-concentration on the structure of polyvinyl alcohol hydrogels. Carbohydr Polym. https://doi.org/10.1016/j.carbpol.2020.116612
[43] Shetty SB, Mahin-Syed-Ismail P, Varghese S, Thomas-George B, Kandathil-Thajuraj P, Baby D, Haleem S, Sreedhar S, Devang-Divakar D (2016) Antimicrobial effects of Citrus sinensis peel extracts against dental caries bacteria: An in vitro study. J Clin Exp Dent 8:e71–e77
[44] Zhang J, Liu T, Liu Z, Wang Q (2019) Facile fabrication of tough photocrosslinked polyvinyl alcohol hydrogels with cellulose nanofibrils reinforcement. Polymer (Guildf) 173:103–109

https://doi.org/10.31436/iiumej.v25i2.3030

MONITORING CHANGE DETECTION OF VEGETATION VULNERABILITY USING HOTSPOTS ANALYSIS

Basheer S. Jasim[1,2], Oday Z. Jasim[2], Amjed N. AL-Hameedawi[2]*

[1]*Technical Institute of Babylon, Al-Furat Al-Awsat Technical University, Iraq*
[2]*Civil Engineering Department, University of Technology - Iraq, Baghdad, Iraq*

**Corresponding author: basheer.jasim@atu.edu.iq*

(Received: 21 September 2023; Accepted: 30 January 2024; Published online: 15 July 2024)

ABSTRACT: Because of the ever-shifting nature of the weather conditions, which are made even more difficult by the dynamic relationship between the environment and the vegetation, one of the most important aspects is the vegetation. Landsat satellite imagery, TM sensor for 2002 and 2012, and OLI-TIRS sensor for 2022 were used for vegetation vulnerability. The Normalized Difference Vegetation Index (NDVI) method and hotspots analysis method were used for image classification, and the land cover map was obtained in three different years. The results of the analysis have shown that during 20 years, the extremely vulnerable zone has increased by 0.53%, the very vulnerable zone by 12.04%, and the moderately vulnerable zone has increased by 2.27% in terms of total area, also decreasing the non-significant zone by 5.74%, and the moderately safe zone decreased by 5.42%. The very safe zone decreased during this period by 2.94%. The extreme safe zone decreased by 0.73% in terms of total. Based on the assessment and validation of zone classification data, the overall accuracy value shows that the vegetation vulnerability accuracy for 2022 was equal to 90%. Also, the kappa coefficient for the classification vegetation vulnerability map was equal to 0.88. The research using Landsat data concluded that there had been a reduction in the amount of land covered by thick vegetation, which resulted in widespread drought conditions in some portions of the study region (Babylon Governorate). This research has shown that using satellite images and GIS spatial analysis is very effective in identifying and evaluating the trend of vegetation vulnerability in the Babylon Governorate. These data and techniques can be used for various analytical purposes.

ABSTRAK: Faktor perubahan cuaca yang mendadak, di mana hubungan dinamik antara alam sekitar dan tanaman menjadi lebih sukar, merupakan satu aspek penting bagi tumbuh-tumbuhan. Imej satelit Landsat, penderia TM 2002 dan 2012, dan penderia OLI-TIRS 2022 digunakan untuk tumbuh-tumbuhan yang terdedah. Kaedah Indeks Perubahan Ternormal Tumbuhan (NDVI) dan kaedah analisis kawasan khas digunakan bagi tujuan pengelasan imej, dan peta kawasan tanah berkaitan diperoleh dalam tiga tahun berbeza. Dapatan analisis menunjukkan selama 20 tahun, zon paling teruk terjejas telah bertambah sebanyak 0.53%, zon terjejas sebanyak 12.04%, zon sederhana terjejas bertambah kepada 2.27% berdasarkan total kawasan, juga pengurangan zon tidak penting 5.74%, zon sederhana selamat berkurang sebanyak 5.42%. Zon selamat telah berkurang selama tempoh ini sebanyak 2.94%. Zon paling selamat berkurang sebanyak 0.73% berdasarkan jumlah keseluruhan. Nilai ketepatan keseluruhan menunjukkan ketepatan tumbuh-tumbuhan terdedah pada 2022 bersamaan 90%, iaitu berdasarkan data klasifikasi zon pada ujian dan validasi. Juga, pekali kappa bagi klasifikasi peta tumbuh-tumbuhan terdedah bersamaan 0.88. Kesimpulan terhadap kajian menggunakan data Landsat ini adalah terdapat pengurangan pada bilangan tanah yang ditutupi oleh tumbuh-tumbuhan tebal, di mana menyebabkan keadaan kemarau yang berleluasa di

https://doi.org/10.31436/iiumej.v25i2.3030

sebahagian kawasan yang dikaji (Babylon Governorate). Kajian ini menunjukkan dengan menggunakan imej satelit dan analisis ruang GIS, ianya sangat berkesan dalam mengenal pasti dan menganalisa perkembangan tumbuh-tumbuhan yang terdedah di Babylon Governorate. Data dan teknik ini boleh digunakan untuk pelbagai tujuan analisis.

KEYWORDS: *Vegetation vulnerability, NDVI, Vegetation Cover, Change detection, Hotspot analysis.*

1. INTRODUCTION

The change detection process requires using many temporal datasets to do a quantitative analysis of the phenomenon's impacts on the period [1][2]. Monitoring local, regional, and global resources and environments requires information on changes occurring on the world's surface [3][4]. The term "land cover" describes the various things that may be found on top of the Earth, both naturally and artificially [5]. The importance of land use lies in its relation with people's lives because it offers services directly in contact with people's needs [6][7]. The average temperature of the Earth's surface has increased, the average amount of precipitation has dropped, and the amount of land decertified has increased due to climate change and rising levels of human activity [8][9]. The remote sensing technique is crucial for creating maps showing land use and land cover from a procedure known as image classification [10]. The vegetation dynamics are essential in revealing how ecological systems respond to climate change [11]. Changes in vegetation may be caused by a wide variety of causes, including temperature and precipitation, ecological movement, grazing, and many more [12]. The environmental phenomenon known as "vegetation vulnerability" refers to how different driving factors work at varying degrees to cause the loss of vegetation. Natural fires, pests, illnesses, natural catastrophes, and even human conflict are all potential causes of vegetation loss. Agriculture is considered one of the most vulnerable sectors to drought compared to other sectors [13].

Several nations have extensively used remote sensing technology for effective environmental monitoring. In addition, many studies were conducted in various locations, including the United States, Europe, and others [14]. Multiple independent remote sensing indicators have been developed to evaluate the current status of the environment [15]. It is feasible to use conventional techniques for measuring the degree of appropriateness and susceptibility in any area, but doing so is very challenging and time-intensive [16]. In recent decades, geographic information systems (GIS) and satellite remote sensing (RS) technologies have gained substantial interest in suitability and vulnerability assessment. It has been proposed that these techniques are extremely effective [17].

Abdullah and Barua's. [18] A hybrid model was utilized to evaluate the model's accuracy in creating the hot spot map for 2021. The exact weight values have been used to forecast the hotspot map for 2026 and 2031, and the kappa statistics have been determined to be greater than 0.86 with 88.75 percent overall accuracy. The whole Nijhum Dwip region is split into seven zones using Gi* analysis, with the 90, 95, and 99% cold confidence levels being designated as moderately vulnerable zone (MVZ), very vulnerable zone (VVZ), and extremely susceptible zone (EVZ), respectively. However, the moderately safe zone (MSZ), very safe zone (VSZ), and extreme safe zone (ESZ), respectively, have been approved at the 99, 95, and 90 percent heat confidence levels. Additionally, it was considered that the remaining proportion of earth is an insignificant area (NSZ). The latter three kinds reflect safe places with the least to most vegetation clusters, whereas the first three indicate vulnerable areas with little or no vegetation. The NSZ is a barrier between the safe and vulnerable areas, which may change into other categories based on how well this boundary is maintained.

https://doi.org/10.31436/iiumej.v25i2.3030

The hotspots of deforestation and forest fires in Andhra Pradesh, India, were identified by Reddy et al. [19] using historical patterns of deforestation (from 1975 to 1985 to 1995 to 2005 to 2014) and fire histories (2009-2014). The priority regions were selected as degradation and deforestation hotspots using a combination of deforestation danger metrics. In 2014, there were 25985 km^2 (16.2 percent of the entire geographical area) of forests in Andhra Pradesh. Between 1975 and 2014, a net deforested area of 2390 km^2 (8.4 percent) was calculated. Based on the prioritizing findings, deforestation and degradation hotspot-1 were identified in around 10.5 percent of existing forests due to severe ecological collapse.

According to research by Singh et al. [20], deforestation hotspots have been modeled as the forest regions with the highest densities of negative changes. The findings of the fragmentation analysis showed that core forests decreased during the research period. According to an examination of forest changes, between 1975 and 2015, the amount of land covered by forests decreased from 23749 km^2 (or 30.3 % of the state's geographical region) to 21145 km^2 (or 27 % of the state's geographical region). Assam's anticipated total forest area in 2030 is 20189 km^2 (25.7 %).

According to previous research findings, rainfall and temperature are the two most essential climate aspects contributing to vegetation changes. Increased human activities may further complicate the processes that influence vegetation changes due to climate change. NDVI is used to detect vegetation. A geographic statistical analysis converts each index pixel into points, indicating clustering based on a value in NDVI. The IDW method converts points into a surface that can be measured, and its areas are calculated. The problem is that no standard determines the division of NDVI, so the researcher chose a method that helps the division geographically and statistically. This study aims to explore vegetation vulnerability between 2002 and 2022 in the Babylon Governorate. The issues that need to be answered by the study are as follows: where are the sites of vegetation vulnerability change concentrated, how can places be identified in which the vulnerability of vegetation is most widespread, and where are the vegetation vulnerability and degradation hotspots?

2. THEORETICAL BACKGROUND

2.1. Image Preprocessing

The processing of images using satellite imagery is a significant difficulty. One of the most important duties in this field is recognizing individual objects inside a satellite image. Image restoration and rectification methods correct radiometric and geometric data distortions specific to certain sensors and platforms. These distortions might occur because of the sensor and platform interaction. Variations in scene lighting and viewing geometry, as well as changes in the conditions of the atmosphere, sensor noise, and responsiveness, may need radiometric adjustments [21]. These will differ based on the particular sensor and platform used to gather the data and the conditions present during the data acquisition [22].

2.1.1. Atmospheric Correction

Radiance data obtained by the sensor from the surface of the Earth are evaluated and corrected for distortions produced by the atmosphere and topography [23]. This correction aims to obtain surface reflectance values and restore the Earth's surface's physical characteristics by removing the atmosphere's effects from satellite images [24].

https://doi.org/10.31436/iiumej.v25i2.3030

2.1.2 Radiometric Calibration

This procedure aims to convert the observed brightness into top-of-atmosphere reflectance. When constructing a multispectral or multitemporal mosaic, it is vital to follow this technique because it eliminates variations between images caused by variances in sensors, solar zenith angle, and distance between the Earth and the sun [25]. The following is a description of the process that must be followed to convert DN data to reflectance values. In the first step, the DN values are changed into radiance. Eq. (1) shows how the radiance value is converted to reflectance [26] where P is the TOA reflectance, $R\lambda$ is the spectral radiance, D is the Earth-sun distance in astronomical units from metadata, $E\lambda$ is the mean solar exoatmospheric irradiance (https://landsat.usgs.gov) and θs is the solar zenith angle from metadata.

$$P = [\frac{(\pi * R\lambda * d^2)}{(E\lambda * cos\theta s)}] \quad (1)$$

Temperature data is collected by satellite sensors like Landsat and recorded as a Digital number (DN) between 0 and 255. DNs from Landsat thermal bands are converted to spectral radiance using the image information and the following method to calculate brightness temperature [27] as shown in Eq. (2) where $R\lambda$ is the spectral radiance [W/(m^2 ×sr×μm)], $Q.cal$ is the value of a pixel in a digital number, $Lmax$ is the spectral radiance (it is scaled to $Q.calmax$), $Lmin$ is the spectral radiance (it is scaled to $Q.calmin$), $Q.Calmax$ is the maximum calibrated value for a pixel and $Q.Calmin$ is the minimum calibrated value for a pixel.

$$R\lambda = \left[\frac{Lmax.\lambda - Lmin.\lambda}{Q.calmax. - Q.calmin.}\right] * [Q.cal - Q.calmin.] + Lmin.\lambda \quad (2)$$

2.2. Normalized Difference Vegetation Index

Computation of Normalized Difference Vegetation Index (NDVI) values for a Landsat image by the following formula [28]:

$$NDVI = \frac{(\rho.NIR - \rho.RED)}{(\rho.NIR + \rho.RED)} \quad (3)$$

where $\rho.RED$ and $\rho.NIR$ are the red and near-infrared reflectance, respectively. NDVI values are within the range −1 to 1, where negative values present loss of vegetation or contaminated vegetation, whereas positive values explain healthy and dense vegetation [29].

2.3. Hotspots Analysis

A "hotspot" is a region with a greater concentration of occurrences than expected from a random distribution of those events. This analytic tool identifies seven geographical patterns and the percentage of confidence levels for either warm or cold conditions (Gi Bin). The G_i^* analysis is a technique developed from studying point distributions or spatial groupings of points in a location [30]. The following equation is used for the calculation of Getis-Ord G_i^* according to the Eq. (4) (5) and (6) where, x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between features i and j, n equals the total number of features, and the G_i^* is a z-score.

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\ \sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}} \quad (4)$$

https://doi.org/10.31436/iiumej.v25i2.3030

$$s = \sqrt{\frac{\sum_{j}^{n=1^{x^2}} j}{n}} - (\bar{X})^2 \quad (5)$$

$$\bar{X} = \frac{\sum_{j}^{n=1^{x}} j}{n} \quad (6)$$

2.4. Inverse Distance Weighted

The Inverse Distance Weighted (IDW) approach is one of the most important geostatistical and mathematical interpolation methods [31]. The formula that is used to determine the value of the unknown variable $Z(S_0)$ at the position So according to Eq. (7) and (8) where n is the number of the monitoring station, and $Z(S_i)$ is the value at each sampled place. S_i and W_i is the weight of S_i, which may be described as follows:

$$Z(S_0) = \sum_{i=1}^{n} W_i Z(S_i), \quad (7)$$

$$W_i = \frac{\frac{1}{d_i^k}}{(\sum_{i=1}^{n} \frac{1}{d_i^k})}, i = 1,2,\dots,n \quad (8)$$

In this equation, di represents the horizontal distance that separates the interpolation points from those observed, and k is the power of the space.

2.5. Accuracy Assessment

Validation of the accuracy of the obtained categorized images using a kappa index and an error matrix [32]. The formula used to find kappa coefficients is:

$$\text{Kappa coefficient} = \frac{\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k} n_{ii}(G_i C_i)}{n^2 - \sum_{i=1}^{k} n_{ii}\,(G_i C_i)} \quad (9)$$

where i is the number of the class, n is the total number of pixels that have been classified and will be compared with the actual data, n_{ii} represents the number of pixels that are part of the introductory data class i, that were classified as belonging to class i, C_i is the total number of pixels that correspond to class i that have been organized. G_i represents the total number of actual data pixels that belong to class i. Table 1 shows the accuracy rank for each kappa coefficient.

Table 1. Classification accuracy is based on the kappa value [33].

Kappa coefficient	Accuracy level
0.41–0.60	Moderate
0.61–0.80	High
≥0.80	Very high

2.6. Available Data and Methodology

Satellite data was used for this investigation. Landsat-5 TM images (for the years 2002 and 2012) and Landsat-8 OLI/TIRS images for the year 2022 were collected from the USGS (https://earth explorer.USGS.gov/), details of which can be found in Table 2. These images and data were then used for the necessary geostatistical and geospatial analysis. This study used ENVI 5.3 and ArcGIS 10.8 software to process and analyze the datasets. Fig. 1 illustrates the methodology flowchart.

https://doi.org/10.31436/iiumej.v25i2.3030

Table 2. Satellite information was used for this study.

No.	Satellite	Sensor	Path/Row	Date of Acquisition	Resolution	Product Type	Cloud Cover
1	Landsat-5	TM	168/38	22/10/2002	30 m	Landsat Collection 1 Level-1	0 %
2	Landsat-5	TM	168/37	22/10/2002			
3	Landsat-5	TM	168/38	18/10/2012			
4	Landsat-5	TM	168/37	18/10/2012			
5	Landsat-8	OLI/TIRS	168/38	13/10/2022			
6	Landsat-8	OLI/TIRS	168/37	13/10/2022			

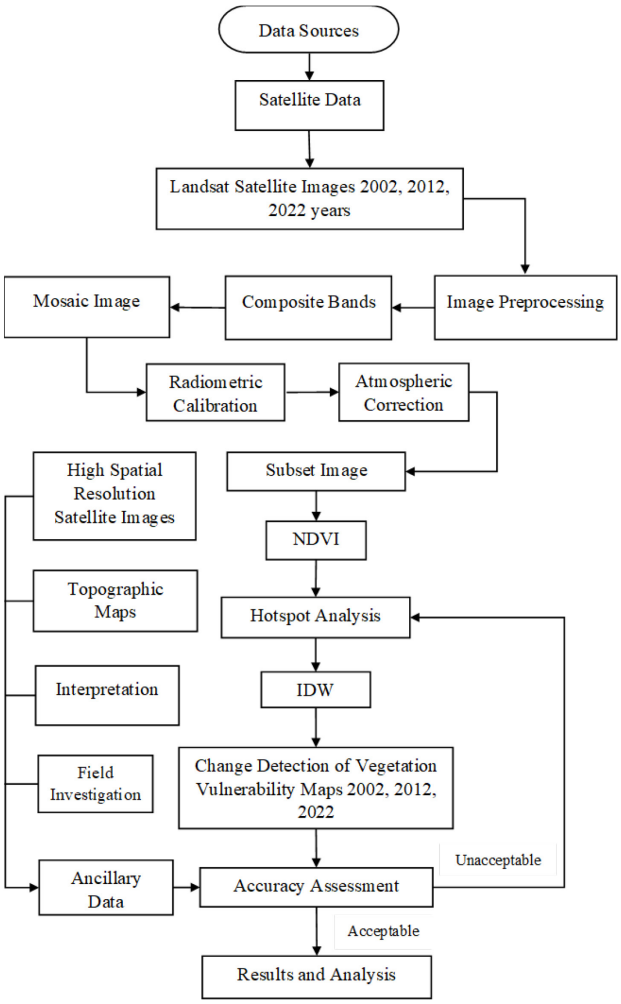
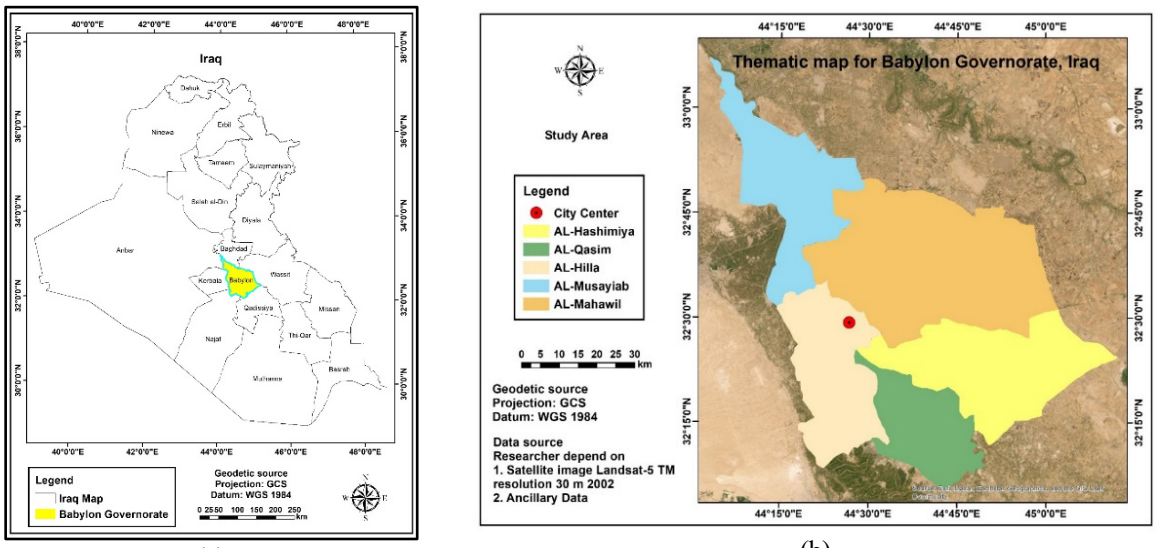


Figure 1. Flowchart for the general work.

3. STUDY AREA

The Babylon Governorate may be considered to be in the geographic middle of Iraq, located around 100 kilometers to the south of Baghdad, the capital of Iraq, between the longitudes of 44°2'43" East and 45°12'11" East, as well as the latitudes of 32°5'41" North and 33°7'36" North. Administratively, five major cities make up the Governorate as a whole [34]. These significant cities include Al-Hillah, Al-Qasim, Al-Hashimiyah, Al-Mahawil, and Al-Musayiab. According to calculations made using the program ArcGIS 10.8, the research region covers an area of about 5338 km^2. Fig. 2(a) and Fig. 2(b) show the study area map.

https://doi.org/10.31436/iiumej.v25i2.3030



(a) (b)

Figure 2. (a) The location of the study area in Iraq. (b) Geographical location map of Babylon Governorate.

4. RESULTS AND DISCUSSION

4.1 Change Detection

The production of digital maps is the first stage in building a database for all the measurements and information of a location [35][36]. To better understand the links and interactions between people and the natural environment, it is vital to collect information concerning changes in land cover. In this regard, remote sensing data has been one of the most important sources for studying the spatial and temporal changes in land cover in different regions. For processing and analysis, remote sensing multi-temporal datasets provide the possibility of mapping and identifying landscape changes and provide an effective step for sustainable local planning and management. In particular, by integrating RS and GIS techniques, it is possible to analyze and classify the changing patterns of land cover over a long period and, as a result, better understand the changes in the target range. This study adopts vegetation vulnerability classification in Table 3.

Table 3. Classification vegetation vulnerability classes following values proposed by [18].

No.	Hotspots	Zone
1	Cold spot 99% Confidence	Extremely vulnerable zone
2	Cold spot 95% Confidence	Very vulnerable zone
3	Cold spot 90% Confidence	Moderately vulnerable zone
4	Not Significant	Non-significant zone
5	Hot spot 90% Confidence	Moderately safe zone
6	Hot spot 95% Confidence	Very safe zone
7	Hot spot 99% Confidence	Extreme safe zone

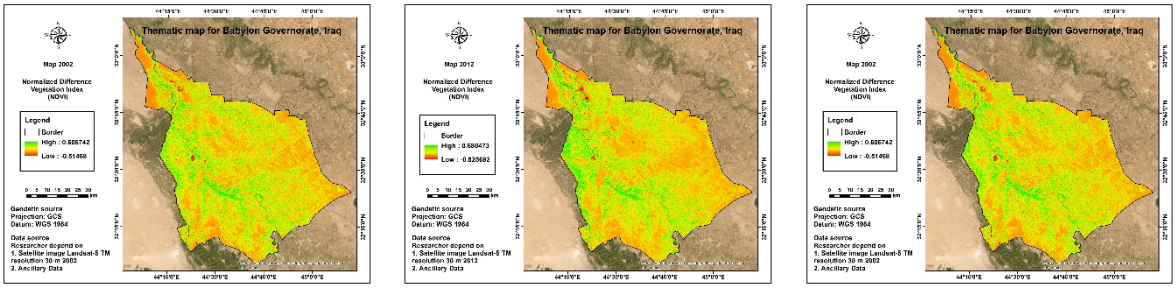
This study computed the NDVI values using October Landsat images (TM and OLI/TIRS) for each year from 2002, 2012, and 2022. Fig. 3 represents the results of the NDVI index in the study area. Fig. 4 represents the results of the hotspot analysis based on the NDVI index in the study area. The calculated NDVI, hotspot change map, and IDW for the years 2002, 2012, and 2022 are presented in Fig. 5. Using the G_i^* method of analysis, the Babylon Governorate area is divided into seven different zones, where the 99%, 95%, and 90% cold confidence levels

https://doi.org/10.31436/iiumej.v25i2.3030

have been categorized as extremely vulnerable, very vulnerable, and moderately vulnerable, respectively.

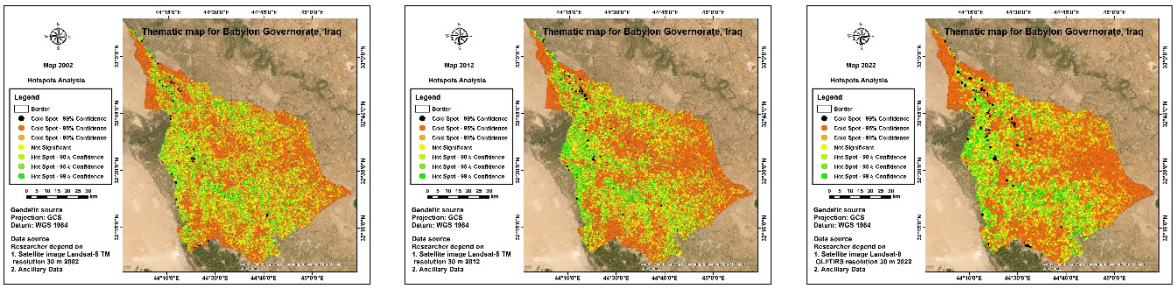
On the other hand, the 99%, 95%, and 90% hot confidence levels have been permitted as an extreme safe zone, very safe zone, and moderately safe zone, respectively. Besides, the middle category has been assumed to be a Non-significant zone. The first three kinds denote hazardous places with little to no vegetation, whereas the final three types denote safe areas with the fewest to the most plant clusters. It is possible for a non-significant zone, which serves as a buffer between a vulnerable zone and a safe zone, to transition into any of the other classifications depending on how well this balance is protected.

The information presented in Table 4 illustrates that the extremely vulnerable zone has increased from (13.47 to 41.85) km^2 in the study area between 2002 and 2022. Likewise, the very vulnerable zone also changed by approximately (733.77 to 1376.21) km^2 in the same period. Also, for the moderately vulnerable zone, there was an increase within the area it covered between the years (2002 and 2022), and it changed by approximately (1352.17 to 1473.23) km^2. At the same time, a decrease in four zones, the non-significant zone, moderately safe zone, very safe zone, and extreme safe zone, from the area it covered between (2002 to 2022) from (1334.31 to 1027.77) km^2 (1050.01 to 760.72) km^2, (638.73 to 481.61) km^2, and (214.60 to 175.66) km^2 respectively. According to the results of the extreme safe zone, there was an increase and then a decrease between the years (2002-2012) and (2012-2022) from (214.60 to 223.31) km^2, (223.31 to 175.66) km^2, respectively. Comparable with many researchers, such as [18][37][38].



(a) 2002 (b) 2012 (c) 2022

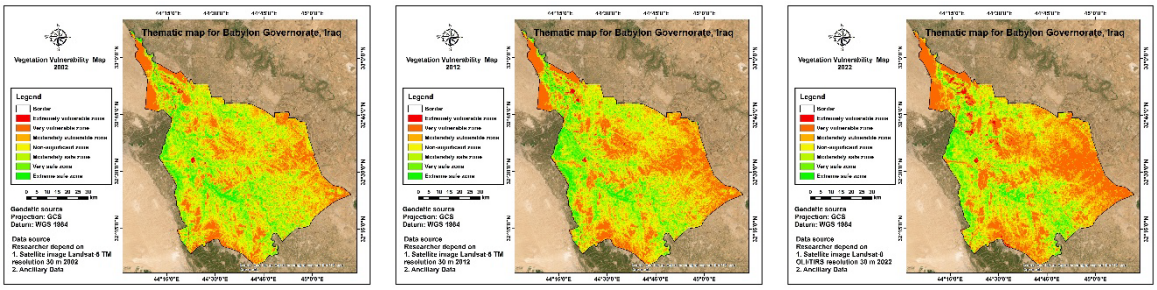
Figure 3. NDVI for years 2002, 2012, and 2022.



(a) 2002 (b) 2012 (c) 2022

Figure 4. Hotspot analysis method for years 2002, 2012, and 2022.

https://doi.org/10.31436/iiumej.v25i2.3030



(a) 2002 (b) 2012 (c) 2022

Figure 5. IDW method interpolation for years 2002, 2012, and 2022.

In Fig. 6, it can be seen that all three vulnerable zones increased between 2002 and 2022. In contrast, the three safe and non-significant zones decreased in 2002. This decrease was due to the lack of exploitation of water resources for agriculture, despite the increase in water between 2012 and 2022; according to the year 2002, the use of agricultural lands to be residential and the spread of residential slums. The exploitation of territory and natural resources is an essential component in the process of achieving sustainable development in the Governorate of Babylon. The destruction of green land can cause major harm to a region's environment and biodiversity. Again, alterations in the kinds of land present a significant challenge to regulators and managers, particularly in the Governorate of Babylon, where infrastructure construction is challenging and expensive. Therefore, those responsible for formulating policies and making plans must recognize these changes present a significant danger and act accordingly.

Table 4. Vegetation vulnerability changes in Babylon Governorate (2002-2022).

Zone	2002		2012		2022		Change trend
	Percentage (%)	Area (km^2)	Percentage (%)	Area (km^2)	Percentage (%)	Area (km^2)	
Extremely vulnerable zone	0.25	13.47	0.42	22.43	0.78	41.85	+ → +
Very vulnerable zone	13.75	733.77	18.32	977.75	25.79	1376.21	+ → +
Moderately vulnerable zone	25.34	1352.17	26.11	1393.26	27.60	1473.23	+ → +
Non-significant zone	25.00	1334.31	23.06	1230.80	19.26	1027.77	- → -
Moderately safe zone	19.67	1050.01	17.16	915.73	14.25	760.72	- → -
Very safe zone	11.97	638.73	10.75	573.78	9.02	481.61	- → -
Extreme safe zone	4.02	214.60	4.18	223.31	3.29	175.66	+ → -
Total	100.00	5337.05	100.00	5337.05	100.00	5337.06	

https://doi.org/10.31436/iiumej.v25i2.3030

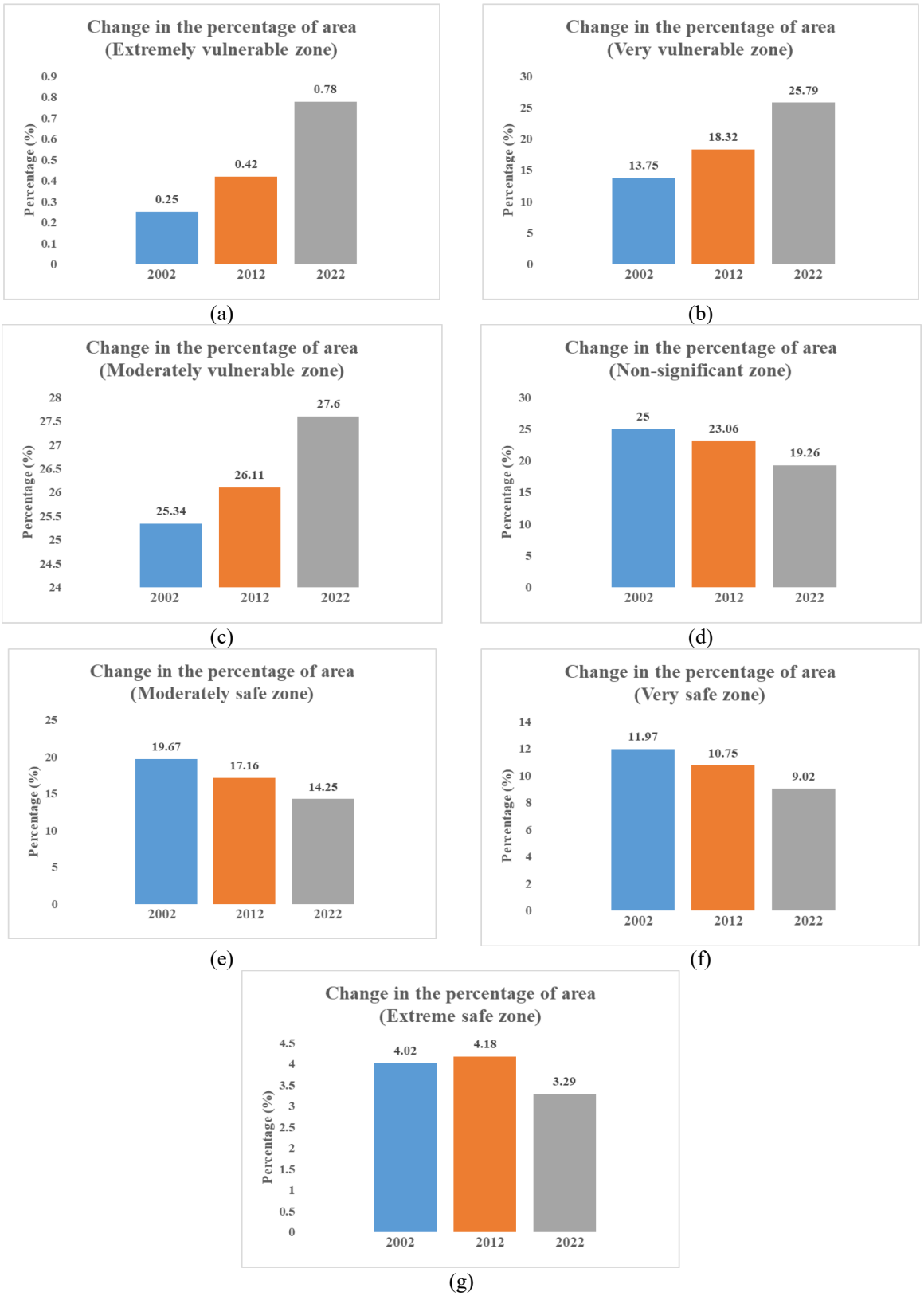


Fig. 6. Change in the percentage of the area (a to g) from 2002 to 2022.

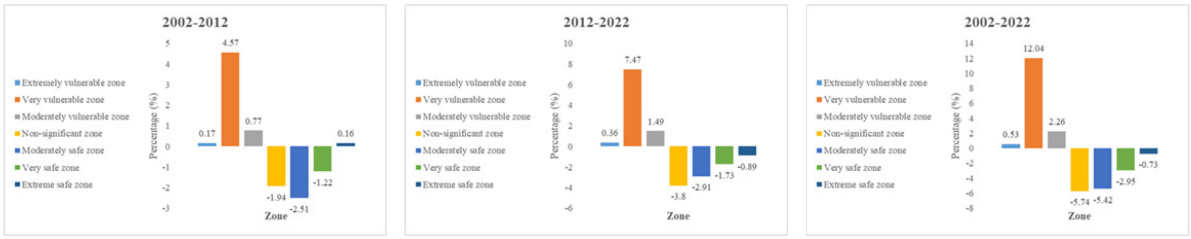
Table 5 shows that the change occurred in the study area in the very vulnerable zone and the lowest change in the extreme safe zone between the years (2002-2012) and about 4.57% and 0.16%, respectively. In contrast, the change occurred in the study area in a very vulnerable zone and the least in an extremely vulnerable zone, about 7.47% and 0.36% between the years

https://doi.org/10.31436/iiumej.v25i2.3030

(2012-2022), respectively. Finally, between the years (2002 and 2022), the change occurred in the study area in a very vulnerable zone, and the least change in an Extremely vulnerable zone was about 12.04% and 0.53%, respectively. The above analysis notes that the most changed zone is the very vulnerable zone. The area of gains and losses of each zone with vegetation vulnerability in Babylon Governorate for the years (2002-2012, 2012-2022, and 2002-2022). These changes may be linked directly to the interaction of physical and social variables such as the increase in population, lack of rain and water resources, climate change, population migration from the countryside to the city, and overgrazing. Fig. 7 elaborates on each category's gained and lost areas over each time interval.

Table 5. Gains and losses of each zone.

Zone	2002 - 2012		2012 - 2022		2002 - 2022	
	Percentage (%)	Area (km^2)	Percentage (%)	Area (km^2)	Percentage (%)	Area (km^2)
Extremely vulnerable zone	0.17	8.96	0.36	19.42	0.53	28.38
Very vulnerable zone	4.57	243.98	7.47	398.46	12.04	642.44
Moderately vulnerable zone	0.77	41.10	1.49	79.97	2.26	121.06
Non-significant zone	-1.94	-103.52	-3.8	-203.03	-5.74	-306.55
Moderately safe zone	-2.51	-134.28	-2.91	-155.01	-5.42	-289.28
Very safe zone	-1.22	-64.95	-1.73	-92.17	-2.95	-157.12
Extreme safe zone	0.16	8.71	-0.89	-47.64	-0.73	-38.93
Total	0.00	0.00	0.00	0.00	0.00	0.00



(a) 2002 (b) 2012 (c) 2022

Fig. 7. Gains and losses in the percentage of the area (a to c) from 2002 to 2022.

The analysis of the results and achievements of the studies related to the identification of land changes has great value because it leads to the justification of the results and the necessity of citing the research findings, and on the other hand, it provides the ground for future research for other researchers. Collecting information about land cover changes is essential to better understanding the relationships and interactions between humans and the natural environment. In this regard, remote sensing data has been one of the most important data sources for studying the spatial and temporal changes of land cover in different regions. Remote sensing multi-temporal datasets for processing and analysis allows mapping and identifying landscape changes and provide an effective step for sustainable local planning and management. In particular, by integrating RS and GIS techniques, it is possible to analyze and classify the changing patterns of land cover over a long period and, as a result, better understand the changes in the target range. Today, it is well established that satellite remote sensing and GIS are the most common methods for quantifying, mapping, and detecting land use and land cover

https://doi.org/10.31436/iiumej.v25i2.3030

change patterns because they have precise spatial analysis techniques and provide a suitable digital format for computer processing and repetitive data collection. Due to the capabilities of remote sensing and geographic information systems, scientists and researchers can detect large-scale changes in land use patterns with the knowledge they have acquired at the temporal and spatial levels, allowing politicians and regional authorities to make decisions.

4.2 Accuracy Assessment

The procedure was validated using reference data in 2022 and verified to be accurate. The methodology for the validation was based on the overall accuracy and kappa coefficient. A total of 280 validation points were surveyed based on an analysis of vegetation vulnerability maps and interpretation based on field observation and ancillary data such as topographic maps, high spatial resolution satellite images, and interpretation. According to the number of training samples analyzed to perform zone classification, the proportion of pixels classified in the correct or incorrect coverage class can be seen in Table 6, which indicates the acceptable error rate. On the other hand, the classification accuracy evaluation matrix was used to evaluate the overall classification accuracy, which is the main criterion for judging the results of the evaluations. Based on the assessment and validation of zone classification data, the overall accuracy value shows that the vegetation vulnerability accuracy for 2022 was equal to 90%. Also, the kappa coefficient for the classification map was 0.88. Therefore, according to the results of the accuracy evaluation, it can be said that the vegetation vulnerability classification process has been carried out with good accuracy and very high quality.

Table 6. Classification accuracy assessment matrix to vegetation vulnerability.

Zone	User's accuracy (%)	Producer's accuracy (%)
	2022	2022
Extremely vulnerable zone	100.00	77.50
Very vulnerable zone	89.19	82.50
Moderately vulnerable zone	81.25	97.50
Non-significant zone	82.93	85.00
Moderately safe zone	88.37	95.00
Very safe zone	100.00	97.50
Extreme safe zone	92.68	95.00
Overall accuracy = 90%		
Kappa coefficient = 0.88		

5. CONCLUSION

Analyzing the results and achievements of studies on identifying vegetation vulnerability is valuable because it justifies the results and the necessity of citing the research findings, while also providing a foundation for future research. The research results have shown that over the twenty years studied in Babylon Governorate, a significant portion of vegetation and crop cover has been lost and replaced by built-up lands. These findings can guide and inspire vegetation vulnerability studies and evaluation ideas in other regions. The region experienced a high percentage of vegetation drought from 2002 to 2022, highlighting the importance of monitoring vegetation changes. Remote sensing techniques have proven effective in observing vegetation vulnerability changes over substantial regions, and the hotspot analysis method has been

particularly useful in identifying areas of vulnerability. This research not only sheds light on past and present vegetation changes but also offers valuable methodologies for future studies in similar contexts.

REFERENCES

[1] Kadhum ZM, Jasim BS, Obaid MK. (2020) Change detection in city of Hilla during period of 2007-2015 using Remote Sensing Techniques. IOP Conf Ser Mater Sci Eng. doi: 10.1088/1757-899X/737/1/012228

[2] Al-Helaly MH, Alwan IA, Al-Hameedawi AN. (2021) Land covers monitoring for Bahar-Al-Najaf (Iraq) based on sentinel-2 imagery. J Phys Conf Ser. doi: 10.1088/1742-6596/1973/1/012189

[3] Al-Helaly MH, Alwan IA, Al-Hameedawi AN. (2021) Assessing land cover for Bahar Al-Najaf using maximum likelihood (ML) and artificial neural network (ANN) algorithms. J Phys Conf Ser. doi: 10.1088/1742-6596/1973/1/012190

[4] [Kadhum ZM, Jasim BS, Al-saedi ASJ. (2023) Improving the spectral and spatial resolution of satellite image using geomatics techniques Improving the Spectral and Spatial Resolution of Satellite Image Using Geomatics Techniques. 040011

[5] Jasim O, Hasoon K, Sadiqe N. (2019) Mapping LCLU Using Python Scripting. Engineering and Technology Journal, 37(4A): 140–147. doi: 10.30684/etj.37.4a.5.

[6] Al-Anbari R, Oday Zakariya, Mohammed ZT. (2016) Environmental and Urban Land Use Analysis by GIS in AL-Shaab of Baghdad as a case study. Eng. &Tech.Journal 34

[7] Husein H, Jasim O, Mahmood S. (2018) Proposal of building a standard geodatabase for urban land use. MATEC Web of Conferences, 1621–5. doi: 10.1051/matecconf/201816203024.

[8] Hasan SH, Al-Hameedawi ANM, Ismael HS. (2022) Supervised Classification Model Using Google Earth Engine Development Environment for Wasit Governorate. IOP Conf Ser Earth Environ Sci. doi: 10.1088/1755-1315/961/1/012051

[9] Jabbar HK, Hamoodi MN, Al-Hameedawi AN. (2023) Urban heat islands: a review of contributing factors, effects and data. In: IOP Conf. Ser. Earth Environ. Sci. IOP Publishing. p 12038.

[10] Al-Saedi ASJ, Kadhum ZM, Jasim BS. (2023) Land Use and Land Cover Analysis Using Geomatics Techniques in Amara City. Ecol Eng, 9161–169.

[11] Muhammad R, Zhang W, Abbas Z et al. (2022) Spatiotemporal Change Analysis and Prediction of Future Land Use and Land Cover Changes Using QGIS MOLUSCE Plugin and Remote Sensing Big Data: A Case Study of Linyi, China. Land. doi: 10.3390/land11030419

[12] Yu L, Wu Z, Du Z et al. (2021) Insights on the roles of climate and human activities to vegetation degradation and restoration in Beijing-Tianjin sandstorm source region. Ecol Eng. doi: 10.1016/j.ecoleng.2020.106105

[13] Alwan I, Ziboon AR, Khalaf A. (2019) Monitoring of Agricultural Drought in the Middle Euphrates Area, IraqUsing Landsat Dataset. Engineering and Technology Journal, 37(7A): 222–226. doi: 10.30684/etj.37.7a.1.

[14] Al-Hameedawi ANM, Abdulkhudhur R, Abdulkareem AO. (2022) Ground penetration radar based digital image processing for reinforcement corrosion in concrete. Innovative Infrastructure Solutions, 7(4): 241.

[15] Sun CJ, Zhang WQ, Li XG, Sun JL. (2019) Evaluation of ecological effect of gully region of loess plateau based on remote sensing image. Trans Chin Soc Agric Eng, 35(12): 165–172.

[16] Frazier PS, Page KJ. (2000) Water body detection and delineation with Landsat TM data. Photogrammetric engineering and remote sensing, 66(12): 1461–1468.

[17] Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR. (2016) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural hazards, 83(2): 947–987.

[18] Abdullah S, Barua D. (2022) Combining Geographical Information System (GIS) and machine learning to monitor and predict vegetation vulnerability: An Empirical Study on Nijhum Dwip,

Bangladesh. Ecological Engineering, 178(January): 106577. doi: 10.1016/j.ecoleng.2022.106577.
[19] Reddy CS, Manaswini G, Satish K V. et al. (2016) Conservation priorities of forest ecosystems: Evaluation of deforestation and degradation hotspots using geospatial techniques. Ecological Engineering, 91333–342. doi: 10.1016/j.ecoleng.2016.03.007.
[20] Singh S, Reddy CS, Pasha SV et al. (2017) Modeling the spatial dynamics of deforestation and fragmentation using Multi-Layer Perceptron neural network and landscape fragmentation tool. Ecological Engineering, 99543–551. doi: 10.1016/j.ecoleng.2016.11.047.
[21] Ziboon A, Albayati M, Dalhel F. (2022) Monitoring Soil Degradation in The Mesopotamian Plain Using GIS and Remote sensing Techniques. Engineering and Technology Journal, 40(5): 649–660. doi: 10.30684/etj.v40i5.2121.
[22] Russ JC. (2006) The image processing handbook. CRC press.
[23] Vargas-Cuentas N, Roman-Gonzalez A, Muñoz LA. (2017) Use of satellite images for droughts studying: The bolivian case.
[24] Hadjimitsis DG, Papadavid G, Agapiou A et al. (2010) Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. Natural Hazards and Earth System Sciences, 10(1): 89–95.
[25] Bruce CM, Hilbert DW. (2006) Pre-processing methodology for application to Landsat TM/ETM+ imagery of the wet tropics. Rainforest CRC Cairns, Australia.
[26] Chittimalli SK. (2016) Reflectance-based calibration and validation of the landsat satellite archive. South Dakota State University.
[27] Zaharaddeen I, Baba II, Zachariah A. (2016) Estimation of land surface temperature of Kaduna metropolis, Nigeria using Landsat images. Science World Journal, 11(3): 36–42.
[28] Suresh S, Mani K. (2017) Application of remote sensing in understanding the relationship between NDVI and LST. Int. J. Res. Eng. Technol. 6
[29] Jasim OZ. (2019) Using of machines learning in extraction of urban roads from DEM of LIDAR data: Case study at Baghdad expressways, Iraq. Periodicals of Engineering and Natural Sciences, 7(4): 1710–1721. doi: 10.21533/pen.v7i4.914.
[30] Chakravorty S. (1995) Identifying crime clusters: The spatial principles. Middle States Geographer, 2853–58.
[31] Yang W, Zhao Y, Wang D et al. (2020) Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang river in Huangshan, China. International journal of environmental research and public health, 17(8): 2942.
[32] Jasim O, Hamed N, Abdulgabar T. (2018) Change detection and building spatial geodatabase for Iraqi marshes. MATEC Web of Conferences, 1623–6. doi: 10.1051/matecconf/201816203021.
[33] Alwan IA, Aziz NA. (2021) An accuracy analysis comparison of supervised classification methods for mapping land cover using sentinel 2 images in the al-hawizeh marsh area, southern iraq. Geomatics and Environmental Engineering, 15(1): 5–21. doi: 10.7494/geom.2021.15.1.5.
[34] Chabuk A, Al-Ansari N, Hussain HM et al. (2019) Landfill sites selection using MCDM and comparing method of change detection for Babylon Governorate, Iraq. Environmental Science and Pollution Research, 26(35): 35325–35339. doi: 10.1007/s11356-019-05064-7.
[35] Jasim BS, Al-Bayati ZMK, Obaid MK. (2018) Accuracy of horizontal coordinates of cadastral maps after geographic regression and their modernization using gis techniques. International Journal of Civil Engineering and Technology, 9(11): 1395–1403.
[36] Al-Hameedawi ANM. (2020) Comparison between Saaty's approach and Alonso and Lamata's approach in site selection process. IOP Conf Ser Mater Sci Eng. doi: 10.1088/1757-899X/737/1/012217
[37] Alam KF, Ahamed T. (2022) Assessment of Land Use Land Cover Changes for Predicting Vulnerable Agricultural Lands in River Basins of Bangladesh Using Remote Sensing and a Fuzzy Expert System. Remote Sensing, 14(21): 5582.
[38] Moradi E, Darabi H, Heydari Alamdarloo E et al. (2023) Vegetation vulnerability to hydrometeorological stresses in water-scarce areas using machine learning and remote sensing techniques. Ecol Inform. doi: 10.1016/j.ecoinf.2022.101838

https://doi.org/10.31436/iiumej.v25i2.3040

AN INVESTIGATION ON THE RESISTANCE OF RUBBERIZED ASPHALT MIXTURE TO AGING AND MOISTURE DAMAGE

AYMAN HASSAN AL-QUDAH[1], SUHANA KOTING[1*], MOHD RASDAN IBRAHIM[1], MUNA M. ALIBRAHIM[2], NISHANTHINI JEGATHEESAN[1]

[1]*Center for Transportation Research, Department of Civil Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[2]*Department of Architectural Engineering, Hijjawi Faculty of Engineering Technology, Yarmouk University, 21163 Irbid, Jordan*

**Corresponding author: suhana_koting@um.edu.my*

(Received: 26 September 2023; Accepted: 15 April 2024; Published online: 15 July 2024)

ABSTRACT: Asphalt pavements in Malaysia suffer severe deterioration, such as raveling and stripping, because the bonding and adhesion between asphalt and aggregate is poor. Heavy rains throughout the year are the main cause of deterioration or a contributing cause to accelerated failure of the pavement structure. Therefore, it is very important to inspect the moisture damage and permeability behavior of asphalt pavement, which can effectively help in a pre-preservation program for asphalt pavement. This study aims to evaluate the effect of adding crumb rubber (CR) on the pavement permeability and moisture damage in highly unfriendly environmental conditions. Laboratory investigation was employed to achieve this aim. The laboratory investigation involved the preparation of two different ACW 14 rubberized asphalt mixtures (unaged and aged). Each type of rubberized mixture contained four CR contents (0%, 5%, 10%, and 15%). The mixtures were tested for air voids, Marshall stability and flow, moisture susceptibility, permeability, and water absorption. The laboratory investigation showed that the CR was appropriate for use up to 10% in asphalt mixtures. There were improvements in the performance of the asphalt mixtures including the Marshall stability, stiffness, durability and resistance against moisture damage. Furthermore, the use of CR prompted better permeability and water absorption. ACW 14 mixture is considered a practically impervious mixture. In conclusion, the application of rubberized pavement can enhance various phases of pavement life and structure by improving the structural performance and reducing environmental impact and landfill issues in the future.

***ABSTRAK*:** Turapan asfalt di Malaysia mengalami kemerosotan yang teruk seperti permukaan turapan jalan yang terpecah (rancak) dan berlubang (pelucutan) kerana ikatan dan lekatan antara asfalt dan agregat adalah lemah. Hujan lebat sepanjang tahun adalah punca utama kemerosotan atau penyumbang kepada struktur turapan rosak dengan cepat. Oleh itu, adalah sangat penting bagi memeriksa kerosakan disebabkan lembapan dan sifat kebolehtelapan turapan asfalt ini dapat membantu dengan berkesan dalam program pra-pemeliharaan turapan asfalt. Kajian makmal telah digunakan bagi mencapai tujuan ini. Kajian melibatkan penyediaan dua campuran asfalt bergetah ACW 14 yang berbeza (tidak berumur dan berumur). Setiap jenis campuran getah mengandungi empat kandungan CR (0%, 5%, 10%, dan 15%). Campuran telah diuji samada mempunyai lompang udara, kestabilan dan aliran Marshall, kerentanan lembapan, kebolehtelapan, dan penyerapan air. Kajian makmal menunjukkan CR sesuai digunakan sehingga 10% dalam campuran asfalt. Terdapat peningkatan dalam prestasi campuran asfalt termasuk kestabilan Marshall, kekakuan, ketahanan dan rintangan terhadap kerosakan lembapan. Tambahan pula, penggunaan CR mendorong kebolehtelapan dan penyerapan air. Campuran ACW 14 dianggap sebagai campuran praktikal tidak telap air. Kesimpulan, aplikasi turapan bergetah dapat meningkatkan

pelbagai fasa hayat dan struktur turapan dengan menambah baik prestasi struktur dan mengurangkan kesan alam sekitar dan isu tapak pelupusan pada masa hadapan.

KEYWORDS: *Crumb rubber modified bitumen, aged asphalt, dense-graded asphalt mix, rubberized asphalt mixture, moisture damage.*

1. INTRODUCTION

Roadways are an important part of any country's infrastructure systems and contribute significantly to economic growth by providing for the movement of commodities and freight, and social benefits and quality of life through access to services, work, and schools, general mobility, and leisure travel [1]. Based on the structural performance, road pavements are classified into two broad categories with numerous advantages and disadvantages, which are flexible and rigid pavements. Flexible pavements, commonly known asphalt pavements, are widely applied as highway, expressway, and airport pavements primarily owing to their flexibility, high riding quality, high resistance to skidding, and good serviceability [2]. Asphalt pavements are constructed with bituminous materials and require transverse joint-free characteristics that require less maintenance efforts resulting in lower construction and maintenance costs when compared to rigid pavements. In Malaysia, nearly 99% of paved roads are asphalt pavements. However, due to the rapid increase in traffic and extreme environmental factors, asphalt pavements are exposed to many distresses such as rutting, cracking, corrugation, shoving, and stripping after long-term operation, which can negatively affect the performance and service life of pavements. Unfortunately, even the best-designed asphalt pavements deteriorate over time due to axial wheel loads, moisture, and temperature.

Despite the fact that asphalt and aggregate materials used in the construction of asphalt pavements are relatively cheap materials [3] and perform good serviceability [4], the disruption to traffic and costs of maintenance and replacement of deteriorated pavement surfaces are an important issue, leading to a demand for more durable materials. Since asphalt is a temperature-sensitive material, it does not have sufficient resistance against pavement distresses that are caused by moisture and temperature [3]. Therefore, road engineers and designers must consider safety requirements, especially in wet conditions, along with economic aspects. To achieve this goal, four basic requirements must be considered when designing: construction cost, traffic loads, environmental factors, and the quality and availability of materials. Failure to comply with these requirements may result in economic losses resulting from maintenance and rehabilitation work or premature failure of the pavement. In Malaysia, asphalt pavements deteriorate rapidly and do not meet the anticipated design life due to heavy vehicle owners not adhering to the maximum permissible weights. Furthermore, asphalt pavements are very sensitive infrastructure elements, and the performance of the pavements are then affected by environmental and climatic factors such as moisture, temperature, heavy rains, and floods. Pavements are vulnerable to surface damage due to increased temperature, but more severe structural damage is caused by water damage arising from wetter winters and more frequent intense rainfall events [5].

Pavement permeability is a physical property, which is essential and crucial to characterizing the performance of pavement and its quality. Pavements with high permeability will rapidly deteriorate, and be exposed to oxidation (aging), raveling, and stripping due to air and water penetration, allowing the water to infiltrate the underlying layers [6]. The presence of water in pavement structure plays a very important role in accelerating pavement distress. Water damage occurs by breaking the bonds between the asphalt binder and the fine and coarse aggregates, thereby reducing durability and strength. A study by Omar et al. [7] showed that water causes a loss of adhesion between the asphalt and the aggregate interfaces (stripping).

https://doi.org/10.31436/iiumej.v25i2.3040

Referring to AL-Qudah [6], water deteriorates the underlying layers of roads. Accordingly, the maintenance of roads affected by wet tropical climatic conditions is a high percentage of the total road construction costs annually, i.e., in 2021, Malaysia was spent RM900 million on repairing roads damaged by rains and floods [8]. Therefore, pavement distresses prevent the optimal function of public road transport and are considered one of the largest factors contributing to fatal motor vehicle crashes.

In Malaysia, the design life of asphalt pavements is 15 years. Due to many advantages such as good strength and durability, adequate skid resistance, low permeability and low binder content, and applicability in any layer of pavement structure, the dense-graded asphalt mix is adopted as a wearing course intended for general use.

Statistics from Transport Ministry show that the number of road accidents in Malaysia is increasing almost annually. There were 370286 road accidents in 2021, and 545588 accidents in 2022, whereas approximately 1.1% of these accidents involved fatal accidents. According to a research study published by the Malaysian Institute of Road Safety in 2021, the design and condition of road infrastructure are one of the main causes of road accidents. Moreover, the pavements deteriorate over time due to traffic and environmental factors. This, in turn, requires regular road maintenance.

Overcoming the deficiencies in the performance of conventional asphalt pavements has resulted in a need for alternatives that can improve the performance of asphalt pavement mixtures. One viable alternative is rubberized asphalt mixture.

Several studies have been carried out to evaluate the use of CR in asphalt pavement mixtures. AL-Qudah [6] indicated that rubberized mixture is a term that refers to the blending of CR with an asphalt binder to enhance the durability characteristics of the following asphalt pavement mixture during service life. Wulandari and Tjandra [9] showed that the use of CR improved asphalt mixture strength and consistency. Sarsam and Hasan [10] evaluated the effect of water on asphalt mixtures containing CR based on the indirect tensile strength (ITS) test. Zhu et al. [11] assessed the performance of rubberized asphalt mixture containing reclaimed asphalt pavement. The penetration test, viscosity test, ductility test, softening test, compressive strength test, Marshal stability test, wheel tracking test, ITS test, dynamic creep, fatigue test, environmental impact were employed by different researchers [6, 12-15] to evaluate the performance evolution of crumb rubber modified bitumen (CRMB) binders and rubberized mixtures, respectively.

The existing studies have mainly evaluated the performance of rubberized mixtures through mechanical property tests, and there needs to be more research on the effect of CR on asphalt mixtures in terms of permeability behavior and water absorption before and after the aging process. To address this research gap in the body of knowledge, this study aims to investigate the effect of CR in asphalt mixtures to resist thermal oxidative aging and moisture damage. It also includes studying the effect of pavement permeability and moisture damage in unfriendly conditions.

Several influential factors affect pavement quality and permeability. The significant factors are aggregate, asphalt, temperature, moisture, and traffic load. Aggregate forms approximately 95% of the total weight of the mix and 85% of the volume. Thus, the properties of combined aggregates highly affect road quality, which offers the possibility of investment in these properties towards resisting varying ranges of challenging environmental and geotechnical conditions and externally applied loads. Aggregates used in pavements' asphalt layers (surface and binder) are constantly stressed due to traffic axial loads that cause wear and tear and premature crushing. For a high-quality pavement, grain particles should have good

resistance to withstand unnecessary rushing and stresses and influence pavements' traffic loading transfer capability to a greater extent. The suitability of aggregates for use in asphalt pavement mixtures is determined by several tests, including specific gravity, water absorption, soundness, abrasion, crushing value, flakiness, and elongation index. The type and size of aggregate are some of the most influential properties. Not only should the totals be strong, but they also require a proper size to make the pavement act monolithically. The aggregate size plays a crucial role in determining each property of pavement, such as durability, skid resistance, resilient modulus, indirect tensile fatigue, indirect tensile strength, stiffness, aging, stability, permeability, rutting, and resistance to moisture damage [3, 6, 16-18]. The aggregate size shall be in the allowable range to achieve strong mechanical interlocking in asphalt mixtures. Pavement containing high coarse aggregate content has high flexibility and porosity, and a faster rate of oxidation binder and loss of adhesion due to contact with water and air, and ultimately pits and exhibits some series of damage states such as raveling, stripping, and potholes. On the contrary, when the proportion of fine aggregate is high, the flexibility, porosity, and skid resistance of the pavement are low, consequently making the pavement more vulnerable to cracking, abrasion, and bleeding. Furthermore, the content of asphalt is mainly related to aggregate size. Not only does the size of the aggregate affect the quality of the pavement, but its shape also plays a very important role. The form of the aggregate may have angular, cubical, rounded, flaky, or elongated particles. Compared to angular, spherical, or cubical particles of the same aggregate, the flaky and elongated particles have less strength and durability. Pavement quality depends on the surface texture [7] and mineralogy of the aggregate [19]. The smoother the surface texture of the aggregate, the lower the absorption of the asphalt binder, and the lower the bond strength.

The type and content of asphalt are the second important consideration that affects the quality of the pavement. Asphalt properties vary over time and from one type to another, and are affected by weather (such as moisture, wind and air temperature) and pavement material parameters (such as physical and thermal properties). A binder serves two major functions; first, to hold the aggregates firmly (adhesion properties), and another function, to act as a sealant to water infiltration (waterproofing). Hence, the asphalt content and its properties significantly influence the pavement quality in terms of stiffness, strength, durability, raveling, rutting, fatigue life, and moisture damage [20]. Accordingly, the incorporation of solid waste materials to asphalt mixtures requires laboratory testing (such as Marshall stability and flow, volumetric properties, ITS, stiffness and Cantabro durability) to ensure that requirements are met. Several factors may contribute to asphalt deterioration, such as oxidation, volatilization, and polymerization, which are caused by moisture and temperature variations over time. Therefore, it is necessary to understand asphalt aging since the asphalt is fully observed under different temperature and moisture conditions. This study gives a quantitative measure for rubberized mixtures with and without oxidative aging.

The third factor that influences road quality is temperature. The viscous-elastic plastic characteristic of asphalt pavements varies with temperature. Therefore, the durability problem of the pavements is prominent because bituminous materials are a kind of temperature sensitive material [3]. Bituminous materials will contract as they fall in temperature and expand as they rise in temperature. Small amounts of expansion and contraction are typically accommodated without excessive damage, however extreme temperature variations cause severe to catastrophic failures. Small variations in the temperature have an unnoticeable effect on pavements. Asphalt pavements can suffer large transverse cracks due to excessive contraction in cold weather. On other hand, asphalt pavement is prone to permanent deformation (rutting) due to excessive expansion in hot weather [5]. Temperature and freeze-thaw cycling lead to acceleration of pavement distresses [21].

https://doi.org/10.31436/iiumej.v25i2.3040

In addition, moisture is the main cause of deterioration or a contributing cause giving accelerated failure of the pavements [7] and can also lead to large economic and human losses. It is true that asphalt pavements are susceptible to surface damage due to temperature valuations, but more severe structural damage is caused by increased water damage arising from wetter winters and more frequent intense rainfall events. As a result, the presence of water inside the asphalt pavement structures causes premature failure and damage even with the absence of traffic loading, especially in humid-tropical regions that are subjected to high rainfall intensity and floods throughout the year. This would lead countries to spend millions of dollars on maintenance and rehabilitation of network roads; e.g. Malaysia spends millions each year on repairing road damage caused by rains and floods.

Finally, one of the primary functions of asphalt pavements is load distribution. However, traffic loads and environmental factors (heavy rains and temperature variations) deteriorate pavement over time. This deterioration begins very slowly and is cumulative, exacerbated by repeated traffic, aging, environmental factors, and construction deficiencies. Not knowing the quantification of all expected loads on the pavement under different vehicle weights and environmental conditions at any operational stage will fail. Various distresses such as rutting, fatigue cracking, and roughness would occur if the pavement cannot sustain stress due to frequent heavy wheel loads. Past studies have reported that various pavement distresses, such as severe depressions, potholes, and alligator cracks, are linked to excessive wheel loads.

2. MATERIAL AND EXPERIMENTAL PROGRAM

2.1. Materials

The materials used for the production of unaged and aged rubberized mixtures included aggregates, asphalt, filler, and ordinary Portland cement (OPC). At the same time, the CRMB binders were made from the use of CR and asphalt.

Granite aggregate supplied from Pens Industries, located in Kuala Perlis, Malaysia were used as coarse and fine aggregates. The dense-graded aggregate with a maximum nominal size of 14 mm (ACW 14) per JKR-2008 standard [22] was adopted. The basic physical properties of aggregates and the grain size distribution curve are presented in Table 1 and Fig. 1, respectively.

Table 1. Basic physical properties of aggregate

Property	Specification	Value	
		Coarse Aggregate	Fine Aggregate
Water absorption (%)	< 2	0.84	0.48
Bulk specific gravity	-	2.56	2.63
Apparent specific gravity	-	2.62	2.67
LA abrasion loss (%)	< 25	19.35	-

In Malaysia, asphalt PEN (60/70) is used to construct asphalt pavements because it is more resistant to traffic loads and heavy rainwater. The asphalt PEN (60/70) was used and sourced by Stolthaven Sdn. Bhd., which is located in Port Klang, Malaysia. Table 2 shows the physical and rheological properties of virgin asphalt.

2% of OPC with a specific gravity of 3.11 was added as an anti-stripping agent [22]. Moreover, CR used with a specific gravity of 1.10 was obtained through the mechanical shredding at ambient temperature, and collected from Gcycle Company located in Kedah,

Malaysia. The CR-sized 0.60-mm (30 mesh) originally produced from recycled vehicle tires was introduced as a reinforcing material.

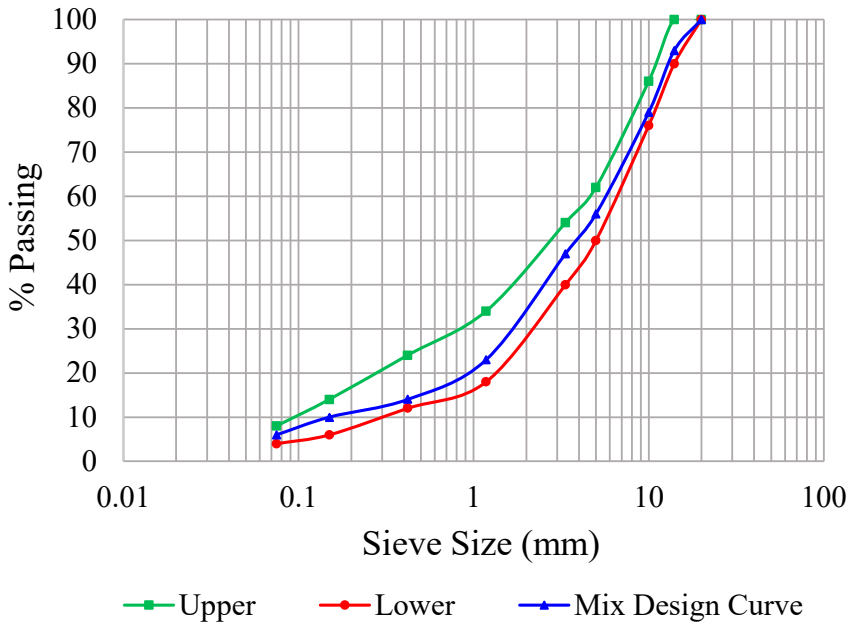


Figure 1. Mix design fractions of dense-graded aggregate (ACW 14).

Table 2. Physical properties of asphalt PEN (60/70)

Property	Specification	Value
Penetration at 25 ºC (0.1mm)	60 - 70	63.2
Ductility at 25 ºC (cm)	> 100	109.7
Softening Point (ºC)	48 - 56	52.1
Specific Gravity at 25 ∘C	1.01 - 1.06	1.032
Viscosity at 135 ºC (mPa.s)	-	491.3

2.2. Production of CRMB Binders

The CRMB binder comprises a bituminous material compound in which rubber in a proper form and proportion is incorporated using an appropriate technique. To this end, the wet process was adopted in this study to prepare CRMBs, where a suitable and effective dispersion of CR particles takes place in the asphalt binder phase. Unlike the wet process, in the dry process, the aggregate is directly mixed with CR before the asphalt binder is added to the asphalt mixture, which means that CRMBs and aged CRMBs cannot be used in the dry process. The CRMB binders were first produced by heating the virgin asphalt up to 180ºC before adding the CR. The required amount of CR was slowly added to the heated asphalt, and then asphalt was blended with CR using a propeller blade mixer at a high speed of 4000 rpm for 60 min, as shown in Fig. 2. The mixing temperature was maintained at 180ºC, following the same procedure reported by AL-Qudah et al. [3]. Four unaged CRMB binders were made, replacing virgin asphalt with the CR at 0%, 5%, 10%, and 15% of asphalt weight. These low CR replacement levels were adopted because the high replacement levels negatively affect the mechanical properties of the mixture by reducing the interfacial adhesion properties.

https://doi.org/10.31436/iiumej.v25i2.3040

Asphalt, like any organic material, is susceptible to temperature fluctuations. Aging is one of the significant problems facing asphalt during the production and service life of pavements due to high temperatures. To evaluate the effect of temperature (aging), four different types of aged binders were produced, in which virgin asphalt and three CRMB binders were subjected to short-term simulation aging using a rolling thin film oven, representing the aging process during production. Eight different CRMB binders were produced and used to prepare eight different ACW 14 rubberized mixtures, and the contents of the binders and mixtures are summarized in Table 3.

Figure 2. Blending process of CR with virgin asphalt

Table 3. Contents of CRMB binders and rubberized mixtures

CR (%)	CR Weight (g)	Asphalt Weight (g)	Aging	CRMB Designation	Weight of Aggregates (g)	ACW 14 Rubberized Mixture	
						Acronym	No. of samples for each test
00	0.0	63.4	no	CRMB-00	1156.6	unaged CR-00	3
05	3.2	60.2	no	CRMB-05	1156.6	unaged CR-05	3
10	6.4	57.0	no	CRMB-10	1156.6	unaged CR-10	3
15	9.5	53.9	no	CRMB-15	1156.6	unaged CR-15	3
00	0.0	63.4	yes	aged CRMB-00	1156.6	aged CR-00	3
05	3.2	60.2	yes	aged CRMB-05	1156.6	aged CR-05	3
10	6.4	57.0	yes	aged CRMB-10	1156.6	aged CR-10	3
15	9.5	53.9	yes	aged CRMB-15	1156.6	aged CR-15	3

2.3. Preparation of Rubberized Asphalt Mixtures

Unaged and aged ACW 14 rubberized mixtures were designed by adhering to the Marshall mix design method [23] in conformity with JKR-2008 standard [22]. All ACW 14 varieties were produced from combined aggregates blended with asphalt at 5.2% (OBC). The mixing process continued until the binder coated all the aggregates. The mixing was done at a temperature of 150°C. Each sample was compacted by applying 75 blows per face at 45°C.

https://doi.org/10.31436/iiumej.v25i2.3040

To prepare unaged asphalt mixtures, the unaged-CRMBs were used. Meanwhile, the aged mixtures were produced with aged CRBMs. Eight different rubberized mixtures were designed and divided into two equal groups: unaged and aged mixtures, as presented in Table 3.

2.4. Test Methods

2.4.1. Air Void Test

The air voids in the total mix (VTM) are important parameters because they permit the performance and properties of the mix to be predicted for the lifespan service of asphalt pavements. Eq. (1) was used to calculate the VTM for the ACW 14 rubberized mixtures in accordance with ASTM D3203 [24].

$$\mathrm{VTM} = \left[1 - \frac{\mathrm{G_{mb}}}{\mathrm{G_{mm}}}\right] \times 100\% \quad (1)$$

where VTM is the voids in the total mix (%), G_{mb} is the bulk specific gravity of the compacted mix and G_{mm} is the maximum theoretical specific gravity of the paving mix.

2.4.2. Marshall Stability Test

Marshall stability and flow test was carried out as per ASTM D6927 [25], which involved determining the resistance and displacement of Marshall specimens in a direction perpendicular to the applied load at a deformation rate of 50 mm/min. The Marshall specimens were cured in a water bath for 30 min at 60°C before testing.

2.4.3. Moisture Susceptibility

The retained Marshall stability (RMS) was used to investigate the moisture susceptibility. In this test, the Marshall specimens were cured at 60ºC for 1 day, and at 25ºC for 28 days to achieve saturation levels up to 80 %. The RMS was measured by AASHTO T283 [26]. The load was applied until the specimen failure at a rate of 50.8 mm/min.

2.4.4. Permeability Test

The falling head permeability test was performed according to FM 5-565 standard [27]Marshall specimens were immersed in water at 27ºC for 2 hours before the test started. The permeability of the samples was calculated based on the falling head principle, as illustrated in Eq. (2).

$$\mathrm{k} = \left(\frac{\mathrm{aL}}{\mathrm{At}}\right) \ln\left(\frac{\mathrm{h_1}}{\mathrm{h_2}}\right) \quad (2)$$

where, k is the coefficient of permeability (cm/sec), L is the height of the sample (cm), a is the pipe cross-sectional area (cm), A is the sample cross-sectional area, t is the time of flow between h_1 and h_2 (sec), and h_1 and h_2 are the initial and final head of water (cm), respectively.

2.4.5. Water Absorption Test

Water absorption before and after the aging process was evaluated as an important physical property of pavement more related to moisture-induced damages. The water absorption test was measured according to AASHTO T165 standard procedure [28]. The weights of dry samples were measured and recorded to the nearest 0.1g as W_{dry}. The samples were then immersed in water at 25ºC for 28 days. Every 7 days, all the samples were taken out, wiped, and the SSD condition weights of the samples were measured and recorded as W_{SSD1}, W_{SSD2},

https://doi.org/10.31436/iiumej.v25i2.3040

W_{SSD3}, W_{SSD4} after 7 days, 14 days, 21 days, and 28 days of immersion process, respectively. The average percentage of water absorption (P_{wa}) was calculated as given Eq. 3.

$$P_{wa} = \left[\frac{W_{SSDi} - W_{dry}}{W_{dry}}\right] x\ 100\% \quad (3)$$

2.4.6. Durability Test of Optimal Rubberized Mixture

The durability performance of optimal rubberized mixture was tested for stiffness, Cantabro durability, ITS, and tensile strength ratio (TSR). The ratio of Marshall stability to Marshall flow is an indication of the stiffness of the asphalt mixture, and is also adopted as a standard by JKR [22]. It is well documented that stiffness is a measure of the material resistance to permanent deformations and shear stresses [3, 29].

The Cantabro durability is crucial to substantiate the production of a good quality wearing course by resisting crushing, and degradation. This test was conducted using a Los Angeles machine and without steel balls in accordance with EN 12697 standard [30].

As stipulated in AASHTO T283 [26], the purpose of the ITS test is to determine the durability of asphalt mixtures which depend on the cohesiveness of the asphalt films, and also to identify the sensitivity of the mix to moisture-induced damage by estimating the TSR for dry and wet conditioned specimens. This test was done by loading a compacted cylindrical specimen across its diametric vertical plane at a constant deformation rate of 50 mm/min. The ultimate load for each sample at failure was noted and utilized to estimate the tensile strength.

3. RESULTS AND DISCUSSION

3.1. Effect of Permeability

Permeability is the physical property that describes how water flows through a particular medium. The permeability is controlled by the size and shape distribution of aggregates and interconnections of the air voids. A study by Zaniewski and Yan [31] showed that the permeability of asphalt pavement is influenced by several factors including air voids, size and surface texture of aggregate, lift thickness and asphalt content. Air void content is the predominant factor that controls the permeability of pavement, where permeability increases when the air voids increase. This is due to the fact that increasing the void content leads to an increase in the interconnectivity of internal voids, which accelerates the process of infiltration and penetration of water into the pavement structure. Also, permeability increases with the increase of aggregate size. Conversely, when lift thickness increases and the potential for interconnected air void decreases, thus permeability decreases. Giompalo [32] reported that permeability decreases with increasing asphalt content because high asphalt content helps in reducing the interconnectivity of internal voids within the pavement, then decreasing the permeability. Some of the photos related to typical pavement distress due to moisture damage, one of Malaysia's main issues, are shown in Fig. 3 and Fig. 4.

https://doi.org/10.31436/iiumej.v25i2.3040

Figure 3. Pavement distresses caused by moisture and traffic loads.

The high permeability of asphalt pavements causes thermo-oxidative aging, raveling, stripping, and rapid deterioration of the underlying layers. This occurs due to the penetration of air and water. As a result, pavement surfaces become less safe, less serviceable, and of lower riding quality. A correlation has been found between poor pavement surfaces and higher road crash rates (larger economic and human losses). Poor pavement surfaces contribute to fatal motor vehicle crashes [33] and are, therefore, a concern for road infrastructure owners and users. At this juncture, an important question arises: How can pavement surface conditions contribute to road accidents? The pavement surface can deteriorate to the level where it becomes dangerous. This includes broken asphalt concrete, sinkholes, uneven road surfaces, potholes, and road cracks. If a driver hits a pothole, the tire could burst, causing the vehicle to veer into another lane and collide. In addition, over time, the pavement surface becomes smoother, and thus, skid resistance decreases. This, in turn, increases road accidents, especially in wet conditions (rainy days).

Figure 4. Infiltration and penetration of water into pavement layers.

3.2. Effect of Water Absorption

Water absorption is the process of absorbing water by materials either rapidly or gradually with the macromolecular structure of the materials. Water absorption is the major problem of asphalt pavements today because it is affected by direct rainfall intensity. The threat of the fluctuation in moisture is widely recognized as one of the main causes of long-term failure of pavement exposed to the atmosphere or in contact with an aqueous media. There are several recognized modes of water absorption problems, such as differential swelling related to concentration gradients, swelling linked to the degradation of the macromolecular structure of the material, and plasticization of the aggregate [34]. The main cause of the failure of asphalt pavement structure is its affinity with high water absorption, especially for underlying layers, which has serious consequences for its durability characteristics. Water damage to asphalt surface layers occurs by breaking the bonds between the asphalt binder and the fine and coarse

https://doi.org/10.31436/iiumej.v25i2.3040

aggregates (stripping), leading to reduced durability and strength of pavements. Water below the asphalt layers is also an important issue. Poor drainage can saturate, swell, and deteriorate the supporting materials in the aggregate base and subgrade, allowing axial wheel loads to cause excess damage to the pavement structure.

3.3. Air Void Content

The VTM values of rubberized asphalt mixtures for different conditions are graphically drawn in Fig. 5. It is seen that the increment in the CR content led to the increment of the VTM values for unaged and aged asphalt mixtures. The unaged CR-15 and aged CR-15 mixtures possessed higher VTM values by 6.9% and 7.3%, respectively, in comparison to unaged CR-00 and aged CR-00 mixtures. This trend occurs because the point of contact between the aggregates is lower when the CR content is increased. The CR particles absorbed the oily components of the binder while the CR blended with the asphalt [6]. This resulted in increased CR mass (swelling CR), which is required to encapsulate the aggregate and subsequently fill the voids between the aggregate. Fig. 5 also shows that VTM, after the aging process, enhanced slightly. This behavior is related to the fact that when CRMB is subjected to aging, it becomes challenging to compact as its mass and stiffness increase. Furthermore, the swelling of CR particles continued to occur when they were placed in the oven. However, although the VTM of the mixtures increased as CR content increased, the VTM values were within the acceptable level (3-5%). Compared with control mixtures, the performance of rubberized varieties is better due to the increased VTM required to allow the pavement to undergo additional compaction under traffic loads while providing space for asphalt to flow without flushing onto the pavement surface.

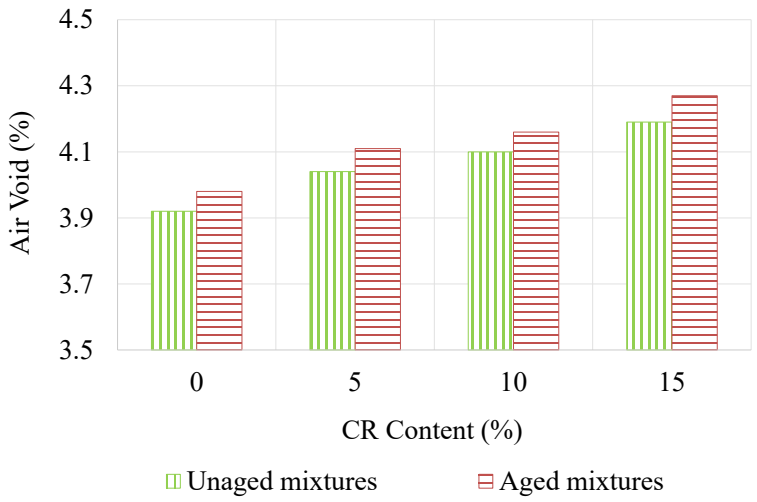


Figure 5. Air voids results before and after the aging process at different CR content

3.4. Marshall Stability

Fig. 6 shows the effect of CR on the stability of the unaged and aged mixtures. With the increase of CR in the asphalt, the stability of mixtures increased. The stability value before the aging process was at its maximum at 10% by 21.4% compared to that of control mixture. This occurs because the asphalt binder became more viscous with the addition of CR, thereby increasing the interfacial adhesion properties and stiffness [6]. The inclusion of more CR decreases the convenience adherence between the mixture compositions, thus deteriorating the stiffness performance. A high replacement level of asphalt with CR results in a brittle, stiff

https://doi.org/10.31436/iiumej.v25i2.3040

mixture due to the loss of large volatiles from the binder at high temperatures, making the mixture more susceptible to cracking and abrasion. According to Izaks and Haritonovs [35], the combination of higher stiffness and more brittle behavior will likely lead to a shorter fatigue lifespan and a greater probability of thermal cracking occurring in the field.

As expected, the stability values of the mixtures with the aging process significantly increased due to the increased stiffness of aged mixtures. It can be explained that the binders lost the volatile materials and became more viscous after exposure to high temperatures. This strengthened the adhesion properties between the aggregate and the asphalt. All of the aged rubberized mixtures had higher stability values than that of the aged control mixture. The aged CR-15 mixture recorded the highest stability value of 29.6% compared to the aged CR-00 mixture. However, more than the stiffness alone is needed to give the complete picture because the ductility and brittleness of the mix will also affect its performance concerning cracking. Therefore, although the aged CR-15 mixture has the highest stability value after aging, this gives a negative indicator and reduces the potential resistance to thermal cracking in the field. Nevertheless, all asphalt mixtures with different CR levels exhibited higher values than the minimum allowable limit, i.e., 8 kN. This indicates that the CR can increase the rutting resistance under heavy traffic loads. Such a result agrees with findings study by Nashruddina, et al. [13].

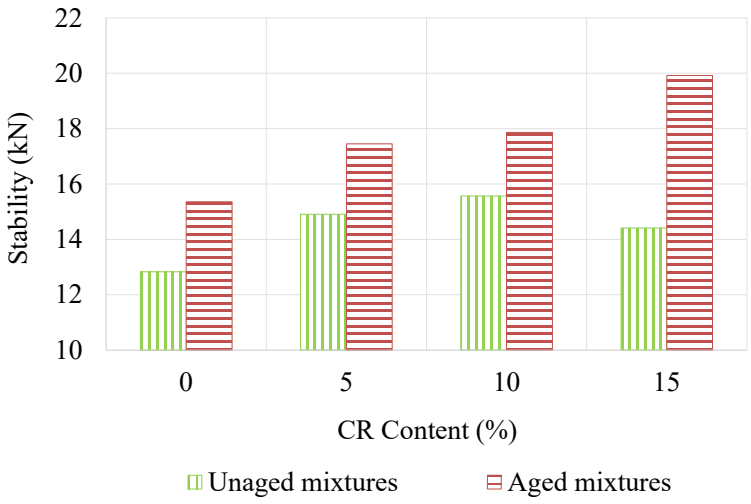


Figure 6. Results of Marshall stability before and after aging at different CR contents

3.5. Marshall Flow

The Marshall flow results at different CR contents with and without thermo-oxidative aging are presented in Fig. 7. The flow values of unaged mixtures decreased with the addition of CR to 10%, and then a noticeable increment in flow value was observed thereafter. After the aging process, there was an inconsistent trend in the relationship between the CR contents and the flow values. However, all the Marshall flow values were within the acceptable range of values (2-4 mm) recommended by JKR for ACW 14 [22]. Compared with unaged mixtures, the aged asphalt mixtures had lower flow values. This reduction could be related to the stiffer mixtures obtained and the decreased fluidity of the aged binders because the binders were very viscous due to the loss of the volatiles.

https://doi.org/10.31436/iiumej.v25i2.3040

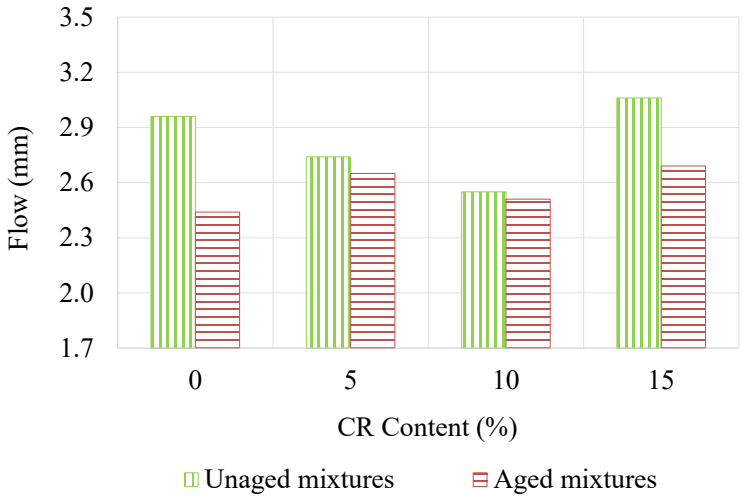


Figure 7. CR content and Marshall flow results before and after aging

3.6. Moisture Susceptibility

The results of RMS versus different conditions and aging for all ACW 14 mixtures are presented in Fig. 8. The RMS values of the rubberized mixtures appeared to be higher than those of the control mixtures. This signified that the mixtures replaced with CR had higher resistance to moisture damage than the control. This trend can be attributed to the increased viscosity of CRMBs because of exposure to high temperatures, which resulted in the formation of thick films of CRMB around the aggregate, reducing the effect of water on the bonding between materials. The RMS results showed that the mixtures with 10% CR without oxidative aging had higher RMS values at 1 and 28 days of curing by an increase of up to 37.7% and 38.2%, in comparison to those mixtures without CR under the same conditions. In the meantime, the highest RMS values for aged mixtures were at 15% and 10% of CR by 43.2% and 38.1% when compared to control mixtures at 1 and 28-day condition, respectively. This is an indication that the application of CR in the asphalt mixture can significantly improve the performance of asphalt pavement by improving the interfacial adhesion properties and reducing the degree of thermo-oxidative aging and moisture-induced damage. A similar trend was reported in past studies using CR as fine aggregate in asphalt mixtures [12, 36]. Water damage resistance varies greatly with the method of incorporating CR into asphalt. Similar to past studies, rubberized mixtures prepared using asphalt pre-blended with CR have greater resistance to water damage than mixtures prepared by directly blending CR with aggregates and then asphalt. The variation in water damage resistance values of rubberized mixtures is evidence that CRMB has good resistance to damage from aging and moisture.

https://doi.org/10.31436/iiumej.v25i2.3040

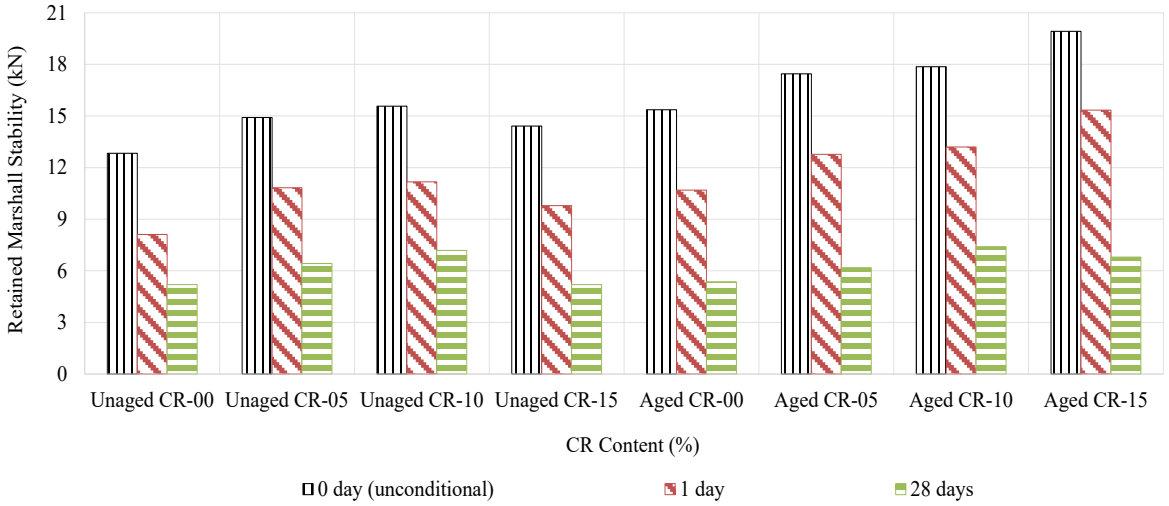


Figure 8. RMS results before and after aging versus CR content

3.7. Permeability

The permeability values of all the mixtures versus CR for different aging conditions are graphically presented in Fig. 9. With an increase in the CR replacement level, the permeability values of the mixtures increased, except for the unaged CR-05 mixture, which had the lowest value. This trend is due to increased VTM values of rubberized asphalt mixtures. The figure also depicts that the aging process causes a slight increase in the permeability values of the mixtures due to the increased VTM after the aging process, which is attributable to the explanation described earlier in section 4.3. Compared to the permeability values of control mixtures before and after aging, the mixtures with 15% CR achieved the highest increase values of 8.1% and 7.8%. Based on the results, the effect of CR on permeability was higher than that of aging. This behavior indicates that the physical and chemical properties of asphalt affected by adding CR were higher compared to the effect of temperature. In other words, the resistance of CRMBs to the impact of high temperature was improved. This proposed potential application of rubberized pavement will perform better than conventional pavement.

Depending on the permeability, asphalt mixtures are classified into several categories. According to the permeability scale proposed by Vardanega and Waters, all the mixtures with CR replacement are classified in category A. This category (A) confirms the mixtures have very low permeability. Moreover, Mullen's categorization indicates that the mixtures are classified as practically impervious mixtures [37]However, when air voids increase, the permeability increases. Unlike the interconnected air voids, isolated air voids do not correlate with high permeability as there are no flow channels that allow water to penetrate. In other words, the pavement structure can be impervious even with some air voids within it.

https://doi.org/10.31436/iiumej.v25i2.3040

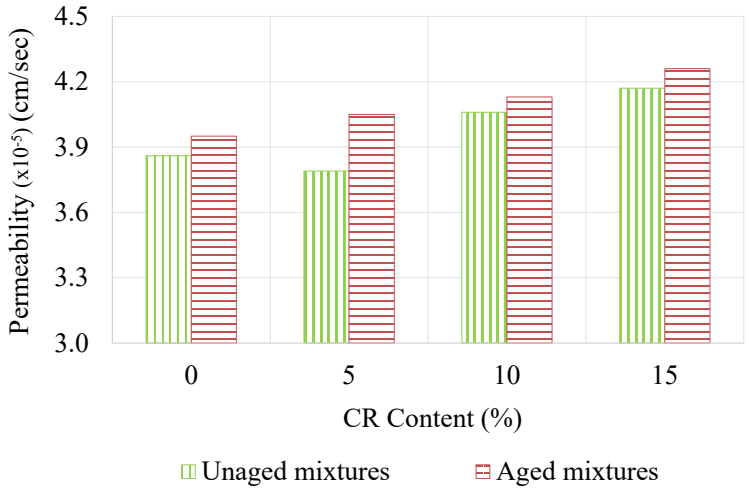


Figure 9. CR content and permeability relationship before and after aging

3.8. Water Absorption

The presence of water in asphalt pavements is unavoidable. The water absorption values of the mixtures versus the CR for oxidative aging conditions are graphically drawn in Fig. 10. The results showed that the CR content affected the water absorption of the mixtures and had a similar trend and that, for any aging condition, the water absorption values of the compacted mixture specimens increased with an increase in the CR replacement level. Moreover, all the rubberized mixtures had higher water absorption values than the control mixtures because of the higher VTM of rubberized mixtures with and without aging conditions. This confirms that VTM plays an important role in the water infiltration and absorption of pavements, which results in moisture damage and loss of interface adhesion, thereby affecting the strength and durability of pavement. However, CR has a significant role in reducing the degree of oxidative aging and leads to a good coat, subsequently decreasing the water absorption by aggregate in asphalt mixtures in the long run and during its service life.

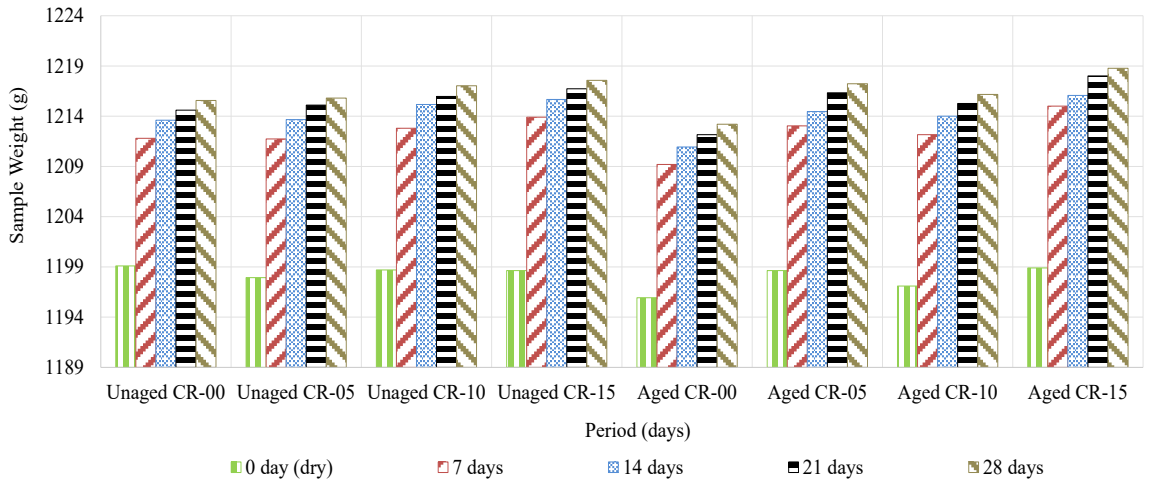


Figure 10. Weights of unaged and aged rubberized mixtures affected by water content

3.9. Durability of Optimal Rubberized Mixture

Asphalt mixtures will undergo thermo-oxidative aging during production and service life. Therefore, compared to mixtures without aging, the performance investigation of asphalt mixtures after aging is the most important consideration. The laboratory investigation for the aged mixtures revealed that the aged mixture with an additional 10% CR (aged CR-10) appeared to have the best performance properties compared to other aged mixtures. The analysis results of the durability of the aged CR-10 mixture are given in Table 4. The results show the stiffness of the aged CR-10 mixture is superior to the aged-control mixture by up to 15.43% and is greater than the JKR requirement (2 kN/mm) [22].

As shown in Table 4, the aged mixture with 10% CR had a higher Cantabro durability value than the control mixture. This is caused by the addition of CR, which made the mixture stiffer and brittle due to the volatile loss of the asphalt components. The increased loss of Cantabro durability can also be attributed to the increased VTM with the addition of CR. The Cantabro loss requirement must be less than or equal to 15% [22], and the results showed that all fulfilled this requirement.

Table 4. Durability characteristics of aged mixtures

Mixture	Stiffness (kN/m)	Cantabro loss (%)	ITS (kPa)	TSR (%)
Aged CR-00	6.35	5.41	1236.1	84.21
Aged CR-10	7.33	6.56	1304.8	86.07

The average values of ITS and TSR for the two aged mixtures are also attached in Table 4. The aged rubberized samples in a diametric strength direction perform similarly, as shown for the Marshall stability. Compared with the aged CR-00 mixture, the ITS of the aged CR-10 mixture increased up to 5.6%, mainly due to the more viscous CRMB resulting from the thermo-oxidative aging process. This in turn led to the formation of thick films of the CRMB around the aggregates, thus strengthening the interfacial adhesion properties. Moreover, the TSR of the aged CR minus 10 mixture was higher than TSR of the aged CR-00 mixture, indicating sufficient resistance against moisture damage. Besides, all the aged mixtures were above the minimum required value (80%) of TSR prescribed by AASHTO T283 [26].

4. CONCLUSION

This study investigated the effect of CR in asphalt mixtures to resist thermal oxidative aging and moisture damage. Moreover, the causes and effect of pavement permeability and moisture damage in unfriendly conditions was examined. Based on the results and analysis, the following conclusions can be drawn:

1. The permeability and water absorption values of the mixtures with and without aging increased when CR content increased due to the increased VTM. The ACW 14 rubberized mixture is considered a practically impervious mixture if designed properly.
2. Asphalt mixture with CR had high resistance to aging and moisture damage due to the improved adhesion properties. Therefore, rubberized asphalt pavement can enhance various phases of pavement life and structure by improving structural performance and reducing moisture and temperature (aging).
3. The permeability and water penetration of asphalt pavements increases over time because of the failure caused by wheel loads and severe environmental factors. A high permeability of asphalt pavement causes thermo-oxidative aging, raveling, stripping, and

rapid deterioration of the underlying layers, which have serious consequences for its durability characteristics.

ACKNOWLEDGEMENT

The authors would like to acknowledge University of Malaya, Malaysia, for providing support to this research paper through the KW IPPP-Research Maintenance Fee (RMF) Individual (no. RMF0476-2021).

REFERENCES

[1] Administration, FH. (2018) Sustainable Pavement Program. U.S. Department of Transportation, Washington, USA.
[2] Hassani A, Taghipoor M, Karimi MM. (2020) A state of the art of semi-flexible pavements: Introduction, design, and performance. Construction and Building Materials, 253: p. 119196.
[3] AL-Qudah A, Abdul Rahim M, Ghazaly ZM, et al. (2018) Effect of aged crumb rubber bitumen on performance dense graded mix in Malaysia. International Journal of Civil Engineering and Technology, 9(4): p. 1356-1369.
[4] Zhang H, Chen Z, Zhu C, Wei C. (2020) An innovative and smart road construction material: thermochromic asphalt binder, in New Materials in Civil Engineering, Editors. Butterworth-Heinemann. p. 691-716.
[5] Qiao Y. (2015) Flexible pavements and climate change: impact of climate change on the performance, maintenance, and life-cycle costs of flexible pavements. University of Nottingham.
[6] AL-Qudah AH. (2018) Mechanical Properties of Asphaltic Concrete Mixture with Crumb Rubber and Aged Binder. School of Environmental Engineering, Universiti Maaysia Perlis, Malaysia.
[7] Omar HA, Yusoff NIM, Mubaraki M, Ceylan H. (2020) Effects of moisture damage on asphalt mixtures. Journal of Traffic and Transportation Engineering (English Edition), 7(5): p. 600-628.
[8] New Strait Times Agency. (2021) Flood-damaged road repairs to cost RM900 million. Kuala Lumpur, Malaysia.
[9] Wulandari PS, Tjandra D. (2017) Use of crumb rubber as an additive in asphalt concrete mixture. Procedia engineering, 171: p. 1384-1389.
[10] Sarsam SI, Hasan EA. (2017) Evaluating Water Damage Resistance of Recycled Asphalt Concrete Mixtures. Journal of Engineering, 23(1): p. 156-168.
[11] Zhu J, Ma T, Dong Z. (2020) Evaluation of optimum mixing conditions for rubberized asphalt mixture containing reclaimed asphalt pavement. Construction and Building Materials, 234: p. 117426.
[12] Mashaan N. (2016) Use of waste tyre rubber in enhancing the mechanical properties of stone mastic asphalt mix. Civil Engineering Department, University of Malaya, Malaysia.
[13] Nashruddin TF, Ramadhansyah PJ, Norhidayah AH, et al. (2014) Evaluation of Binder Absorption in Asphalt Mixture with Various Aging Conditions Using Rice Method. Jurnal Teknologi (Sciences & Engineering), 71(3): p. 53-56.
[14] Bansal S, Misra AK, Bajpai P. (2017) Evaluation of modified bituminous concrete mix developed using rubber and plastic waste materials. International Journal of Sustainable Built Environment, 6(2): p. 442-448.
[15] Wangn T, Xiao F, Zhu X, Huang B, Wang J, Amirkhanian S. (2018) Energy consumption and environmental impact of rubberized asphalt pavement. Journal of Cleaner Production, 180: p. 139-158.
[16] Pouranian MR, Haddock JE. (2020) Effect of aggregate gradation on asphalt mixture compaction parameters. Journal of Materials in Civil Engineering, 32(9): p. 04020244.
[17] Garcia VM, Barros L, Garibay J, Abdallah I, Nazarian S. (2020) Effect of aggregate gradation on performance of asphalt concrete mixtures. Journal of Materials in Civil Engineering, 32(5): p. 04020102.

[18] Bharath G, Reddy KS, Tandon V, Reddy MA. (2021) Aggregate gradation effect on the fatigue performance of recycled asphalt mixtures. Road Materials and Pavement Design, 22(1): p. 165-184.
[19] Shivakumara SN, Mallesh SK, Anjum MI. (2015) Effect of particle shape on the workability and properties of conventional bituminous mixes. Int. J. Res. Eng. Technol, 4(4): p. 50-54.
[20] Asghar MAA. (2015) Evaluation of strength characteristics of recycled bituminous pavement materials, in School of Civil Engineering & Surveying. University of Southern Queensland.
[21] Din IM, Mir MS, Farooq MA. (2020) Effect of freeze-thaw cycles on the properties of asphalt pavements in cold regions: a review. Transportation Research Procedia, 48: p. 3634-3641.
[22] Malaysian Standards. (2008) Standard Specification for Road Works, Section 4, Flexible Pavement. Jabatan Kerja Raya (JKR), Kuala Lumpur, Malaysia.
[23] Annual Book of ASTM Standards. (2002) Standard Test Method for Preparation of Bituminous Mixtures Using Marshall Apparatus, in ASTM D1559, American Society for Testing and Materials (ASTM) International: West Conshohocken, PA, US.
[24] Annual Book of ASTM Standards. (2005) Standard Test Method for Percent Air Voids in Compacted Dense and Open Bituminous Paving Mixtures, in ASTM D3203, American Society for Testing and Materials (ASTM) International: West Conshohocken, PA, US.
[25] Annual Book of ASTM Standards. (2005) Standard Test Method for Marshall Stability and Flow of Bituminous Mixtures, in ASTM D6927 American Society for Testing and Materials (ASTM) International: West Conshohocken, PA, US.
[26] Annual Book of AASHTO Standards. (2007) Standard Method of Test for Resistance of Compacted Asphalt Mixtures to Moisture-Induced Damage, in AASHTO T283, American Association of State and Highway Transportation Officials (AASHTO): Washington, US.
[27] Florida Standards. (2004) Florida Method of Test for Measurement of Water Permeability of Compacted Asphalt Paving Mixtures, in FM5-565, Florida Department of Transportation: Gainesville, Florida, US.
[28] Annual Book of AASHTO Standards. (2002) Standard Method of Test for Effect of Water on Cohesion of Compacted Bituminous Mixtures, in AASHTO T165, American Association of State and Highway Transportation Officials (AASHTO): Washington, US.
[29] Babalghaith AM. (2021) Development of Stone Mastic Asphalt Pavement Design with Palm Oil Clinker as Fine Aggregate Replacement, in Civil Engineering. University of Malaya. Malaysia.
[30] Annual Book of EN Standards. (2017) Standard Test Method for Particle loss of porous asphalt specimens, Bituminous mixtures, in European Standards (EN) 12697, British Standards (BS) Institution: London, UK.
[31] Zaniewski JP, Yan Y. (2013) Hot Mix Asphalt Concrete Density, Bulk Specific Gravity, and Permeability.
[32] Giompalo JA. (2010) Permeability of hot mix asphalt concrete as affected by binder content. West Virginia University.
[33] Mkwata R, Chong EEM. (2022) Effect of pavement surface conditions on road traffic accident-A Review. in E3S Web of Conferences. EDP Sciences.
[34] Fadhil TH, Mohammed HA, Ahmed TY. (2015) An empirical relationship between asphalt and water absorption of coarse aggregates in HMA. Journal of Engineering and Sustainable Development, 19(1): p. 132-146.
[35] Izaks R, Haritonovs V, Klasa I, Zaumanis M. (2015) Hot Mix Asphalt with High RAP Content. Procedia Engineering, 114: p. 676-684.
[36] Cao W. (2007) Study on properties of recycled tire rubber modified asphalt mixtures using dry process. Construction and building materials, 21(5): p. 1011-1015.
[37] Vardanega P, Waters T. (2011) Analysis of asphalt concrete permeability data using representative pore size. Journal of Materials in Civil Engineering, 23(2): p. 169-176.

https://doi.org/10.31436/iiumej.v25i2.3211

GEOMETRIC QUALITY ANALYSIS OF TERRESTRIAL LASER SCANNING DATA FOR INDUSTRIAL USAGE

ALI SALAH J. AL-SAEDI [1,2*], FANAR M. ABED [3], IMZAHIM A. ALWAN [2]

[1] *Surveying Department, Amara Technical Institute, Southern Technical University, Misan, Iraq*
[2] *Civil Engineering Department, University of Technology, Baghdad, Iraq*
[3] *Surveying Engineering Department, College of Engineering, University of Baghdad, Baghdad, Iraq*

**Corresponding author: ali.alsaedi@stu.edu.iq*

(Received: 4 March 2024; Accepted: 28 June 2024; Published online: 15 July 2024)

ABSTRACT: Terrestrial laser scanning is a potential emerging technology increasingly used in several applications, including reverse engineering, digital reconstruction, deformation monitoring, forensic crime scene preservation, and construction (AEC) applications. The data tolerance accepted in these applications ranges from tens of millimeters (e.g., historical monument digitization) to tens of micrometers (e.g., industrial high-precision manufacturing and assembling). Instrument mechanism, atmospheric conditions, object surface characteristics, and scan geometry are the four main factors that affect the laser point clouds produced by the Time of Flight (TOF) 3D laser scanner. Consequently, research groups worldwide have put a significant effort into modeling the sources of TOF-TLS errors and design-specific performance evaluation methodologies. This paper investigated the influence of scanning geometry parameterized by incidence angle and the range on the quality of TOF-TLS data in industrial sites. The quality of an indoor sample dataset of an industrial case study was studied and assessed. The results showed that the incidence angle and range parameters substantially impacted the quality of the TOF-TLS data. The suggested methodology can accurately correct the laser data to eliminate the incidence angle and range effects. A revised and optimized point cloud dataset was reconstructed by utilizing these features in conjunction with the approximated quality of the individual points. Furthermore, when assessing the quality of individual point clouds, the accuracy validation obtained through the RMSE value was 3 mm based on ground-truth reference points. On the other hand, the standard deviation values computed through the Multi-Scale Model-to-model cloud (M3C2) analysis were revealed to reach 1mm, which shows better performance results than the Cloud-to-Cloud (C2C) and Cloud-to-Model (C2M) comparison analysis. However, the proposed method may result in the elimination of several significant laser points. These points of high incidence angle values are not eliminated in every instance. The effect of scanning geometry, represented by the angle of incidence with the normalized intensity of the scanning points, should be studied intensively in future studies.

ABSTRAK: Terestrial pengimbas laser adalah teknologi yang berpotensi besar dalam pelbagai aplikasi, termasuk kejuruteraan balikan, rekonstruksi digital, pengawasan perubahan bentuk, pemeliharaan tempat kejadian jenayah forensik, dan aplikasi konstruksi (AEC). Penerimaan data toleransi dalam aplikasi ini adalah dalam julat lingkungan beberapa milimeter (cth: digitalisasi monumen sejarah) sehingga beberapa mikrometer (cth: industri pembuatan berkejituan tinggi dan pemasangan). Mekanisme alatan, keadaan atmosfera, ciri-ciri permukaan objek, dan geometri pengimbas adalah empat faktor utama yang memberi kesan terhadap gumpalan asap titik laser terhasil daripada masa terbang (TOF) 3D pengimbas laser. Oleh itu, kumpulan pengkaji sedunia telah berusaha bersungguh-sungguh dalam pemodelan sumber ralat TOF-TLS dan mereka kaedah penilaian prestasi reka bentuk tertentu. Kajian ini dipengaruhi oleh geometri pengimbas berparameter sudut tuju dan julat kualiti data

TOF-TLS dalam kawasan industri. Kualiti sampel set data dalam bangunan menggunakan kajian pembelajaran industri telah diteliti dan dinilai. Dapatan kajian menunjukkan sudut tuju dan julat parameter sangat mempengaruhi kualiti data TOF-TLS. Kaedah yang dicadangkan ini dapat memperbetulkan arah secara tepat iaitu kesan pengurangan sudut tuju dan julat. Set data rawak titik yang diubah dan dioptimumkan telah dibina semula dengan menggunakan ciri-ciri ini bersempena kualiti anggaran titik-titik individu. Tambahan, apabila menaksir kualiti titik rawak individu, pengesahan ketepatan yang diperoleh melalui nilai RMSE dijumpai sebanyak 3 mm berdasarkan titik-titik panduan kesahihan bumi. Di samping itu, nilai sisihan piawai yang dikira melalui analisis Model-ke-model Skala-Pelbagai Rawak (M3C2) didapati mencapai 1 mm, di mana menunjukkan dapatan prestasi lebih baik berbanding analisis Rawak-ke-rawak (C2C) dan Rawak-ke-Model (C2M) secara bandingan. Walau bagaimanapun, kaedah yang dicadangkan ini mungkin menyebabkan penyisihan beberapa titik laser penting. Titik-titik ini yang mempunyai nilai sudut tuju yang tinggi adalah tidak disisihkan pada setiap masa. Kesan pengimbas geometri, merupakan sudut tuju dengan keamatan ternormal titik-titik pengimbas, perlu dikaji dengan lebih lanjut pada masa depan.

KEYWORDS: *Terrestrial Laser Scanning, Industrial, Quality Analysis, Scanning Geometry, Range, Incidence Angle.*

1. INTRODUCTION

In the last two decades, Terrestrial Laser Scanners have become frequently used systems in reverse engineering and the quality control of facilities and infrastructures [1–3]. TLS device advantages include fast acquisition possibilities, sizeable spatial coverage, and high measurement accuracy [4]. Using a spherical coordinate system, the TLS provides a three-dimensional scene visualization by measuring the distances to the surfaces of the scanned objects. In addition to recording a horizontal and vertical angle for each point, it also measures the range from the hit point on the surface to the scanner location [5]. Noise affects scan measurements, even though TLS measurements are precise. Ultimately, point cloud measurements contain both systematic and random unavoidable errors. During data pre-processing stages, such as registration or segmentation, spread errors of individual points are propagated into end products, such as 3D reconstructions [6,7]. Individual point quality is influenced by four main factors: (i) the mechanism of the scanner, (ii) the atmospheric conditions and environment, (iii) the properties of the object being scanned, and (iv) the scanning geometry [5,8]. All of the factors affect the measurement of individual points, potentially by introducing noise or by altering the intensity and shape of the signal that is being emitted. There are many factors over which the user has limited control, such as the scanner, the atmospheric, and the object properties. The only component the user may control is the scanning geometry, as the user can decide where the scan should take place and, as a result, the perspective from which a point cloud is viewed [5]. In addition, when discussing the data, precision, outliers, and other incorrect points must be considered. Therefore, it is necessary to identify and eliminate untrusted points and outliers from the dataset or follow an adjustment procedure to clean the data for further accuracy analysis and more robust measurement data [5].

Scanning geometry is defined by the range and the point incidence angle computed to a surface [5]. This factor is well studied and investigated in Airborne Laser Scanning data [9–11]. On the other hand, the scanning geometry of the TLS data has been studied in numerous studies. For instance, in the investigation of intensity return using a variety of scanning geometry settings, the effects of range and incidence angle on the intensity measurements of (TLS) have been analyzed by Kaasalainen et al. [12], who investigated correction methods for

various TLS instruments and targets. Other studies by Tan and Cheng [13] suggested a new way to eliminate these range and incidence angle effects on the TLS intensity data. This method is based on linear interpolation of intensity values from reference targets. In this context, Pesci and Teza [14] designed an experiment to investigate the effect of surface irregularities on intensity data, depending on the TLS survey. This experiment involved acquiring TLS-based data depending on a rotating artificial target with flat and irregular surfaces. The results showed that the differences in the intensity of the backscattered signal depend on surface type and the incidence angle. Some previous studies proposed various strategies and modeled for eliminating the influence of inappropriate scan geometry on TLS data. Soudarissanane et al. [5] focused on understanding the influence of scan geometry on the quality of individual points in a point cloud, specifically the local measurement noise. Following this, the study successfully modeled the dependence of measurement noise on distance and incidence angle for planar surfaces. Later studies investigated the influence of incidence angle on the precision of terrestrial laser scanning points, where they could identify and correct the noise due to incidence angle, along with other influencing factors, to provide a better measurement quality [15]. Moreover, Pardiñas et al. [3] hypothesized a methodology for building an error model of TLS measurements using object distances and incidence angles. A point cloud of a tunnel section is used to analyze the spatial distribution of errors. The error model is applied in this case. The findings demonstrated that the point density, as a surface fitted to the point cloud, counteracts the angle of incidence's impact on the point cloud quality close to the tunnel gable. Miriam et al. [16], discussed the influence of incidence angle (IA) on TLS distance measurements. They developed a new methodology for comparing TLS distances measured directly with reference distances. The study proved that at close range, other error effects were more pronounced than the influence of IA. On the other hand, the systematic effect of IA at a distance of about 30 m, was obviously recorded. The effect of the scanning geometry is used in many studies, for example, model deformation in man-made structures [3,17] as well as objects in natural sites that use scanning geometry constraints [18–20].

This paper discusses the analysis, assessment, and correction of 3D point clouds generated by terrestrial laser scanning systems for industrial applications. The paper examines the effects of the scanning geometry on the point cloud's quality, focusing especially on the laser beam's incidence angle and range in relation to industrial sites. Furthermore, we looked at the feasibility of assessing a single point's quality. The measurements' received signal level steadily decreases as the incidence angle and range are raised. The paper is structured as follows: Section 2 details the techniques employed. The case study is described in Section 3. The results and discussion are presented in Section 4, and the conclusions are provided in Section 5.

2. METHODS

2.1 Scanning Geometry

The ultimate output of a scan is frequently a point cloud with n observations, each containing the 3D coordinates $[x_i, y_i, z_i]$, where $i = 1, \dots, n$ of a point in the Cartesian coordinate system. The scanner can detect the signal reflected from a surface using the TLS as the coordinate system's center. The two directional angles are then recorded: the vertical angle, or latitude, or φ_i, and the horizontal angle, or longitude, or θ_i. It calculates the distance ρ_i to the reflection's surface. The scanning geometry can impact the point cloud's quality, including the incidence angle and range. For this reason, evaluating and reducing errors and uncertainties in TLS data requires an awareness of the consideration of the scanning geometry [4]. This research defined the surface local orientation in 3D space based on point cloud geometry in a local coordinate system where the vector $P_i = [x_i, y_i, z_i]$ where $i = 1, \dots, n$, defined by a

https://doi.org/10.31436/iiumej.v25i2.3211

specific neighboring radius value [20]. The angle between the normal vector N of the surface and one laser beam vector Pi is called the incidence angle αi, as shown in Figure 1. For more information, see Eq. (1) [10,15].

$$\alpha i = cos^{-1}\left(\frac{Pi \cdot N}{|Pi||N|}\right) \quad (1)$$

The unit vector that points outward and is perpendicular to the surface's tangent plane at Pi is called the normal vector N. Consequently, the incidence angle is always in the interval [0 < αi < 90]. When a perpendicular laser beam strikes a surface, the resulting footprint is expected to be circular. Otherwise, the footprint will get more eccentric and elongated as the incidence angle increases. Their magnitude is proportional to the ellipse's major and minor axes [21].

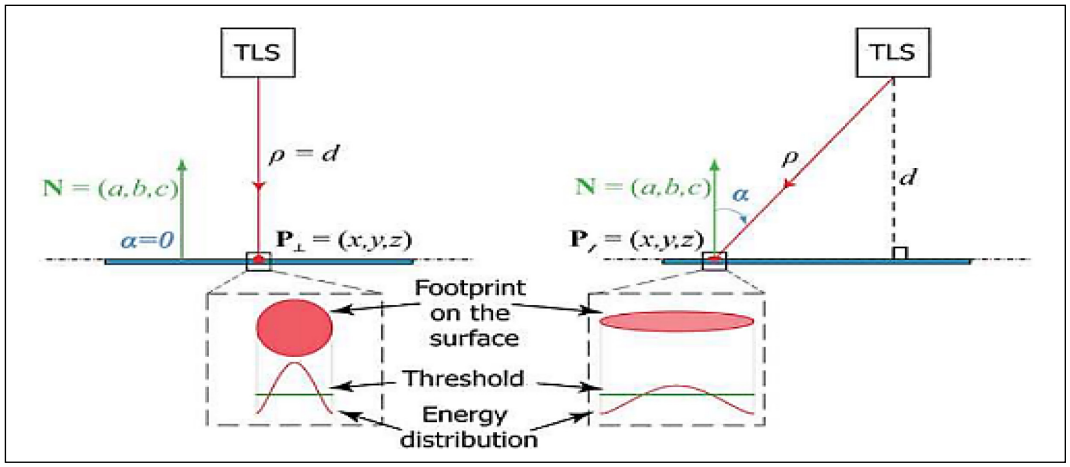


Figure 1. Diagram depicting the reflection geometry.

The TLS technology allows precise surface-to-object distance measurements by sending out laser beams in predetermined angular orientations and measuring the reflected signals in the same direction [5]. This research calculates the effect of scanning geometry depending on the radar range equation, see Eq. (2). It is defined as a function of the scanning geometry [22].

$$P_R(\rho) = P_T \frac{D_R^2 \chi \cos\alpha}{4\rho^2} \eta_{sys}\eta_{atm} \quad (2)$$

In the system settings, P_R is received power and P_T is transmitted power, and they are measured in watts, ρ is the range of the laser beam in meters, D_R is the diameter of the receiver aperture in meters, χ is the cross-section of the target, and η_{sys} and η_{atm} are the transmission factors in the system and the atmosphere, respectively. The SNR of a laser return decreases as the cosine of the incidence angle α increases, as shown in Eq. (2). Additionally, the degradation of SNR is inversely proportional to the square of the range [5].

2.2 Analyzing the Laser Point Quality

The assumption of planar surfaces illustrates the quality of the individual points that make up a point cloud. The estimation method known as total least squares [5], is utilized in this paper to acquire planar features.

2.2.1 Accuracy (Geometric) Analysis

The accuracy is an absolute value and is defined as the mean deviation of the measured coordinates from their reference values. To determine the accuracy coordinates from high-definition TLS, target scans are used with their equivalent 3D points (GCPs and CPs), which are calculated using an accurate reference technique. Geometric accuracy refers to how well

https://doi.org/10.31436/iiumej.v25i2.3211

point clouds match actual or ground-truth data. The 3D coordinate's accuracy can be expressed by Root Mean Square Error (RMSE), as stated in Eq. (3). In this equation, the variable V represents the difference between a measured value and its ground truth value. In contrast, the variable n represents the total number of observations [23–27].

$$RMSE = \pm\sqrt{\frac{\sum_{i=1}^{n} v_i^2}{n}} \quad (3)$$

2.2.2 Precision (Physical) Analysis

In a TLS system, the deviation of a sample of measurements from the mean value is a measure that can be used to determine the precision of individual measurements [28]Through its definition, the standard deviation of an LS plane can be defined as a reference value for a single measurement of noise. Various analyses were utilized to identify and verify the 3D point cloud quality by comparing them to the reference models. These methods included roughness analysis, C2C distance analysis, C2M distance analysis, and M3C2 comparison.

2.2.2.1 Roughness Analysis

Most Time of Flight (TOF) scanners are susceptible to noise (or roughness) during range collection. Point cloud roughness can be measured by computing the distance between each point and the best-fitting plane calculated to its nearest neighbors [29]. Cloud Compare (CC) software (https://www.cloudcompare.org) presents flexible tools where roughness is computed by specifying the radius of a sphere centered on individual points [19, 29].

2.2.2.2 Cloud-to-Cloud Distance (C2C) Analysis

This method is the quickest and easiest direct method for comparing three-dimensional point clouds. This method requires no data gridding, meshing, or surface normal calculation procedures. It is possible to determine the surface change most straightforwardly by calculating the distance between the two points (C2C, Figure 2A). There are two ways in which the local reference surface model can be improved: first, by employing a height function (C2C_HF, Figure 2B), and second, by using a least square fit of the closest point neighbors [30]. Cloud matching algorithms like the ICP also take advantage of this method. However, as shown in Figures 2A and 2B, the spacing of the points, outliers, and roughness of the clouds all impacted the distance measurement [19].

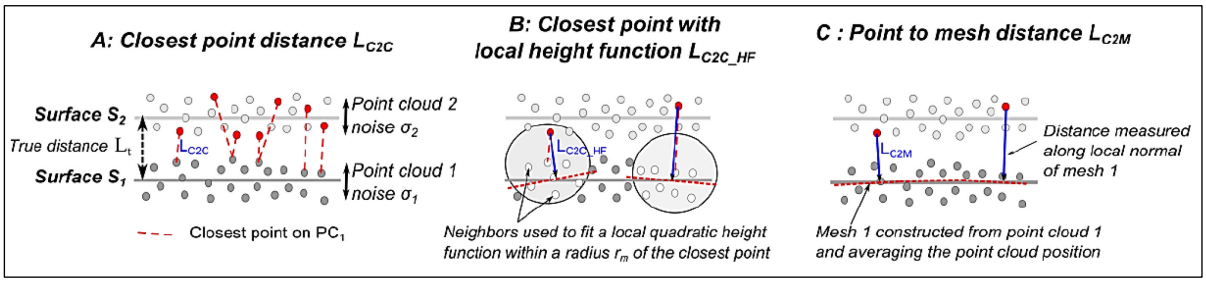


Figure 2. Current 3D comparison techniques for evaluating the similarities and differences between two data points, PC1 and PC2 [19].

2.2.2.3 Cloud to Mesh Distance (C2M) Analysis

This method is the one that is used the most commonly to analyze point cloud precision levels. The distance between a point cloud and a reference three-dimensional theoretical model or mesh determines how much surface change is detected [31,32]. This method is useful for flat surfaces because it allows one to build a mesh that corresponds to the average position of the reference point cloud (Figure 2C). On the other hand, the process of creating surface

https://doi.org/10.31436/iiumej.v25i2.3211

meshes is made more difficult either by point clouds that have a significant amount of roughness across all scales or by missing data related to occlusion. In most cases, a manual inspection that takes a lot of time has to be performed. In addition, mesh construction helps to smooth out specific details that may be significant when evaluating the properties of the local roughness [19].

2.2.2.4 Multi-scale Model to Model Cloud (M3C2) Analysis

The M3C2 approach determines the cloud to cloud distance between the two compared models. Instead of using traditional algorithms' random nearest neighbor distance, the M3C2 method calculates the distance along a local normal direction based on the surface's roughness. This method is better at dealing with point location uncertainty and registration errors. It is possible to provide a general description of the M3C2 algorithm by breaking it down into four steps, as illustrated in Figure 3. Step one begins with selecting the core points used in the reference model. The software then determines each core point's normal direction using a fitting plane under the control of a set of normal scales *D*. In the next step, the projection depth *h* and projection scale *d* of the cylinder will be modified to generate a cylinder that is aligned along the normal direction that is already known. Step three involves separating the three-dimensional model-derived sets of points inside the cylinder into their respective sets. Finally, the distance between two subsets of points is calculated using the average position length along the normal direction. Lague et al. [19] provide a more detailed description of the M3C2 algorithm for further details.

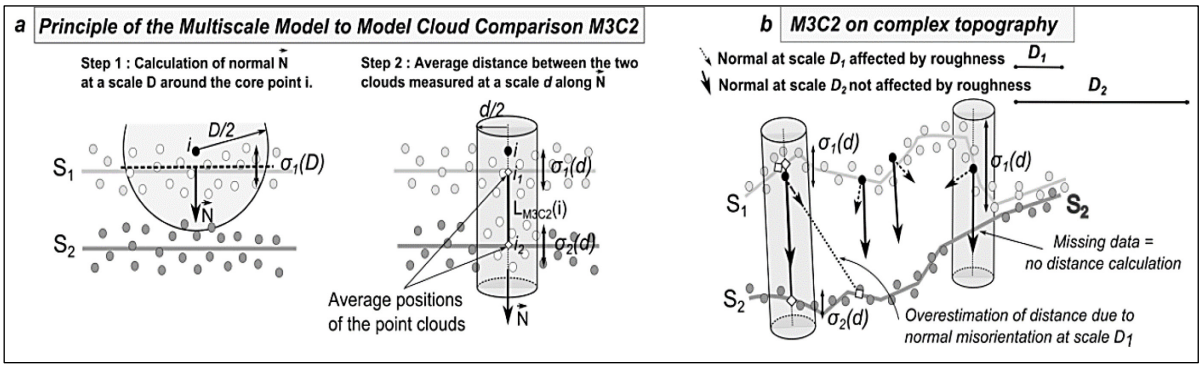


Figure 3. The M3C2 technique calculates the distance between clouds [19].

3. CASE STUDY

The quality analysis approaches described in section 2 are applied in this study using point cloud datasets obtained from the STONEX X300 terrestrial laser scanner. This device was used to collect 3D data in a controlled environment to examine the impact of scanning geometry on the resulting scan quality and analyze the quality of individual points. The scanner is a pulse-detection device that contains two integrated cameras with a resolution of 5 megapixels each. It has a range accuracy of less than 5mm at 50 meters [33]. Figure 4 shows the Area of Interest (AOI) industrial device, which is a SATO SATRONIK C3250 Plasma Cutting Machine fabricated in Germany. It is known for its intricate design and small details used for cutting thin and thick materials. This technology is widely used in industries like fabrication and metalworking for fast and accurate cutting.

https://doi.org/10.31436/iiumej.v25i2.3211

Figure 4. Plasma cutting machine case study.

The site should be planned carefully to achieve actual scanning, as well as determine the optimal scanning positions and targeting locations. The distribution of artificial targets over and surrounding the machine facilitates the alignment of the point clouds acquired from various stations. These targets were used for geo-referencing, the connection of individual scans (registration), and assessing the 3D accuracy of the coordinate measurements obtained from TLS. The trajectory positions were not spacious enough for circle scanning, even though the point clouds were obtained from fifteen scan positions about 3–4 m away from the machine.

The main goal of post-processing is to achieve clean and registered point clouds. Most processing operations for laser point clouds are carried out by JRC 3D Reconstructor 3, which is a native software provided by STONEX (https://www.stonex.it). The process of aligning data from various scans into one common coordinate system is called data registration. JRC 3D-Reconstroctor involves two types of automatic registration: pair-wise (ICP) and global (LM-ICP). LM-ICP with Bundle Adjustment was adopted to refine the alignment, as it allows the registration of all the scans and evenly distributes the registration error. Model alignment was performed before executing the C2C, C2M, and M3C2 methods, as demonstrated in Figure 5.

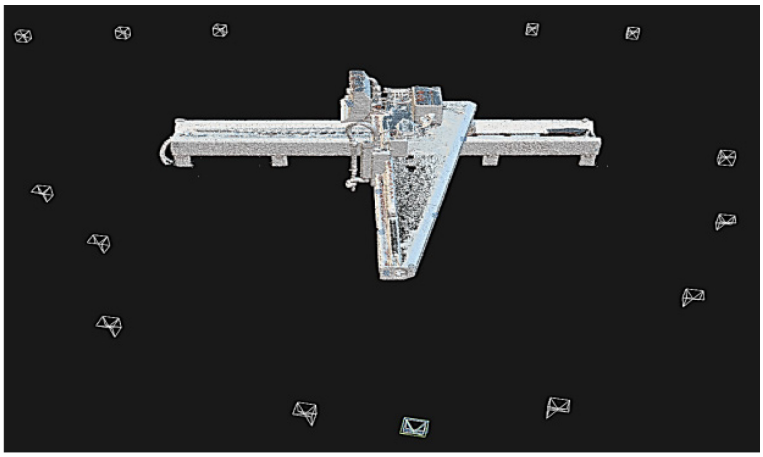


Figure 5. Registered TLS data scans of SATO SATRONIK C3250.

4. RESULTS AND DISCUSSION

4.1 Quality assessment based -geometric and physical aspects

As mentioned, fifteen stations were captured for the regular geometry point cloud. The precision assessment was applied to the plasma cutting machine case study to evaluate the laser scan measurements and eliminate outliers. According to Section 2.2.2, analysis is used to assess individual points' quality. Results are illustrated in Figures 6 and 7, where (a) refers to

https://doi.org/10.31436/iiumej.v25i2.3211

roughness analysis, (b) refers to C2C analysis, (c) refers to C2M analysis, and (d) refers to M3C2 analysis. The precision assessment was applied to a real dataset with outliers, as shown in Figure 6. As a next step, the precision assessment was applied to a dataset without outliers, see Figure 7. The quality analysis methods are carried out using the CC software. In Figure 6, the grey and red points on the bottom of the machine represent the invalid outlier point cloud, mainly because the corresponding points cannot be detected using the small projection scale in the C2C, C2M, and M3C2 methods.

CC contains several filters, including SOR filter and Noise filter, which are applied to remove noise and outliers from the data in this study. In the SOR filter, 6 points were used to estimate the mean distance, while in the noise filter, the k-D tree (k-dimensional tree) value was set at 6 points, and the radius was selected to 0.01 m to include neighbors. Compact spatial data structures like k-D trees are ideal for significant tasks like determining the location of nearby neighbors. Each node in the k-D tree is used to split up one of the k dimensions in a binary search tree. However, some outliers cause deformation even when applied above filters. Therefore, they were removed manually from the target point cloud to simulate the occlusion regions, whose roughness is shown in Figure 6 (a). Whereas the standard deviation of roughness in the datasets with outlier values was equal to 1.2 mm, noting that the local neighborhood radius was set at 0.01 m. However, it was noticed in Figure 7 (a) that the standard deviation was decreased to approximately 0.6 mm when removing the outliers due to a decrease in the local neighborhood radius of 0.005 m. The roughness values are widely dispersed as the outliers are present, according to the Gaussian roughness distribution.

According to C2C analysis, the standard deviation of datasets that include outliers was 0.077 m, Figure 6 (b). Compared with Figure 7 (b), the standard deviation dropped to 0.006 m as the outliers were removed. However, this method can be used only for calculating the relative distance between corresponding points without reflect of signed displacement. On the other hand, a reflection of the change in the direction of the point cloud can be carried out using C2M analysis, as it generates sharp triangles easily when constructing the mesh and TIN model. This is not suitable for complex scanning scenes. Figure 6 (c) and Figure 7 (c) showed the similarities of the standard deviation values for both C2M and C2C methods, which was 0.075 m in the dataset, including outliers, compared to 0.012 m without outliers. To calculate the quality, the M3C2 analysis is performed where Figure 6 (c) indicates the analysis results for the datasets with outliers, as the projection depth was 0.406 m. Obviously, the neighborhood normal vector and initial projection radius are set at 0.1 m and 0.05 m, which are determined according to the point cloud surface roughness. On the other hand, using the M3C2 method, Figure 7 (c) shows the quality analysis results of the datasets without outliers. The projection depth of the latter was 0.407 m, and the normal vector neighborhood and initial projection radius were set to 0.094 m and 0.047 m, respectively. For a comparison between Figure 6(c) and Figure 7(c), a variation has been noticed in the standard deviation with projection depth using the M3C2 according to the Gaussian distribution of roughness, specifically, at the depths 16 and 20 cm, as the maximum standard deviations were 0.006 m and 0.001 m, respectively. It can be concluded that the M3C2 analysis can be used to obtain more accurate quality results compared to C2C and C2M methods.

https://doi.org/10.31436/iiumej.v25i2.3211

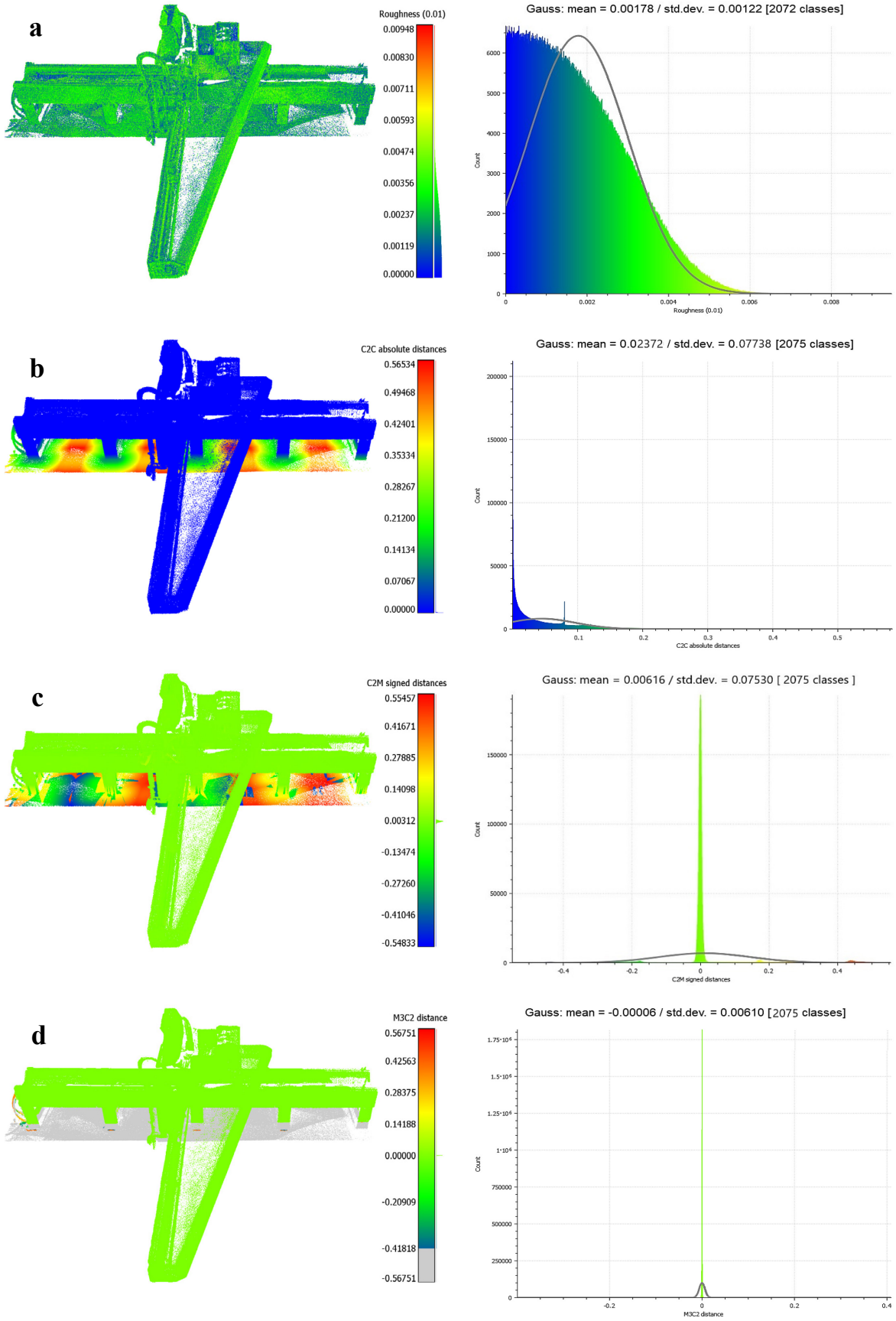


Figure 6. Real dataset of the plasma cutting machine with outliers analysis (a) Roughness, (b) C2C, (c) C2M, and (d) M3C2

https://doi.org/10.31436/iiumej.v25i2.3211

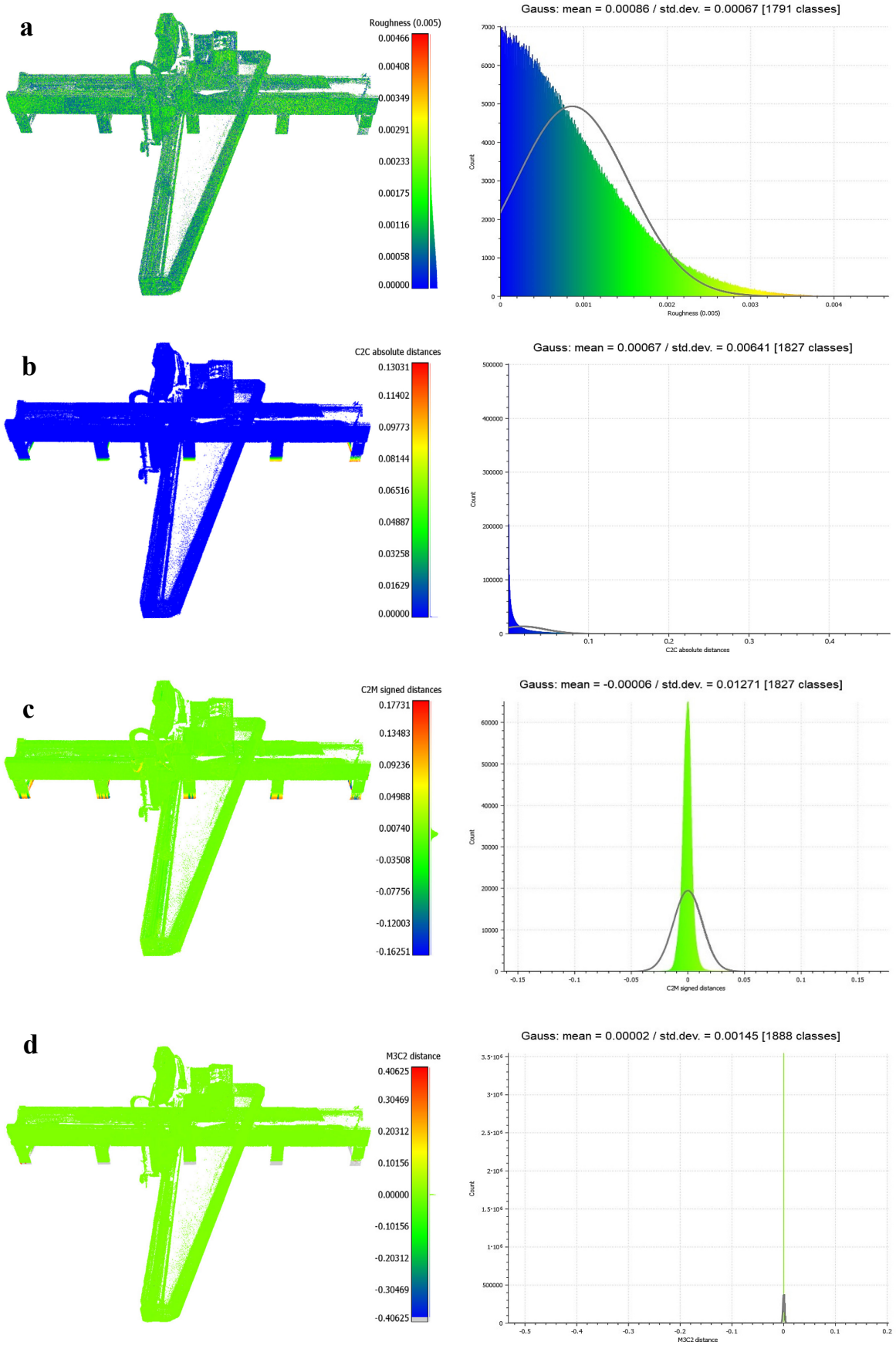


Figure 7. Real dataset of the plasma cutting machine without outliers analysis (a) Roughness, (b) C2C, (c) C2M, and (d) M3C2

https://doi.org/10.31436/iiumej.v25i2.3211

Analysis of the outcome of the results from applied algorithms has been performed to understand how outliers affect data variability, Table 1. As an advantage, this analysis contributes to a precise understanding of the underlying patterns in each parameter and directs researchers in making appropriate decisions about data pre-processing and interpretation. Depending on the results obtained from the current study, we found that the M3C2 method outperformed the C2C and C2M methods for the deformation analysis. Figure 8 illustrates the data standard deviations with and without outliers using the mentioned analysis methods. An increase in standard deviation values indicates that the involvement of outliers will increase data variability. Once outliers are removed from a dataset, the standard deviation values drop, bringing the remaining values closer to the mean. Since outliers may distort statistical metrics, researchers often eliminate outliers for better comprehension of data's underlying trends and patterns.

Table 1. Compares the real dataset before and after outlier removal in different analysis methods.

Case Study	**Real Dataset with Outliers**			**Real Dataset without Outliers**		
No. of Points	**4301853**			**3334525**		
Method	**Classes**	**Mean (m)**	**Std. Dev. (m)**	**Classes**	**Mean (m)**	**Std. Dev. (m)**
Roughness	2072	0.00178	0.00122	1791	0.00086	0.00067
C2C	2075	0.02372	0.07738	1827	0.00067	0.00640
C2M	2075	0.00616	0.07530	1827	0.00006	0.01271
M3C2	2075	0.00006	0.00610	1888	0.00002	0.00145

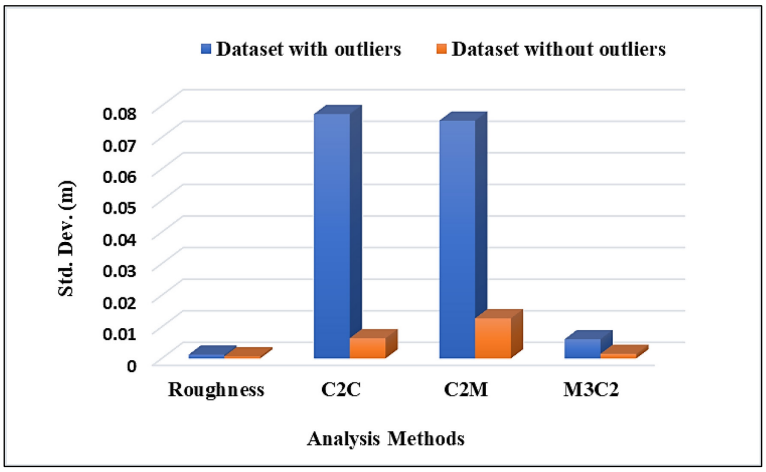


Figure 8. Compares the entire dataset standard deviations with and without the presence of outliers.

Figure 9 illustrates a selected part of the case study machine, (a) and (b) indicate the evaluation quality of TLS data with and without the presence of outliers, respectively. The choice of this part is based on containing small and complex details and estimating the effectiveness of previous evaluation analysis methods depending on the size and other details of the case study. Different results are obtained according to the analysis methods utilized and are similar to those obtained in the entire case study, as shown in Table 2. In addition to having lower highest and lowest standard deviation values, 0.008 m and 0.003 m, respectively, than other approaches, the M3C2 method has demonstrated its efficacy by producing valuable and accurate quality analysis data when compared to the C2C and C2M methods. Refer to Figure 10 for additional clarification.

https://doi.org/10.31436/iiumej.v25i2.3211

Table 2. Compares dataset of part (1) before and after outlier removal using different analysis methods.

Part 1	Real Dataset with Outliers			Real Dataset without Outliers		
No. of Points	732599			490218		
Method	Classes	Mean (m)	Std. Dev. (m)	Classes	Mean (m)	Std. Dev. (m)
Roughness	853	0.00172	0.00124	681	0.00087	0.00068
C2C	856	0.03070	0.06900	701	0.00109	0.00681
C2M	856	0.01038	0.08713	701	0.00052	0.01928
M3C2	856	0.00002	0.00820	701	0.00011	0.00335

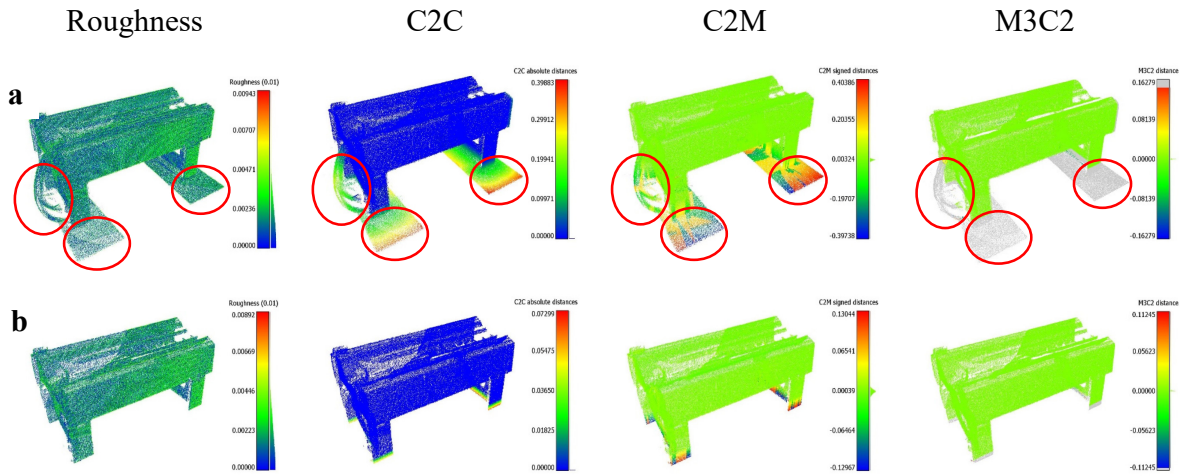


Figure 9. Dataset of the selected part from plasma cutting machine (a) with the presense of outliers, (b) without the presense of outliers, using different analysis methods.

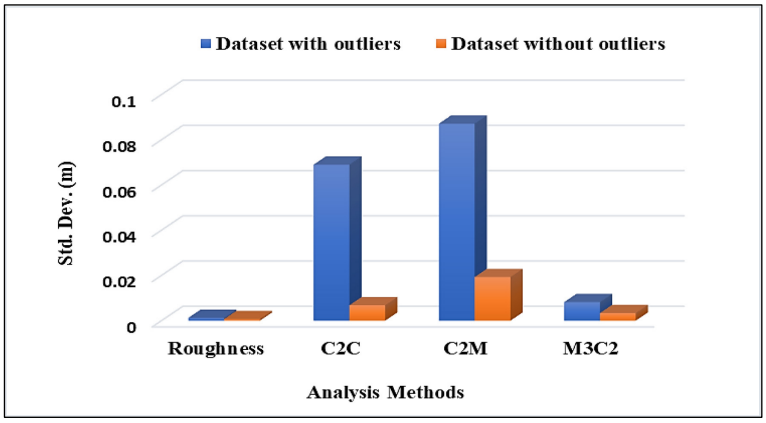


Figure 10. Compares the dataset of the selected part standard deviations with and without outliers.

On the other hand, the point clouds produced by laser scanners often contain numerous points that exhibit significant inaccuracies caused by various variables. Therefore, a geometric evaluation was used to evaluate the laser scan measurements for the plasma cutting machine case study. This was implemented by calculating the RMSE and comparing the results with the values of the associated GCPs and CPs in the laser point cloud. To ensure that TLS point clouds are geometrically accurate, we survey a selection of GCPs and CPs using a total station reference device. The RMSE values delivered from the selected GCPs and CPs were 2.9 mm and 3.2 mm, respectively, as illustrated in Table 3. We can notice that the errors were delivered

in the x direction. However, the largest error was targeted at points 3 and 6. This can be justified due to the position of points 3 and 6, which were covered by only 3 scans, while 5 scans covered other points.

4.2 Quality assessment based -scanning geometry

Two tests are carried out using data from the plasma cutting machine to examine the influence of scanning geometry parameters such as range and incidence angle on the quality of the computed point clouds. The first test was implemented to determine the effect of different ranges per point on the point cloud quality. According to the laser range equation, Eq. (2), the quantity of power received decreases as the scan range increases.

Table 3. RMSE-based control points and checkpoints.

NO.	Type	X error (m)	Y error (m)	Z error (m)	Total error (m)	RMSE
1	Control Point	0.0026	0.0027	0.0016	0.0028	0.0029
2	Control Point	0.0013	0.0011	0.0025	0.0021	
3	Control Point	0.0036	0.0021	0.0027	0.0033	
4	Control Point	0.0028	0.0017	0.0031	0.0030	
5	Check Point	0.0029	0.0021	0.0027	0.0030	0.0032
6	Check Point	0.0030	0.0025	0.0033	0.0035	
7	Check Point	0.0018	0.0017	0.0012	0.0029	

The range's impact on the overall noise level grows as the square of the range does. This outcome can be demonstrated in Figure 11. The range values are illustrated through the horizontal axis, and the standard division is illustrated through the vertical axis. Assuming a constant incidence angle, the standard deviation alters as a function of scanning distance. Compared to other ranges, the point cloud quality registered a lower standard deviation value at the 9-meter range. It has a high-precision standard deviation of 0.006 m for roughness, while M3C2 has a standard deviation of 0.007 m. The degree of detail was maintained by ensuring that all sections and intricacies of the machine were covered completely.

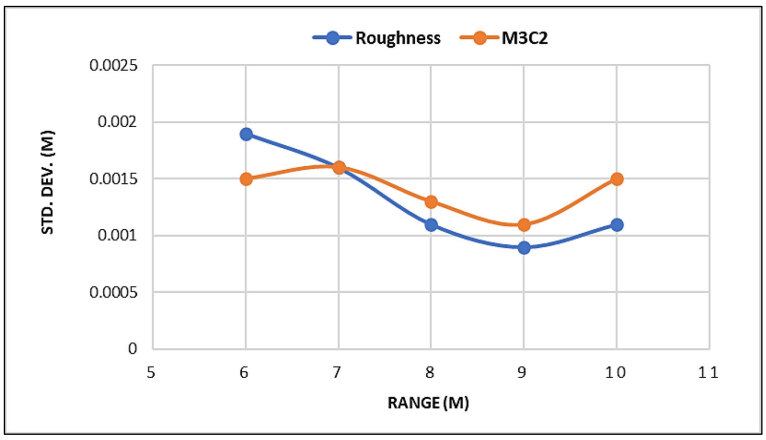


Figure 11. Experiments of range effects on the point cloud quality.

On the other hand, the second experiment was applied to examine the effect of the incidence angle on the range measurement. The scanning distance from the TLS to the experimental plasma-cutting machine was 9 m. As shown in Figure 12, the point cloud quality is observed as the laser incidence angle varies from 0° to 90° in 10° increments.

https://doi.org/10.31436/iiumej.v25i2.3211

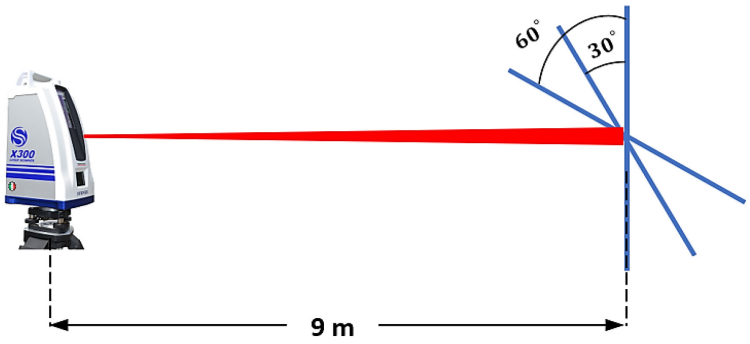


Figure 12. Measurement setup of the incidence angle experiments.

The incidence angles were calculated based on Eq. (1) and using Cyclone 3DR software. Table 4. provides information on the total number of original case study points and the percentage of points filtered at different incidence angles. As the incidence angle increases, the percentage of filtered points generally decreases. For example: At an incidence angle of 10°, 92.4% of the original points have been filtered. This trend continues, and the percentage of the filtered points decreases as the incidence angle increases. No points have been screened at a 90-degree incidence angle (0.00%). Specific data points may be deemed less reliable or credible at lower incidence angles if a higher percentage of points are filtered at those angles.

Table 4. Number and percentage of points filtered at different incidence angles.

Total No. of original case study 1883731 points	**Points have been filtered**	
Incidence angle (°)	**No. of points**	**percentage**
10	1740226	92.4%
20	1511891	80.3%
30	1384072	73.5%
40	1328672	70.5%
50	1292664	68.6%
60	1255326	66.6%
70	1143550	60.7%
80	653468	34.7%
90	0	0.00%

For further explanation, Figure 13 visually represents the relationship of filtered points with different incidence angles. The points in red represent the remaining points at any specific angle, while the remaining points of blue color will be filtered. In addition, Figure 14 illustrates a net-view analysis of the plasma cutting machine with the incidence angles and the number or density of points to demonstrate this effect in different ranges.

Finally, the effect of the incidence angle was calculated for roughness and M3C2 analysis methods. Figure 15 shows the impact of the incidence angle on the quality of the point cloud at individual points, as indicated in Eq. (2), using the estimated planar parameters per segment. As the incident angle increases, the point cloud quality diminishes when the scanning distance remains constant. The point cloud quality is relatively consistent and close together, especially when the angle falls between 0° and 60°. However, it drops if the incident angle is greater than 60 degrees. When the angle hits 80 degrees, achieving the necessary measurement accuracy becomes more difficult [10,34]. This can be explained by calculating the effect of the incidence angle on the measurement according to the range equation Eq. (2), the function cos α. As the angle of incidence increases, the deterioration of the signal increases, which it is directly

https://doi.org/10.31436/iiumej.v25i2.3211

proportional and also due to the dispersion behavior of the different surfaces with respect to the incoming light [5,35]. Figure (15) displays the standard deviation as it changes with the change of the incident angle. Figure 16 and Figure 17 also show how the point cloud quality changes as a function of incidence angle. The degradation of the measurement precision is particularly pronounced for darker-colored angles.

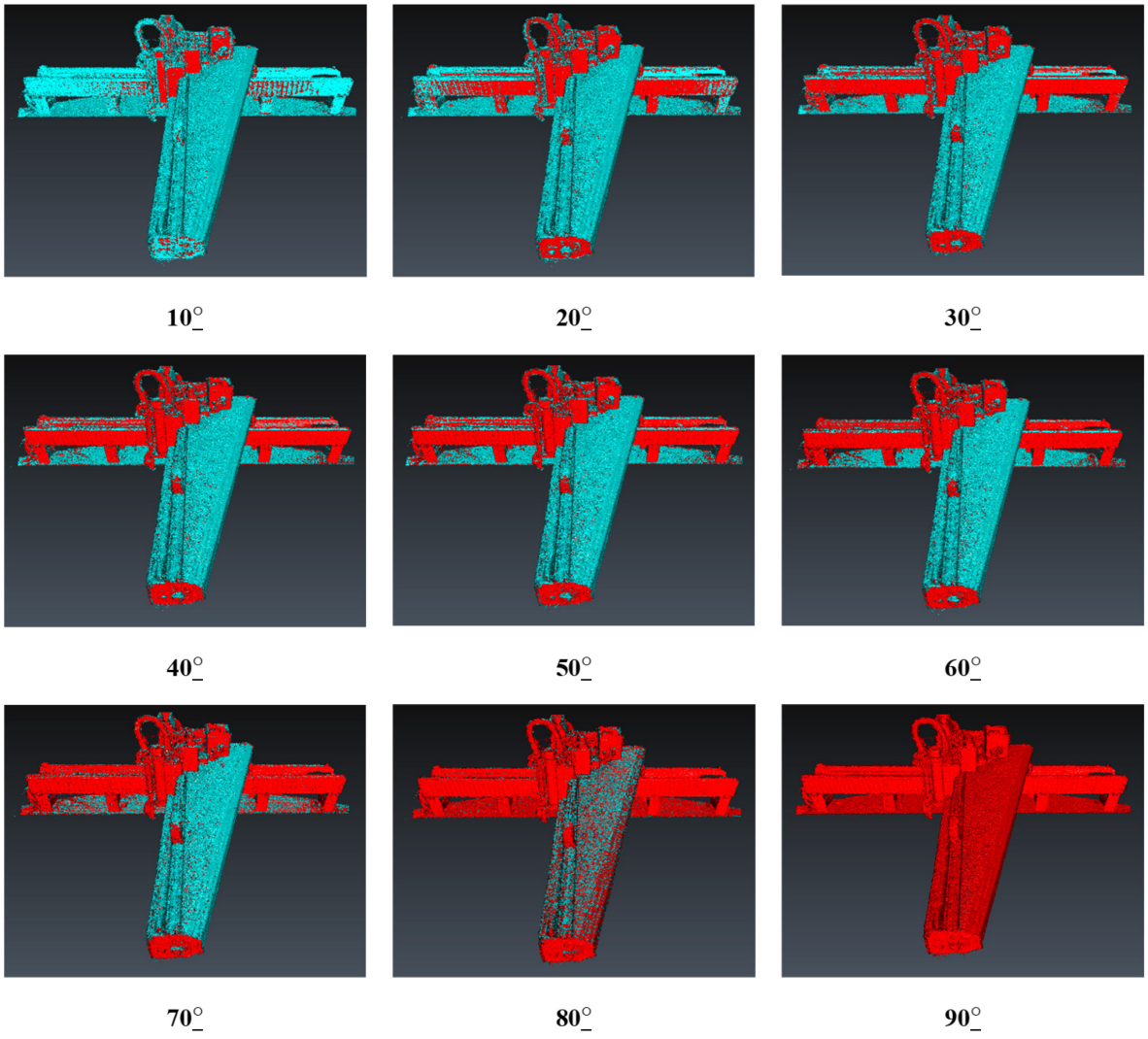


Figure 13. Point clouds of the plasma cutting machine colored with different incidence angles.

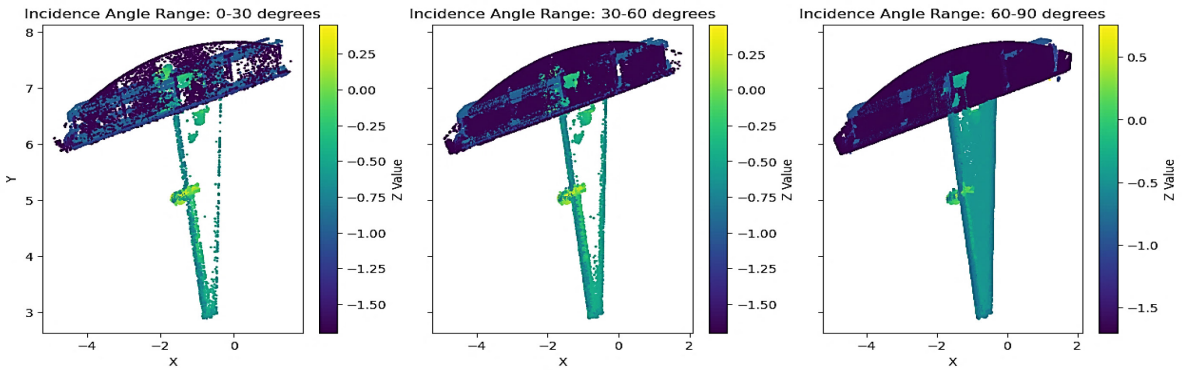


Figure 14. Net-view of the plasma cutting machine with the incidence angles.

https://doi.org/10.31436/iiumej.v25i2.3211

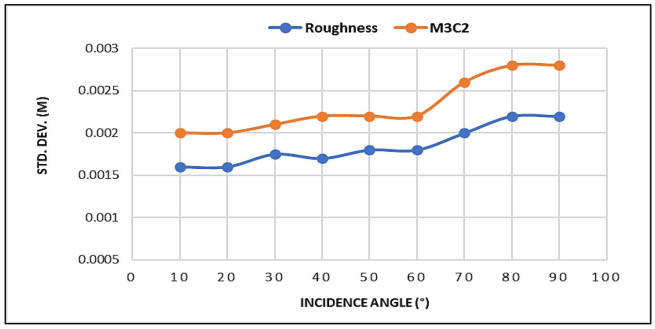


Figure 15. Experiments of incidence angles effects on the point cloud quality.

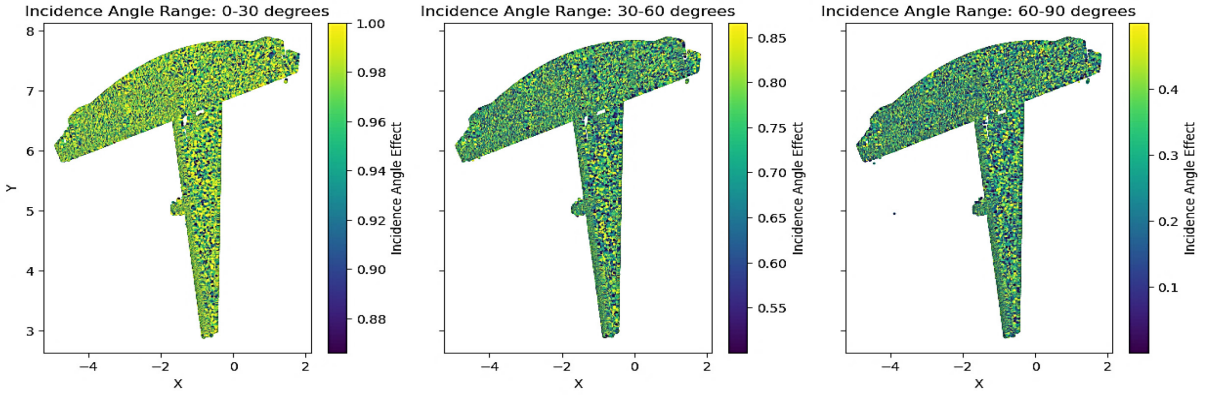


Figure 16. Net-view of the plasma cutting machine with the incidence angle effects.

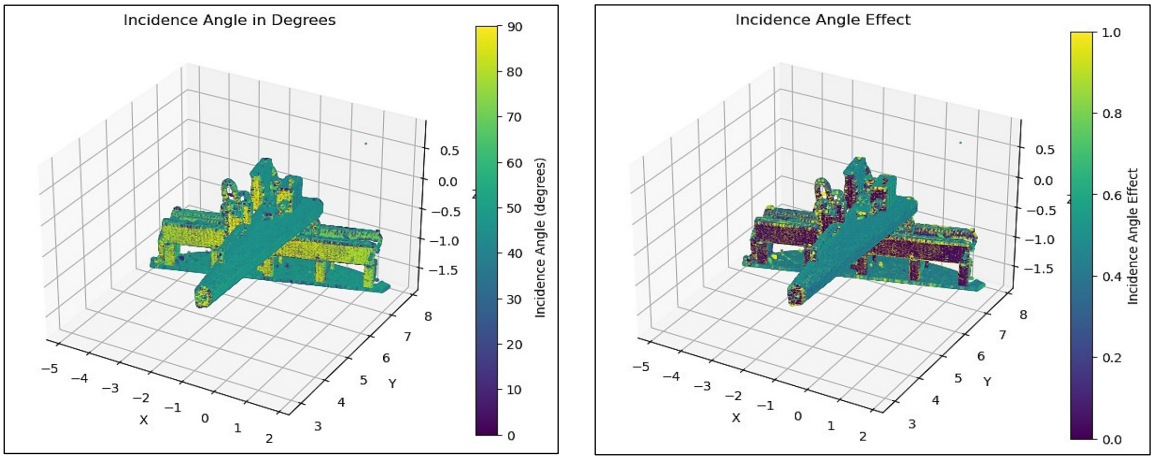


Figure 17. Three Three-dimensional net-views of the plasma cutting machine colored with (a) incidence angles and (b) incidence angle effects.

5. CONCLUSION

This paper presents a workable solution for improving laser point cloud quality in complicated situations by analyzing the impact of scanning geometry, such as incidence angle and range, for industrial applications. The paper also discusses the quality assessment of laser point clouds based on geometric and physical aspects. This includes analyzing the results of multiple statistical analyses in indoor standard conditions. Recognizing and minimizing the effects of these impacts is of the utmost importance for the 3D reverse engineering industry. The proposed method demonstrates a high degree of simplicity and accuracy for laser point cloud analysis.

Quantifying the quality of the point cloud is accomplished by a geometrical analysis of its accuracy. Using RMSE as a metric for evaluation, this study investigated the RMSE value for CPs as 0.003 meters, while for GCPs, it was 0.002 meters. Quantify analysis can be performed by precision analysis, which is also known as physical analysis. Standard deviation is used as a measurement of the evaluation. CC tools are utilized to analyze the roughness, C2C, C2M, and M3C2 approaches. Each approach has advantages and disadvantages, as illustrated by [36]. Hence, this method is robust to changes in point density and point cloud noise. Deformation analysis results proved the efficiency of the M3C2 method compared to C2C, C2M, and roughness methods [19]. The effect of outliers on data variability can be understood by comparing datasets with and without their presence, specifically focusing on the standard deviations. For each parameter's values, removing outliers and extreme values from the dataset generates a dataset more representative of the normal distribution. This approach assists scholars in making a proper decision whenever they are preparing and interpreting their data, as it provides a more detailed picture of the patterns that occurred for each parameter. Assessment of point cloud quality based on scanning distance and laser incidence angle. Regarding the effect of scanning geometry on the signal-to-noise ratio, it is possible to estimate the increased measurement noise due to increasing incidence angle and range. The findings indicate that terrestrial laser scanners with a minor beam divergence can collect high-accuracy point clouds over longer scanning distances if the laser incident angle is less than 60 degrees. However, a point cloud quality analysis must be performed before multi-station registration, as the scanning distance and angle change depending on the origin of the computed coordinates.

The point cloud quality has a substantial relationship with the criteria discussed above for accurate applications such as industrial engineering. By optimizing the incidence angles and ranges of the laser scanning data capture, it is possible to ascertain more stable data and a better-quality positioning outcome. Consequently, it will be feasible to lessen the overall error associated with the measurements using more sophisticated scenarios in future studies.

ACKNOWLEDGEMENT

The authors give their heartfelt thanks to the Head of the Photogrammetry and Laser Scanning Laboratory in the Department of Surveying Engineering, University of Baghdad, for providing the laser scanning device and related data collection tools and the needed technical support. It is also necessary to acknowledge the engineering workshop in the College of Engineering to facilitate collecting the data and working in the study area.

REFERENCES

[1] Tang P, Huber D, Akinci B, Lipman R, Lytle A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. Automation in construction, 19(7): 829-843.

[2] Teza G, Pesci A. (2013). Geometric characterization of a cylinder-shaped structure from laser scanner data: Development of an analysis tool and its use on a leaning bell tower. Journal of Cultural Heritage, 14(5): 411-423.

[3] Roca-Pardiñas J, Argüelles-Fraga R, de Asís López F, Ordóñez C. (2014). Analysis of the influence of range and angle of incidence of terrestrial laser scanning measurements on tunnel inspection. Tunnelling and underground space technology, 43: 133-139.

[4] Vosselman, G., and Maas, H. G. (2010). Airborne and Terrestrial Laser Scanning. CRC press.

[5] Soudarissanane S, Lindenbergh R, Menenti M, Teunissen P. (2011). Scanning geometry: Influencing factor on the quality of terrestrial laser scanning points. ISPRS journal of photogrammetry and remote sensing, 66(4): 389-399.

[6] Muralikrishnan B. (2021). Performance evaluation of terrestrial laser scanners—A review. Measurement Science and Technology, 32(7): 072001.

[7] Thamir ZS, Abed FM. (2020). How geometric reverse engineering techniques can conserve our heritage; a case study in Iraq using 3D laser scanning. In IOP Conference Series: Materials Science and Engineering, 737(1): 012231.
[8] Jurek T, Kuhlmann H, Holst C. (2017). Impact of spatial correlations on the surface estimation based on terrestrial laser scanning. Journal of Applied Geodesy, 11(3): 143-155.
[9] Schaer P, Skaloud J, Landtwing S, Legat K. (2007). Accuracy estimation for laser point cloud including scanning geometry. In Mobile Mapping Symposium 2007, Padova.
[10] Abed FM, Mills JP, Miller PE. (2012). Echo amplitude normalization of full-waveform airborne laser scanning data based on robust incidence angle estimation. IEEE Transactions on Geoscience and Remote Sensing, 50(7): 2910-2918.
[11] Kaasalainen S, Ahokas E, Hyyppa J, Suomalainen J. (2005). Study of surface brightness from backscattered laser intensity: Calibration of laser data. IEEE Geoscience and Remote Sensing Letters, 2(3): 255-259.
[12] Kaasalainen S, Jaakkola A, Kaasalainen M, Krooks A, Kukko A. (2011). Analysis of incidence angle and distance effects on terrestrial laser scanner intensity: Search for correction methods. Remote Sensing, 3(10): 2207-2221.
[13] Tan K, Cheng X. (2016). Correction of incidence angle and distance effects on TLS intensity data based on reference targets. Remote Sensing, 8(3): 251.
[14] Pesci A, Teza G. (2008). Effects of surface irregularities on intensity data from laser scanning: an experimental approach. Annals of Geophysics.
[15] Soudarissanane S, Lindenbergh R, Menenti M, Teunissen, P., & Delft, T. (2009). Incidence angle influence on the quality of terrestrial laser scanning points, ISPRS. In Proceedings ISPRS Workshop Laser scanning (pp. 1-2).
[16] Zámečníková M, Neuner H, Pegritz S, Sonnleitner R. (2015). Investigation on the influence of the incidence angle on the reflectorless distance measurement of a terrestrial laser scanner. Iulia Dana Negula, Violeta Poenaru, 37.
[17] Khalaf AZ, Al-Saedi ASJ. (2016). Assessment of Structure with Analytical Digital Close Range Photogrammetry. Engineering and Technology Journal, 34(11): 2140-2151.
[18] Milenković M, Pfeifer N, Glira P. (2015). Applying terrestrial laser scanning for soil surface roughness assessment. Remote sensing, 7(2): 2007-2045.
[19] Lague D, Brodu N, Leroux J. (2013). Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). ISPRS journal of photogrammetry and remote sensing, 82: 10-26.
[20] Hussein LY, Alwan IA, Ataiwe TN. (2023). Smart city 3D modeling with a total station and a smartphone. In IOP Conference Series: Earth and Environmental Science (Vol. 1129, No. 1, p. 012002).
[21] Sun W, Wang J, Yang Y, Jin F, Sun F. (2023). Accurate deformation analysis based on point position uncertainty estimation and adaptive projection point cloud comparison. Geocarto International, 2175916.
[22] Höfle B, Pfeifer N. (2007). Correction of laser scanning intensity data: Data and model-driven approaches. ISPRS journal of photogrammetry and remote sensing, 62(6): 415-433.
[23] Luhmann T, Robson S, Kyle S, Harley I. (2011). Close range photogrammetry: Principles, techniques and applications.
[24] Khalaf AZ, Ataiwe T, Mohammed I, Kareem A. (2018). 3D Digital modeling for archeology using close range photogrammetry. In MATEC Web of Conferences (Vol. 162, p. 03027). EDP Sciences.
[25] Hadi AH, Khalaf AZ. (2022). Accuracy Assessment of Establishing 3D Real Scale Model in Close-Range Photogrammetry with Digital Camera. Engineering and Technology Journal, 40(11): 1492-1509.
[26] Hussein LY, Alwan IA, Ataiwe TN. (2023). Evaluation of the accuracy of direct georeferencing of smartphones for use in some urban planning applications within smart cities. In AIP Conference Proceedings, 2793 (1).
[27] Khalaf AZ, Alwan IA, Jameel A. (2013). Establishment oF 3D Model with Digital Non-Metric Camera in Close Range Photogrammetry. Engineering and Technology Journal, 40(11): 1492-

1509
[28] Wunderlich TH, Wasmeier P, Ohlmann-Lauber J, Schäfer TH, Reidl F. (2013). Objective specifications of terrestrial laserscanners–A contribution of the geodetic laboratory at the Technische Universität München. Lehrstuhl für Geodäsie.
[29] CloudCompare forum [https://www.cloudcompare.org/forum/]
[30] Girardeau-Montaut D, Roux M, Marc R, Thibault G. (2005). Change detection on points cloud data acquired with a ground laser scanner. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 31 (8): 1601-1611.
[31] Monserrat O, Crosetto M. (2008). Deformation measurement using terrestrial laser scanning data and least squares 3D surface matching. ISPRS Journal of Photogrammetry and Remote Sensing, 63(1): 142-154.
[32] Olsen MJ, Kuester F, Chang BJ, Hutchinson TC. (2010). Terrestrial laser scanning-based structural damage assessment. Journal of Computing in Civil Engineering, 24(3): 264-272.
[33] JRC 3D Reconstructor stonex [https://www.stonex.it/project/x300-laser-scanner/]
[34] Abed FM, Mills JP, Miler PE. (2014). Calibrated full-waveform airborne laser scanning for 3D object segmentation. Remote Sensing, 6(5): 4109-4132.
[35] Lichti DD. (2008). A method to test differences between additional parameter sets with a case study in terrestrial laser scanner self-calibration stability analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 63(2): 169-180.
[36] Teng J, Shi Y, Wang H, Wu J. (2022). Review on the Research and Applications of TLS in Ground Surface and Constructions Deformation Monitoring. Sensors, 22(23): 4109-4132.

https://doi.org/10.31436/iiumej.v25i2.2769

CLASSIFICATION OF C. ANNUUM AND C. FRUSTECENTS RIPENING STAGES: HOW WELL DOES DEEP LEARNING PERFORM?

ZARINA IBRAHIM, MARSYITA HANAFI*, SITI MARIAM SHAFIE, SHARIFAH M. SYED AHMAD

Dept. of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

**Corresponding author: marsyita@upm.edu.my*

(Received: 3 March 2023; Accepted: 21 February 2024; Published online: 15 July 2024)

ABSTRACT: Chilli is one of the world's most widely grown crops. Among all of the chilli variants, C. *annuum* and C. *frustescents* are the most prevalent and consistently liked variants in Asia, where it is appreciated for its strong taste and pungency. Nevertheless, harvesting at the proper ripening stage according to their colour, size, and texture is essential to ensure the best quality, marketability, and shelf life. Currently, visual inspection is the primary method used by farmers, which is time-consuming and complicated. Even though automated chilli classification using computer vision and intelligent methods has received scholars' attention, the classification of C. *annuum* and C. *frustescents* ripening stages using deep learning models has not been extensively studied. Hence, this study aims to investigate the effectiveness of three deep learning models, namely EfficientNetB0, VGG16 and ResNet50, in classifying chilli ripening stages into unripe, ripe, and overripe classes. We also introduce a huge dataset comprising 9,022 images of C. *annuum* and C. *frustescents* chilli under various growth stages and imaging conditions which provides sufficient samples for the deep learning modelling. The experimental results show that the ResNet50 model outperforms other models with more than 95% accuracy for all classes.

ABSTRAK: Cili merupakan salah satu tanaman terbanyak ditanam di dunia. Antara semua varian cili, C. annuum dan C. frustescents adalah yang paling meluas ditanam dan merupakan varian paling pedas di Asia, kerana rasanya yang kuat. Namun begitu, penuaian pada peringkat cili matang mengikut warna, saiz dan teksturnya adalah penting bagi memastikan kualiti, kebolehpasaran dan jangka hayat terbaik. Pada masa ini, pemeriksaan visual adalah kaedah utama yang diguna pakai petani bagi memeriksa cili, tetapi ia memakan masa dan rumit. Walaupun pengelasan cili secara automatik menguna pakai kaedah komputer dan pintar mendapat perhatian sarjana, kajian tentang klasifikasi cili jenis C. annuum dan C. frustescent pada peringkat matang menggunakan model pembelajaran mendalam masih belum begitu meluas. Oleh itu, kajian ini bertujuan bagi mengkaji keberkesanan tiga model pembelajaran mendalam, iaitu EfficientNetB0, VGG16 dan ResNet50, dalam mengklasifikasi kematangan cili pada beberapa peringkat matang cili seperti belum masak, masak dan terlalu masak. Kami juga memperkenalkan set data yang besar terdiri daripada 9,022 imej cili C. annuum dan C. frustescents pada pelbagai peringkat pertumbuhan dan keadaan imej, bagi menyediakan sampel yang cukup untuk membina model pembelajaran mendalam. Hasil dapatan eksperimen mendapati model ResNet50 mengatasi model lain dengan peratusan 95% lebih tepat berbanding semua kelas.

KEYWORDS: *Transfer Learning, Deep Learning, Fruit Classification, Chilli Fruit Dataset, EfficientNetB0, VGG16, ResNet50.*

https://doi.org/10.31436/iiumej.v25i2.2769

1. INTRODUCTION

Chillies have been used as part of the human diet as spice, condiments, and vegetables for their appealing colour, flavour, and spice since the advancement of civilisation. The largest producer and exporter of chilli is India, followed by China and other countries [1-2]. Chilli is a famous cash-value crop in Malaysia and has been recorded as among the top ten crops from 2016 to 2020 [3]. This is due to its massive utilisation in the food industry for producing sauces, soups, processed meats, nibbles, candies, and soft beverages [4]. However, there is a bottleneck in chilli production due to manual sorting and grading to determine whether the chilli is in perfect condition. Chilli fruit ripening stages can be categorised according to the chilli's bioactive compound [4]. L. A. Martínez-López et al. [5], studied the ripening stages of 34,066 chilli genes at four time intervals, which were 10, 20, 40, and 60 days, respectively, after anthesis (DAA). The authors found that 10 and 20 DAA represent the early and middle stages of fruit growth, 40 DAA is when chillies reach the breaking stage (mature-green), and 60 DAA is when the fruits are fully ripened. However, the study by Olatunji & Afolayan [6] revealed that the taxonomic status of C. *frustescents* and C. *annuum* is unclear because they are morphologically related. In the experiment, the authors used the International Board for Plant Genetic Resources Descriptors for Capsicum (IPGRI,1995), in which the maturation stage classification is based on multiple traits, such as fruit colour, fruit shape, fruit surface, seed colour, and seed size. The fruit colour ranges from light yellow to green for the intermediate maturation stage and red for the mature stage. Nevertheless, the chilli fruit sorting differs according to consumer specifications and preferences. This means that chillies can be sorted according to numerous parameters, and based on these facts, the manual classification process can be time-consuming. Furthermore, individual perception and exhaustion may lead to inconsistencies in the selection and sorting decisions [7-8]. Therefore, a rapid, effective, and intelligent system is required to categorise the ripening stage in chilli harvesting.

Automated chilli fruit classification study has received significant attention in recent years. In Khuriyati et al. [9], the authors used artificial neural networks (ANN) and image processing to sort and grade red peppers. The result showed 84.46% accuracy when trained and tested using 190 and 288 images, respectively, whereas Sudianto et al. [7], used You Only Look Once (YOLO) method via transfer learning models to detect and classify the quality of A- and B-grade chillies in real-time. The authors showed that the classification accuracy for grade A chillies was higher than for grade B chillies, with 99.4% and 75.6% accuracies, respectively. The same method was applied by Abdul Manan et al. [8] to classify chilli plants and fruit (bird's eye). A comparison with two other models, Faster R-CNN and RetinaNet, has shown that the YOLO model outperformed other models by achieving 75.69% accuracy. Purwaningsih et al. [10], used a simple CNN network to classify chilli images as feasible or unsuitable. The authors showed that the network was able to produce 80% accuracy when classifying images under a controlled environment. Cruz-Domínguez et al. [11] used ANNs to recognise dried chilli peppers based on their patterns. The study produced an accuracy of only 82.13%, which may be attributed to the simple classification network. Other studies related to transfer learning-based deep learning models, such as EfficientNetB0, VGG16, and ResNet50 for fruit and vegetable classification were discussed in [12-17]. The results produced by the deep models were much better than their simple machine learning based counterparts in [9-11], but the performance of these models in classifying C. *annuum* and C. *frustescents* images under controlled and various imaging conditions into three ripening stages has yet to be explored. Furthermore, there are limited publicly available chilli images under different imaging conditions. Inspired by these facts, this study aims to investigate the efficacy of transfer learning-based EfficientNetB0, VGG16, and ResNet50 models to classify C. *annuum* and C.

https://doi.org/10.31436/iiumej.v25i2.2769

frustescents images with the aforementioned conditions. Additionally, the objective is to develop a chilli image dataset that is large enough to facilitate deep learning modelling. The findings in this paper will create opportunities for producing more robust classifiers for automated chilli classification.

This paper is organised as follows: Section 2 describes the method used in this study. This includes the detail on the compilation of the chilli fruit dataset and explanation of the three deep learning models of EfficientNetB0, VGG16, and ResNet-50. Section 2 also includes deliberation on the experimental setup for the classification. Section 3 presents the experimental results, followed by discussions. Finally, Section 4 concludes the paper.

2. METHOD

2.1. Chili Fruit Dataset

The dataset used in this study consists of 9,022 images of C. *annuum* and C. *frustescents*, of which 1496 are original chili images whereas 7526 are the augmented image counterparts. The original images were captured using an Oppo Reno4 V11.1 smartphone under various growth stages and imaging conditions, such as various chilli shapes, different distances from chilli to the camera (10cm and 15cm) and different camera settings such as ISO 100, ISO 200, white balance, 5000, and 6000. Other varying parameters considered in this study include different image sizes, scales, orientations and flipping effects, which were performed during the image augmentation process.

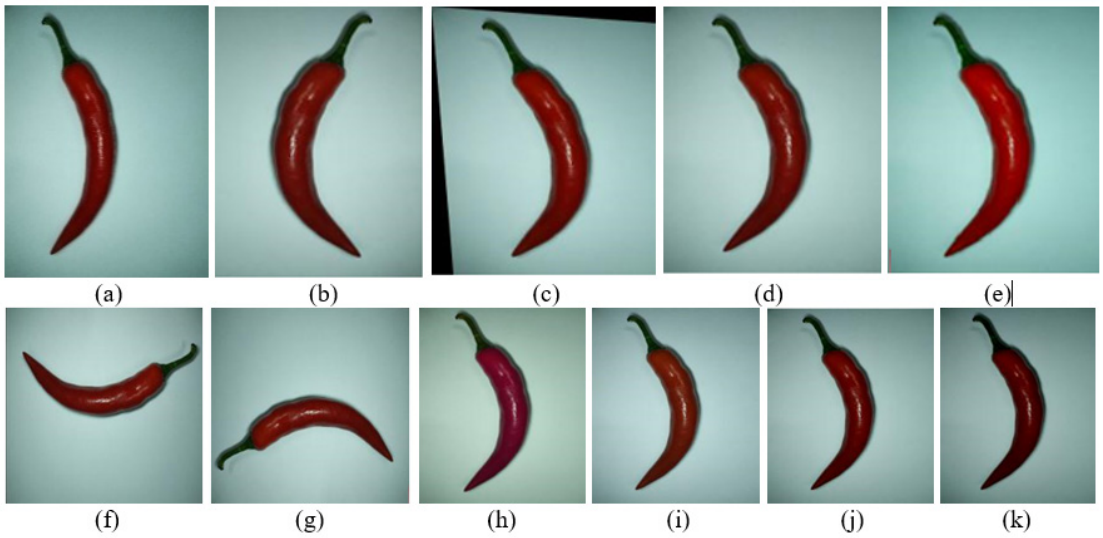


Fig. 1. Samples of ripe C. *annuum* images. (a) Original image; (b) Horizontal flipped; (c) Shear at 20% angle vertical and horizontal (20°,20°); (d) Saturation adjust to -50% (e) Saturation adjust to +50%; (f) 90° rotated clockwise; (g) 90° rotated counter-clockwise; (h) Hue adjust to -25°; (i) Hue adjust to +25°; (j) Brightness adjust to 20% brighter and (k) Brightness adjust to 20% darker.

In this study, the original chili images were resized to 224 x 224 pixels, horizontally flipped, sheared at a 20% angle, image saturated between -50% and +50%, rotated 90° clockwise and anti-clockwise, hue changed at -25° and +25°, and brightness adjusted by 20% brighter and darker, as shown in Fig. 1, using Roboflow software. The reason for producing images under various imaging conditions is to replicate actual varying scenarios and to provide sufficient samples that support deep learning modelling. The images were labelled and grouped

https://doi.org/10.31436/iiumej.v25i2.2769

into three stages of maturity, which are unripe, ripe, and overripe classes, respectively. Table 1 shows the total number of images acquired for each chilli variety.

Table 1: Chillies quantities per type

Images	Train	Test	Total
Original Dataset:			
C. *annuum*	415	103	518
C. *frustescents*	783	195	978
Augmented Dataset:			
C. *annuum*	3251	811	4062
C. *frustescents*	2771	693	3464
Combine Dataset:			
C. *annuum*	3666	914	4580
C. *frustescents*	3554	888	4442

2.2. Deep Learning Models

In this study, a transfer learning approach was applied to the three state-of-the-art deep learning models, namely EfficientNetB0, VGG16, and ResNet-50. The performance of these models in classifying C. *annuum* and C. *frustescents* ripening stages were compared. These models were selected due to their exceptional performance in classifying the maturity of other fruits and vegetables, as discussed in Miraei Ashtiani et al. [12], Duong et al. [17] and Suharjito et al. [18]. The utilisation of the transfer learning approach is expected to produce faster and better results than using the model from scratch, as discussed in Pardede et al. [16], Szyc [19] and Zhu et al. [20]. The EfficientNet was developed by Tan & Le [21] and currently consists of 8 models, ranging in quality from B0 to B7. Each version was upgraded from the previous version using a different compound coefficient, with EfficientNetB0 as the baseline. As a result, the model consistently scales the depth, width, and resolution to produce more efficient results with better accuracy. The first stage in the compound scaling approach is to find a grid to discover the relationship between the various scaling dimensions of the baseline network under a fixed resource limitation. This method determines a reasonable scaling factor for the depth, width, and resolution dimensions. Using these coefficients, the baseline network is then scaled to the desired target network [22]. The convolutional layer is the first layer in the architecture, and the remaining layers are mobile inverted bottleneck convolutional (MBConv) building blocks with squeeze-and-excitation optimisation layers, as shown in Fig. 2. The recommended input image dimension for this model is 224x224x3.

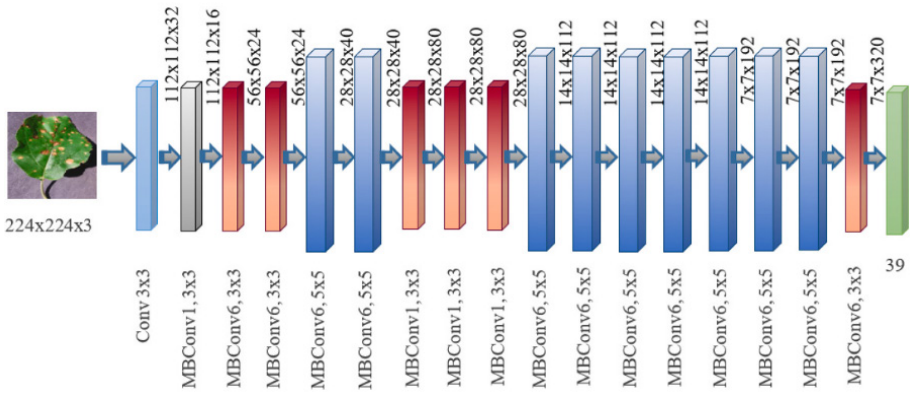


Fig. 2. EfficientNetB0 Schematic Diagram [22].

https://doi.org/10.31436/iiumej.v25i2.2769

The VGG-16 architecture, as shown in Fig. 3, comprises a combination of convolution layers, maximum pooling and fully connected (FC) layers [22]. The convolutional layers have a RELU activation function. All the information from these layers was passed to the max-pooling layer that functions to reduce the dimensions of the feature maps. Finally, the classification output was compiled by the FC layers, which are the last three layers in the architecture. These layers represent the required output classes, which for this study were the three classes of the chilli fruit ripening stages. Nevertheless, the recommended default input image size is 224x224x3.

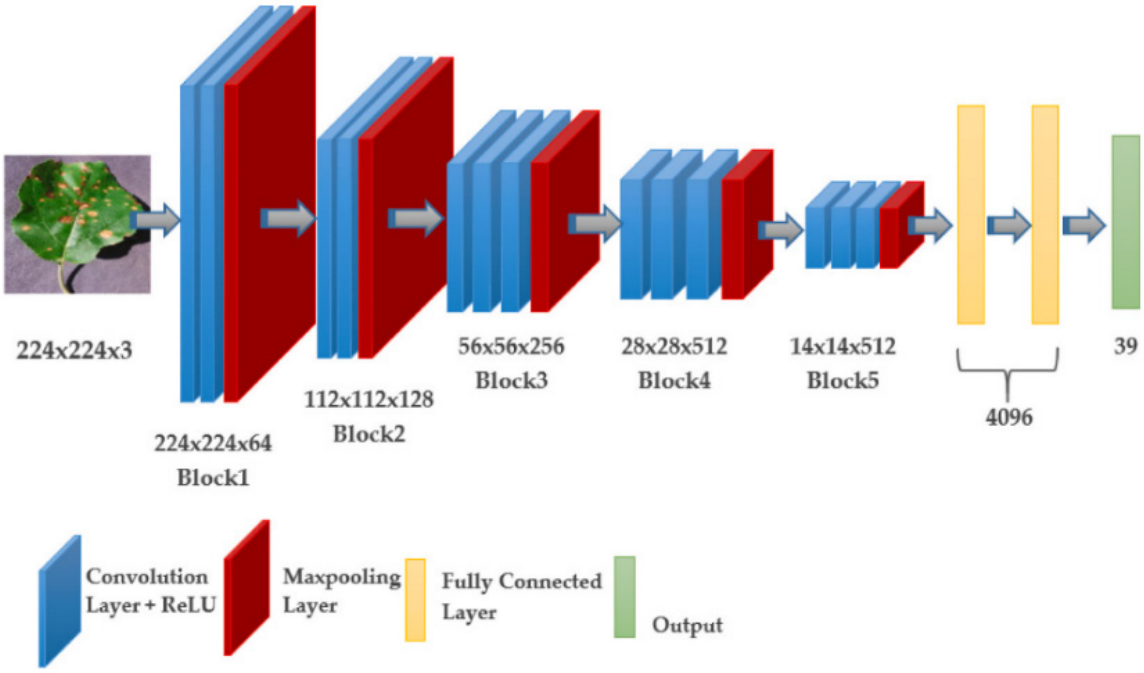


Fig. 3. VGG16 Schematic Diagram [22].

ResNet50 model is a potential remedy for the issue of numerous non-linear layers failing to learn identity mappings and degradation. The network architecture has several stacked residual units, as shown in Fig. 4. The network is constructed using residual units as building blocks and consists of convolution and pooling layers. This architecture accepts 224 x 224 pixels input images, following the recommendation in [21] and uses the same 3 x 3 kernels as the VGG16.

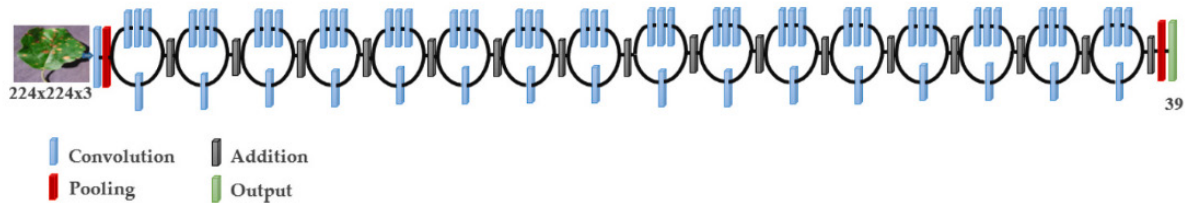


Fig. 4. ResNet50 Schematic Diagram [22].

2.3. Experimental Setup

This experiment was performed using an x64-based processor with an Intel(R) Core(TM) i5-10200H CPU operating at 2.40GHz, an NVIDIA GeForce GTX 1650 graphics card, and 8GB of RAM. These deep learning models were compiled by GPU using Jupyter Notebook alongside the Keras and TensorFlow framework. The proposed flow chart of C. *annuum* and

https://doi.org/10.31436/iiumej.v25i2.2769

C. *frustescents* ripening stages classification is shown in Fig. 5. The transfer learning used in EfficentNetB0, VGG-16, and ResNet50 models improves the classification accuracy and shortens the training time [23-24]. The hyperparameters, namely batch size, learning rate, and the number of epochs used in the experiments, were selected based on the outcome achieved during the training and the recommended values in [24-25]. For this study, the appropriate values for batch size, learning rate, and the number of epochs are 32, 0.0001 and 30 respectively. Normalisation and batch size were a consideration in the experiment setup. All deep learning models were executed using Stochastic Gradient Descent (SGD) for optimisation and softmax function for classification. Holdout cross-validation is applied in the experiment, where the entire dataset is partitioned randomly into a training set and a testing set with a distribution of 80% and 20%, respectively.

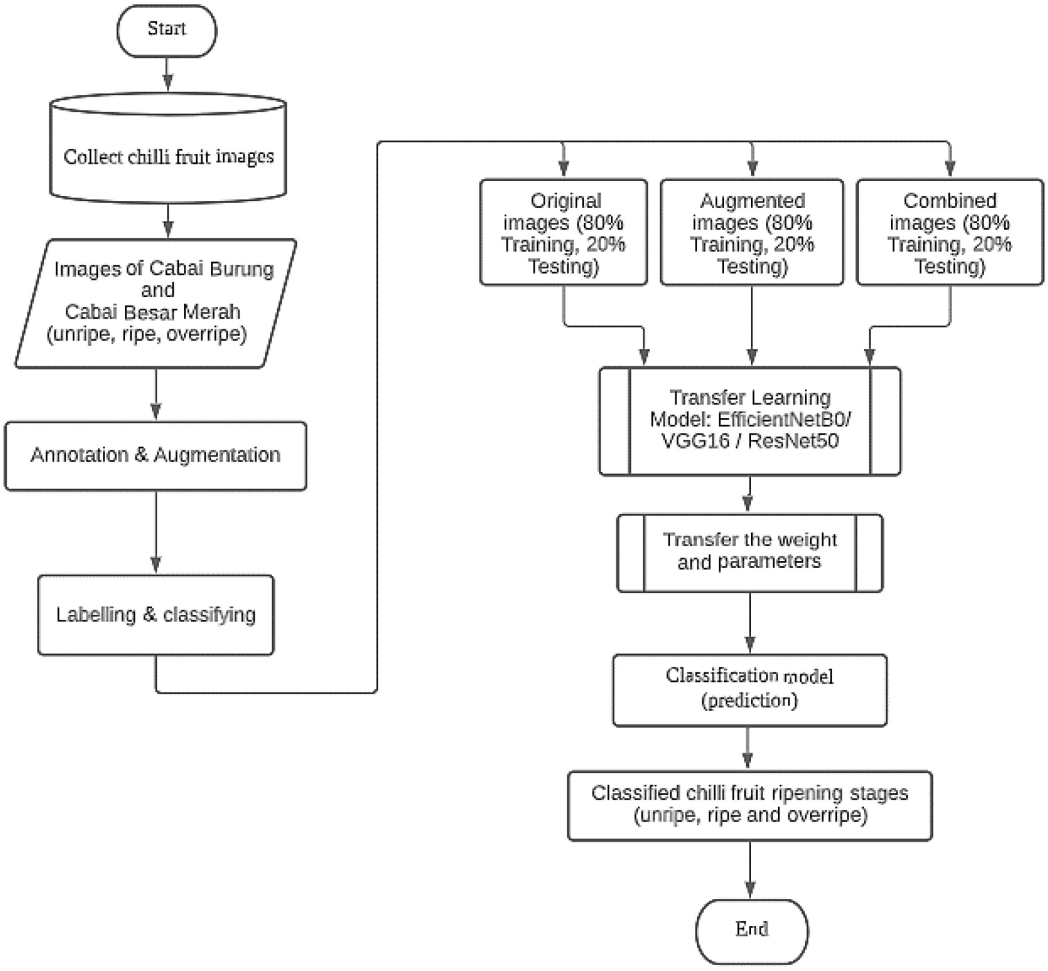


Fig. 5. Chilli Fruit Ripening Stage Classification Flow Chart.

3. RESULTS AND DISCUSSION

The performance of the models is evaluated using accuracy, precision, recall, and F1-score. Accuracy generally assesses the performance of the classification model and precision represents the accuracy of the forecasts. Recall is the true positive rate and the F1 score is a weighted harmonic mean of recall and precision, where 1.0 is the best result and 0 is the lowest. These parameters can be calculated using the values from the confusion matrices, as shown in Tables 2-4. In these tables, P.O., P.R., P.U., A.O., A.R. and A.U represent predicted overripe, predicted ripe, predicted unripe, actual overripe, actual ripe and actual unripe, respectively. The

https://doi.org/10.31436/iiumej.v25i2.2769

experiments were performed using three datasets, where the first dataset consisted of original images only, the second dataset consisted of augmented images only and the third dataset consisted of a combination of original and augmented images. All chilli fruit images were trained and tested with 30 epochs and divided randomly at a ratio of 80:20. Figs. 6-8 show that the ResNet50 model outperforms EfficentNetB0 and VGG-16 models when classifying C. *frustescents* images. The ResNet50 model produced 98% to 99% accuracy, while VGG16 had a slightly lower range of accuracy percentages. The EfficientB0 model produced a wide range of accuracy, ranging from 56% to 96%. It is shown in Fig. 7 that the EfficientNet model has poor stability compared to ResNet and VGG16 models, especially when tested using only the original images. This shows that the EfficientNet model needs more images to work well.

Table 2: Confusion matrix obtained using original images.

Variety	Model	EfficientNetB0			VGG16			ResNet50		
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *frustescents*	A,O.	0.66	0.31	0.03	0.98	0.01	0.01	1.00	0.00	0.00
	A.R.	0.18	0.74	0.08	0.03	0.97	0.00	0.03	0.88	0.09
	A.U	0.00	0.60	0.40	0.01	0.02	0.97	0.00	0.00	1.00
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *annuum*	A.O.	0.63	0.37	0.00	0.98	0.02	0.00	1.00	0.00	0.00
	A.R.	0.45	0.48	0.07	0.00	0.84	0.16	0.06	0.87	0.07
	A.U	0.20	0.60	0.20	0.00	0.40	0.60	0.00	0.10	0.90

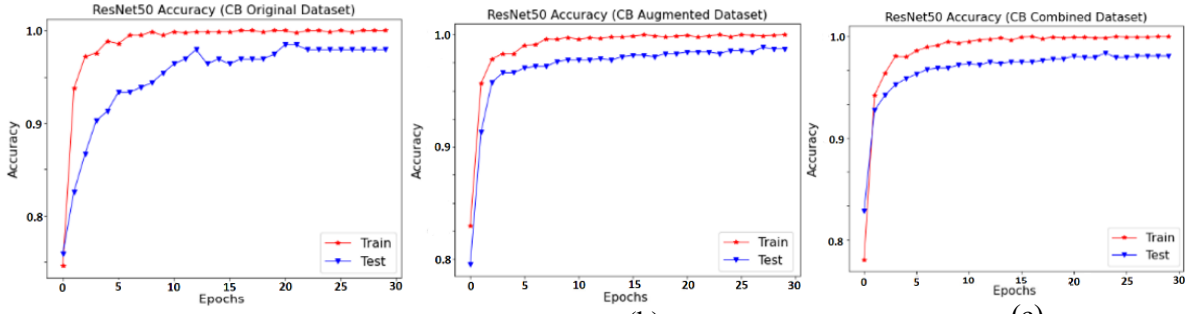
Table 3: Confusion matrix obtained using augmented images.

Variety	Model	EfficientNetB0			VGG16			ResNet50		
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *frustescents*	A.O.	0.97	0.03	0.00	0.99	0.00	0.01	1.00	0.00	0.00
	A.R.	0.03	0.93	0.04	0.00	0.98	0.02	0.00	0.97	0.03
	A.U	0.00	0.02	0.98	0.00	0.01	0.99	0.00	0.01	0.99
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *annuum*	A.O.	0.99	0.01	0.00	0.99	0.01	0.00	1.00	0.00	0.00
	A.R.	0.03	0.90	0.07	0.02	0.81	0.17	0.03	0.94	0.03
	A.U	0.00	0.05	0.95	0.00	0.08	0.92	0.00	0.04	0.96

Table 4: Confusion matrix obtained using combined images.

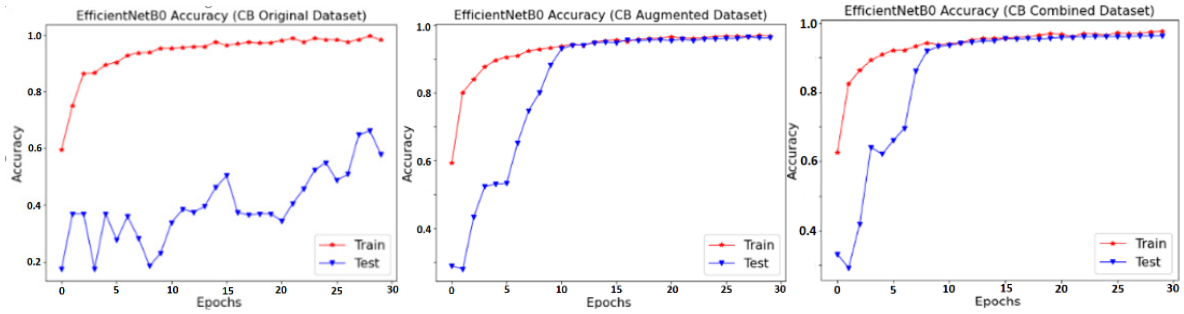
Variety	Model	EfficientNetB0			VGG16			ResNet50		
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *frustescents*	A.O.	0.99	0.01	0.00	0.97	0.02	0.01	1.00	0.00	0.00
	A.R.	0.03	0.91	0.06	0.00	0.98	0.02	0.00	0.97	0.03
	A.U	0.01	0.02	0.97	0.02	0.01	0.97	0.00	0.02	0.98
		P.O.	P.R.	P.U	P.O.	P.R.	P.U	P.O.	P.R.	P.U
C. *annuum*	A.O.	1.00	0.00	0.00	0.98	0.02	0.00	1.00	0.00	0.00
	A.R.	0.02	0.90	0.08	0.02	0.91	0.07	0.03	0.92	0.05
	A.U	0.00	0.06	0.94	0.00	0.06	0.94	0.00	0.05	0.95

https://doi.org/10.31436/iiumej.v25i2.2769



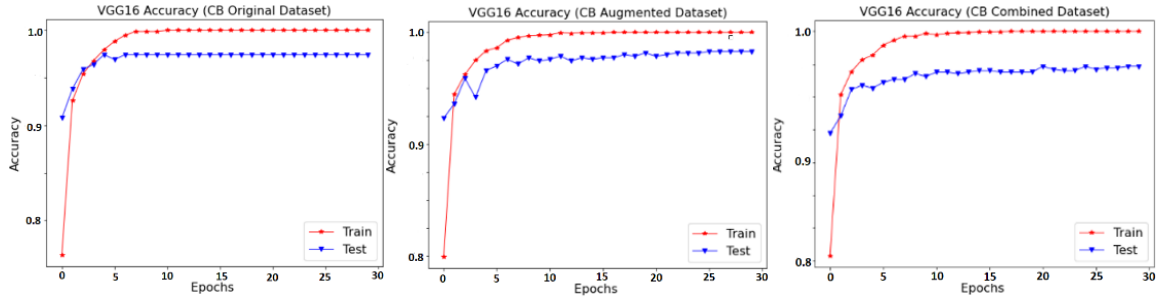
(a) (b) (c)

Fig. 6. The accuracy produced by ResNet50 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *frustescents*.



(a) (b) (c)

Fig. 7. The accuracy produced by EfficientNetB0 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *frustescents*.



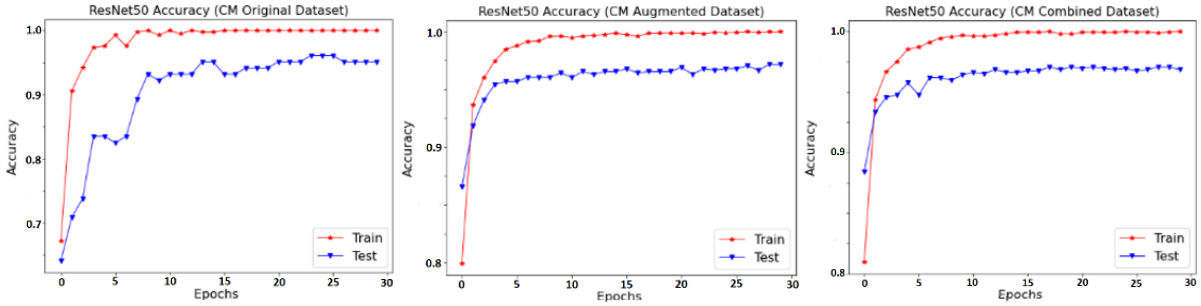
(a) (b) (c)

Fig. 8. The accuracy produced by VGG16 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *frustescents*.

The same trend was observed when classifying C. *annuum*, as shown in Fig. 9-11. The ResNet50 model produced an accuracy ranging from 95% to 97%, while VGG16 had a slightly lower range of accuracy percentages. As expected, the EfficientNetB0 model produced a wide range of accuracy, ranging from 54% to 96%, where the lowest accuracy was produced when tested with only the original images. In terms of stability, it is shown that the ResNet50 and VGG16 models are better than the EfficientNet model. However, it is observed that VGG16 failed to generalise the testing set because the testing accuracy was much lower than the

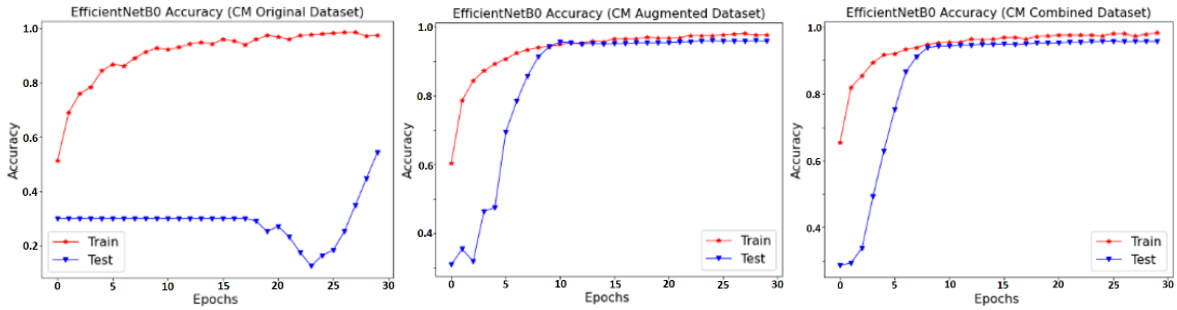
https://doi.org/10.31436/iiumej.v25i2.2769

training accuracy. The performances matrices for all deep learning models for C. *frustescents* and C. *annuum* are presented in Table 2-4.



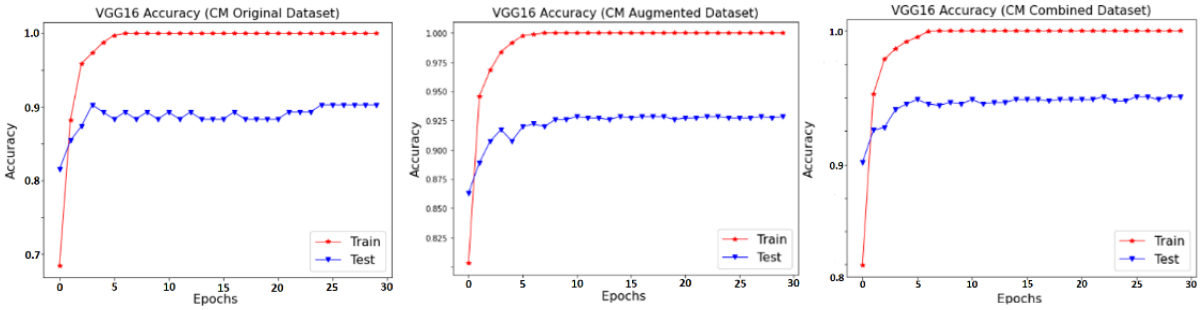
(a) (b) (c)

Fig. 9. The accuracy produced by ResNet50 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *annuum*.



(a) (b) (c)

Fig. 10. The accuracy produced by EfficientNetB0 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *annuum*.



(a) (b) (c)

Fig. 11. The accuracy produced by VGG16 model for classifying (a) original images, (b) augmented images and (c) combined images of C. *annuum*.

Tables 5-7 show that the classifiers produced higher accuracy when classifying C. frustescents than C. annuum. It is also shown that positive predictive values for C. frustescents is slightly higher than C. annuum. The same trend is observed for recall. This is because of the distinctive sizes of C. frustescents fruits used in the data collection, making them easier to be classified. The experiment results also showed that the total number of images used for classification is vital for good classification. The performance of the classifiers when

https://doi.org/10.31436/iiumej.v25i2.2769

classifying 1496 original images is poorer than classifying 7526 augmented images, especially for EfficientNetB0. This means that the data used to train EfficientNetB0 is insufficient.

Table 5: Classification performance using original images

Variety	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score	Loss
C. *frustescents*	EfficientNetB0	57.95	67.00	60.00	0.57	0.197
	VGG16	97.44	97.00	97.00	0.97	0.081
	Resnet50	97.94	98.00	96.00	0.97	0.049
C. *annuum*	EfficientNetB0	54.37	52.00	44.00	0.45	0.267
	VGG16	90.29	79.00	81.00	0.80	0.124
	Resnet50	95.14	92.00	92.00	0.92	0.071

Table 6: Classification performance using augmented images

Variety	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score	Loss
C. *frustescents*	EfficientNetB0	96.16	96.00	96.00	0.96	0.207
	VGG16	98.30	98.00	98.00	0.98	0.065
	Resnet50	98.72	99.00	99.00	0.99	0.037
C. *annuum*	EfficientNetB0	95.68	95.00	95.00	0.95	0.198
	VGG16	92.85	91.00	91.00	0.91	0.064
	Resnet50	97.16	97.00	96.00	0.97	0.046

Table 7: Classification performance using combined images

Variety	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score	Loss
C. *frustescents*	EfficientNetB0	96.33	96.00	96.00	0.96	0.188
	VGG16	97.33	97.00	97.00	0.97	0.057
	Resnet50	98.55	98.00	98.00	0.98	0.035
C. *annuum*	EfficientNetB0	95.73	95.00	95.00	0.95	0.176
	VGG16	95.08	94.00	94.00	0.94	0.052
	Resnet50	96.82	96.00	96.00	0.96	0.042

4. CONCLUSION

The efficacy of transfer learning models, namely EfficentNetB0, VGG16 and ResNet50, in classifying three stages of ripening chilli fruit for C. *frustescents* and C. *annuum* from a dataset consisting of 9,022 images in a controlled environment and under various imaging conditions is demonstrated. The experiment results show that ResNet50 is superior to the other two deep learning models. Nevertheless, the other two models, EfficientNetB0 and VGG16 also presented good results of over 90%. However, EfficientNetB0 has difficulty classifying an original dataset for both chilli fruit types. This is because the dataset used for training is small, consisting of less than 1,000 images, and also due to various illumination effects. In conclusion, transfer learning algorithms have shown the potential to be used in classifying chilli

fruit ripening stages. The performance can be improved by using a large dataset and various augmentation images for training to produce better performance.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support received from the Master of Communication Engineering and Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia. The authors would also want to extend appreciation to Mrs. Noor Azlina Abdul Aziz and Cui Hairu for their support in this study.

REFERENCES

[1] Berke TG, Shieh SC. (2001) Capsicum , chillies , paprika , bird' s eye chilli Introduction : classification and use. Handbook of Herbs and Spices, 1:111–122. http://dx.doi.org/10.1533/9781855736450.111.

[2] Chakrabarty S, Islam AKMA. (2017) Selection Criteria for Improving Yield in Chili (Capsicum annuum) . Advances in Agriculture, Volume 2017. https://doi.org/10.1155/2017/5437870.

[3] Department of Statistics Malaysia Official Portal. (n.d.). Available: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=164&bul_id=OTM1TDMzS1IvYm5mU1JiU1Fwekt3UT09&menu_id=Z0VTZGU1UHBUT1VJMFlpaXRRR0xpdz09.

[4] Vázquez-Espinosa M, Fayos O, González-De-Peredo AV, Espada-Bellido E, Ferreiro-González M, Palma M, Garcés-Claver A, Barbero GF. (2020) Changes in capsiate content in four chili pepper genotypes (capsicum spp.) at different ripening stages. Agronomy 10(9):1337. https://doi.org/10.3390/agronomy10091337.

[5] Martínez-López LA, Ochoa-Alejo N, Martínez O. (2014) Dynamics of the chili pepper transcriptome during fruit development. BMC Genomics 15:143. https://doi.org/10.1186/1471-2164-15-143.

[6] Olatunji TL, Afolayan AJ. (2019) Contributions to the classification of Capsicum annuum L. and Capsicum frutescens L. in West Africa using morphological traits. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 47(1):135–142. https://doi.org/10.15835/nbha47111204.

[7] Sudianto YH, Haristu A, Hardhienata M. (2020) Chilli quality classification using deep learning. International Conference on Computer Science and Its Application in Agriculture, 1-5. https://doi.org/10.1109/ICOSICA49951.2020.9243176.

[8] Abdul Manan AA, Mohd Razman MA, Mohd Khairuddin I, Shapiee MNA. (2020) Chili Plant Classification using Transfer Learning models through Object Detection. Mekatronika, 2(2): 23–27. https://doi.org/10.15282/mekatronika.v2i2.6743.

[9] Khuriyati N, Pamungkas AP, Pambudi AA. (2019) The Sorting and Grading of Red Chilli Peppers (Capsicum annuum L.) Using Digital Image Processing. International Journal of Agriculture and Environmental Science, 6(4):17–23. https://doi.org/10.14445/23942568/ijaes-v6i4p104.

[10] Purwaningsih T, Anjani IA, Utami PB. (2018) Convolutional Neural Networks Implementation for Chili Classification. 2018 International Symposium on Advanced Intelligent Informatics (SAIN), 190–194. https://doi.org/10.1109/SAIN.2018.8673373.

[11] Cruz-Domínguez O, Carrera-Escobedo JL, Guzmán-Valdivia CH, Ortiz-Rivera A, García-Ruiz M, Durán-Muñoz HA, Vidales-Basurto CA, Castaño VM. (2021) A novel method for dried chili pepper classification using artificial intelligence. Journal of Agriculture and Food Research 3:100099. https://doi.org/10.1016/j.jafr.2021.100099.

[12] Miraei Ashtiani SH, Javanmardi S, Jahanbanifard M, Martynenko A, Verbeek FJ. (2021) Detection of mulberry ripeness stages using deep learning models. IEEE Access, 9:100380–100394. https://doi.org/10.1109/ACCESS.2021.3096550.

[13] Saragih RE, Emanuel AWR. (2021) Banana Ripeness Classification Based on Deep Learning using Convolutional Neural Network. 3rd 2021 East Indonesia Conference on Computer and

Information Technology, EIConCIT, 85–89. https://doi.org/10.1109/EIConCIT50028.2021.9431928.
[14] Zhang Y, Lian J, Fan M, Zheng Y. (2018) Deep indicator for fine-grained classification of banana's ripening stages. Eurasip Journal on Image and Video Processing, 46(2018). https://doi.org/10.1186/s13640-018-0284-8.
[15] Foong CC, Meng GK, Tze LL. (2021) Convolutional Neural Network based Rotten Fruit Detection using ResNet50. 2021 IEEE 12th Control and System Graduate Research Colloquium, ICSGRC, 75–80. https://doi.org/10.1109/ICSGRC53186.2021.9515280.
[16] Pardede J, Sitohang B, Akbar S, Khodra ML. (2021) Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection. International Journal of Intelligent Systems and Applications, 13(2):52–61. https://doi.org/10.5815/ijisa.2021.02.04.
[17] Duong LT, Nguyen PT, Di Sipio C, Di Ruscio D. (2020). Automated fruit recognition using EfficientNet and MixNet. Computers and Electronics in Agriculture, 171:105326. https://doi.org/10.1016/J.COMPAG.2020.105326.
[18] Suharjito, Elwirehardja GN, Prayoga JS. (2021) Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches. Computers and Electronics in Agriculture, 188: 106359. https://doi.org/10.1016/j.compag.2021.106359.
[19] Szyc K. (2018). Comparison of Different Deep-Learning Methods for Image Classification. INES 2018 - IEEE 22nd International Conference on Intelligent Engineering Systems, Proceedings, 000341–000346. https://doi.org/10.1109/INES.2018.8523931.
[20] Zhu W, Braun B, Chiang LH, Romagnoli JA. (2021) Investigation of transfer learning for image classification and impact on training sample size. Chemometrics and Intelligent Laboratory Systems, 211:104269. https://doi.org/10.1016/j.chemolab.2021.104269.
[21] Tan M, Le QV. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. 36th International Conference on Machine Learning, ICML, 10691–10700.
[22] Atila Ü, Uçar M, Akyol K, Uçar E. (2021) Plant leaf disease classification using EfficientNet deep learning model. Ecological Informatics, 61:101182. https://doi.org/10.1016/j.ecoinf.2020.101182.
[23] Kumar AC, Mubarak DMN. (2021) Classification of Early Stages of Esophageal Cancer Using Transfer Learning. IRBM, 43(4):251–258. https://doi.org/10.1016/j.irbm.2021.10.003.
[24] Huang ML, Chuang TC, Liao YC. (2022). Application of transfer learning and image augmentation technology for tomato pest identification. Sustainable Computing: Informatics and Systems, 33: 100646. https://doi.org/10.1016/j.suscom.2021.100646.
[25] Masters D, Luschi C. (2018) Revisiting Small Batch Training for Deep Neural Networks. http://arxiv.org/abs/1804.07612.

https://doi.org/10.31436/iiumej.v25i2.2954

ENHANCED ANTENNA PERFORMANCE AT 3.5 GHZ WITH A COMPACT AND INTELLIGENT REFLECTING SURFACE

FIRDAUS TAQIYUDDIN AHAMAD[1], NORUN FARIHAH ABDUL MALEK[1*], FARAH SAKIINAH ROSLAN[1], NORAZLINA SAIDIN[1], MD. RAFIQUL ISLAM[1], NIDAL EL QASEM[2]

[1]Department of Electrical And Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
[2]Department of Communications and Computer Engineering, Faculty of Engineering, Al-Ahliyya Amman University, Jordan.

[*]Corresponding author: norun@iium.edu.my

(Received: 17th June 2023; Accepted: 17th June 2023; Published on-line: 17th June 2023)

ABSTRACT: Intelligent Reflecting Surface (IRS) is an upbound 5G technology capable of intelligently controlling and altering an electromagnetic (EM) wave. IRS is a planar 2D metamaterial or metasurface made up of many passive element reflecting elements connected to a smart controller, which is capable of introducing an independent phase shift and/or amplitude attenuation (collectively termed as "reflection coefficient") to the incident signal at each reflecting element. Hence, in this research, an IRS was designed to operate at 3.5 GHz structured by a compact unit cell size of 21.4 mm x 21.4 mm with Circular Patch and Ring. The metasurface consists of FR-4 substrate with a dielectric constant of 4.3 and copper backplane as the ground plane. Generally, the IRS uses a PIN diode or varactor to achieve the configurability by the ON and OFF state. However in this research, the concept is proven by connecting and disconnecting metal strips to indicate the ON and OFF state. The reflection magnitude and phase are the main parameters that were analyzed in this research. In OFF and ON states, the magnitude of the reflection coefficient is -0.32 dB and -0.38 dB respectively with dynamic reflection range of 325º. A prototype for the OFF state has been fabricated and demonstrated as a reflecting surface for a horn antenna. The measured outcome, employing the reflecting surface positioned approximately 10 cm away from the horn antenna, indicates a decrease in return loss of approximately 72.2%. The results show that the proposed reflecting surface can be used as a good reflector in IRS at 3.5 GHz.

ABSTRAK: Permukaan Pemantul Pintar (IRS) merupakan teknologi terbaru 5G yang mampu mengawal dan mengubah gelombang elektromagnetik (EM) secara pintar. IRS adalah 'bahan meta' 2D satah atau permukaan meta 2D satah yang terdiri daripada sejumlah besar elemen pemantau pasif yang bersambung dengan pengawal pintar. Ia mampu mengadakan pergeseran fasa bebas dan/atau penurunan amplitud (secara kolektif iaitu sebagai pekali refleksi) kepada isyarat insiden pada setiap unsur reflektif. Oleh itu, kajian ini adalah berkenaan IRS yang beroperasi pada 3.5 GHz dengan struktur sel kompak bersaiz 21.4 mm x 21.4 mm seunit dengan tampalan kuprum berbentuk cincin dan bulatan. Permukaan meta ini terdiri daripada substrat FR-4 dengan pemalar dielektrik 4.3 dan satah kuprum di bahagian belakang. Secara umum, IRS menggunakan diod PIN atau varaktor bagi mencapai keboleh konfigurasi mengikut keadaan BERSAMBUNG dan TIDAK. Walau bagaimanapun, konsep ini dibuktikan dengan menyambung dan memutuskan jalur logam bagi menunjukkan keadaan BERSAMBUNG dan TIDAK. Magnitud pantulan dan fasa pekali merupakan parameter utama yang dikaji dalam kajian ini. Dalam keadaan TIDAK dan BERSAMBUNG, magnitud

https://doi.org/10.31436/iiumej.v25i2.2954

pekali pantulan ialah -0.32 dB dan -0.38 dB masing-masing dengan julat pantulan dinamik 325º. Prototaip pada keadaan TIDAK telah dibentuk dan menunjukkan sebagai permukaan pantulan bagi antena jenis tanduk. Dapatan hasil menunjukkan permukaan reflektif yang diukur pada jarak 10 cm dari antena tanduk mengalami penurunan kehilangan refleksi sebanyak 72.2%. Ini menunjukkan permukaan reflektif yang dicadangkan dapat digunakan sebagai reflektor IRS yang baik pada frekuensi 3.5 GHz.

KEYWORDS: ***reconfigurable intelligent surface, 5G, metasurface, Circular Patch and Ring (CPR), unit cell, measurement***

1. INTRODUCTION

In recent years, there has been an increasing trend of research and development for enhancement and improvement in wireless communication systems. This surge is propelled by the increasing number of communication devices and technologies that demand more robust wireless connections. Conventional wireless technologies are inadequate to meet the requirements of upcoming technologies, necessitating exceptionally high data rates, reduced interference, minimized signal attenuation, energy efficiency, and pervasive connectivity. These advanced technologies encompass Internet of Things, Virtual Reality (VR), Augmented Reality (AR), Artificial Intelligence (AI), smart cities, and Autonomous Vehicles (AVs). Consequently, an Intelligent Reflecting Surface (IRS) or so called Reconfigurable Intelligent Surface (RIS) [1] has emerged as a promising new technology anticipated to revolutionize wireless communication especially for the upcoming sixth generation.

Generally, IRS is a smart technology made of 2-dimensional (2D) metamaterials or metasurfaces having the ability to control the propagation of signals in the wireless communication channel. Hence, it potentially could improve spectral efficiency (SE) and energy efficiency (EE) to assist the sixth-generation (6G) and subsequent generations in providing the necessary data rate to assist current and forthcoming technologies [2].

A 2D metamaterial or metasurface consists of negative permeability and permittivity, categorizing it within quadrant 3 of materials, also known as Negative Index Material (NIM) [3]. Within the metasurface concept, the surface can be adjusted to alter the phase of individual reflecting elements, termed unit cells, thereby enabling signal beamforming and directing the signal toward a specific direction. Moreover, the technology behind 2D metamaterials is engineered to function across different frequencies, making it highly efficient for various technologies that operate within a range of frequencies [2].

A metasurface consists of many small reflecting elements called meta-atoms, meta-cells, or unit cells [4]. The unit cells consist of three layers where the middle layer is the dielectric substrate such as Roger RO4350B, FR4, and others while the upper layer consists of a copper patch and bottom part as copper ground are commonly used as the material. Electromagnetic (EM) properties of the unit cells are influenced by structural features of the metasurface such as geometry (split-ring, circular or hexagon), size, orientation, and arrangement [5].

The objective of controlling the electromagnetic wave through IRS can be achieved by modifying the reflection coefficient of individual unit cells. This can be accomplished using various methods, such as mechanical adjustments (rotation and translation), incorporating functional materials (like liquid crystal and graphene), and employing electronic devices (such as PIN diodes and Varactor diodes) [6].

Recent research conducted by [1] involved the design and prototype of an RIS working at 3.5 GHz, comprising 2430 unit cells equipped with varactor diodes for tuning purposes.

https://doi.org/10.31436/iiumej.v25i2.2954

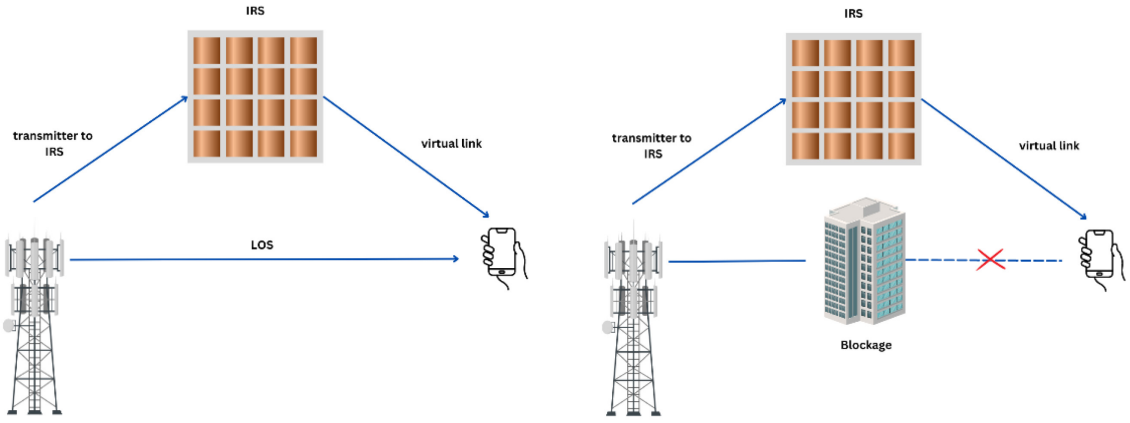
Additionally, [7] proposed achieving configurability in the IRS by integrating PIN diodes into the metasurface. Their design operates at 6 GHz, demonstrating a reflection loss of approximately 0.34 dB. Another IRS model detailed in [8] consisted of 160 meta-cells functioning at 5.8 GHz. This model was proposed and then fabricated to evaluate its performance in an anechoic chamber. By utilizing the PIN diode as an RF switch, the system indicates an almost 180º phase difference across the operating band and incurs an approximate loss of 1.5 dB.

This paper introduces a metasurface employing a circular patch with ring unit cells designed to operate at 3.5 GHz. The selection of the 3.5 GHz frequency is due to medium band frequency for 5G. Furthermore, fabricating an IRS at 3.5 GHz is deemed practical, considering the availability of materials and measurement equipment tailored for 5G applications. This work is a continuity from [9], where the author illustrates the ON state by incorporating metal strips connected to the cells, whereas the strips are omitted for the OFF state of the IRS. This research explores the concept of modeling IRS without utilizing electronic tuning mechanisms for ease of configurations. Section 2 provides a review of existing works. Section 3 delves into the metasurface unit cell, array design, and measurement. Section 4 is dedicated to the analysis of results, while the conclusion is succinctly summarized at the article's end.

2. EXISTING RESEARCH ON IRS

2.1 Fundamental Theory of Intelligent Reflecting Surface

In a conventional wireless communication context, multi-path propagation arises, resulting in significant distortions to the receiving signals. This effect, known as fading, is a significant limitation in future wireless communication systems [5]. IRS is deemed to be a viable solution that can effectively boost the performance of wireless communication networks by smartly reconfiguring the channel environment with numerous passive reflecting features. Fig. 1(a) shows the function of IRS to improve the diversity of an LoS communication link by generating a replica copy [2].



(a) (b)

Fig. 1. Presence of (a) line-of-sight link and (b) blocking element and absence of LoS link.

Fig. 1(b) depicts the scenario of the Non-Line-of-Sight (NLoS) communication link between the BS and EU. IRS serves as a virtual source to create NLoS communication link due to the blockage by some blocking elements including signboards, cars, and high rise buildings

https://doi.org/10.31436/iiumej.v25i2.2954

that can be found in most of the urban areas that have very high population density of people and infrastructure [2].

By controlling the channel environment, a specific metric objective, including the system workable rate or coverage, can be enhanced using the proper IRS phase shifts, amplitude, or both. This is fundamentally distinct from conventional wireless communication research, where optimization and design options are limited to a pair of transceivers [6].

2.2 Existing Research on Reflecting Element using Metamaterial

The metasurface consists of many small reflecting elements called meta-atoms, meta-cells, or unit cells. The term of unit cell will be used in this research. The unit cell consists of three layers where the middle layer is the dielectric substrate such as Roger RO4350B , FR4, and others while the upper layer is a patch and the bottom part, called ground, is commonly made of copper as the material.

The aim for the control of electromagnetic waves of IRS can be idealized by altering the reflection coefficient of each unit cell [10]. The authors in [11] proposed a design of unit cell with a PIN diode as the tuning element to alter the capacitance of each unit cell. The concept of this design is to control the electromagnetic response by changing the state of an element by biasing the voltage. The operating frequency of the proposed design is in the W-band at 100GHz.

The RIS work by [10] proposes a phase-shifting unit cell using PIN diodes to exhibit 8 discrete phase states at 5G mid-band. The substrate used in this design is F4B (lossy) with relative permittivity of 2.65. The authors utilized S-parameter data extracted from PIN diodes, revealing eight different phase states and an average reflection loss of 1 dB. Another work by [11] proposes an IRS in THz band (80-110 GHz) with configurable gain. The material used for the substrate is Taconic Company RF-43, having relative permittivity of $\varepsilon r = 4.3$, The results show that an average of 3.7 dB for amplitude and 300 degrees for phase have been achieved for reflected waves with a steering range of up to ∓ 84 for azimuth or elevation.

The existing IRS [1, 7, 8] show the complexity in fabricating the metasurface due to the very complicated structure and expensive external control board. Most of the research implements PIN diodes, varactors, and functional material (graphene & liquid crystal). Our proposed reflecting surface structure is a passive structure that is easy to fabricate, low cost, and requires no external complicated controller. The difference of our work in comparison with existing research is the representation of PIN Diodes using metal strips as an equivalent circuit in order to simplify the structure for the IRS.

2.3 Design and Working Principles of IRS

An Intelligent Reflecting Surface consists of three layers which can be briefly described as below:

- The first layer: Consists of many reconfigurable metallic patches and the main objective is to manipulate the electromagnetic wave of incident signals. The metallic patches are etched with a dielectric substrate.
- Middle layer: A layer of copper plate is used to avoid and reduce signal leaking when the reflection occurs.
- Third layer/ bottom layer: This layer is comprised of a control circuit board (such as FPGA [2]) that can manipulate the reflection phase and amplitude on a real-time basis.

In the design of IRS, there are some proposed ideas as the mechanism to tune the element in order to dynamically cater to the wireless channel [6]. The approaches can be described in Table 1. In addition, Table 2 summarizes the related works of the IRS, including the theory, design, and historical technologies.

Table 1. List of tuning mechanisms [6].

Mechanism	Description
Mechanical attenuation	Mechanical rotation, Mechanical translation
Functional material	Liquid crystal, Graphene
Electronic device	Positive-intrinsic-negative (PIN) diode, Varactor diode, Field-effect transistors, Micro-electromechanical systems (MEMS) switches

Table 2. List of IRS-related works.

References	Description
[2]	Discussion of the previous technology related to the concept of Intelligent Reflecting Surface. Further explanation of system and the architecture of IRS and IRS-aided wireless communication
[4]	Discussion of Reconfigurable Intelligent Surface (RIS) for the frequencies below 10 MHz. Theory of reflect array antenna and metasurface in IRS application.
[5]	Discussion of 6G wireless network associated with IRS technology. Implementation of IRS comparison between other up bound technologies including relay and backscatter communication.
[6]	Discussion on state-of-the-art of IRS and the fundamental of IRS technology.
[10]	Design of 2.75-bit reflecting unit cell of metasurface
[11]	Design of unit cell of metasurface with tunable element (PIN diode)
[12]	Discussion on RIS for 5G and beyond wireless communication. Analyze the principle of metamaterial and application. The integrating metasurface concept with other technologies.
[13]	Simulation and fabrication of an active metasurface based on PIN diodes that exhibit high absorption, reflection, and transmission in a single design.
[14]	Liquid-crystal based metasurface based real time configuration to achieve beamforming
[15]	Prototype of IRS using PIN diode as tuning element.
[16]	Theory of IRS and prototype of RIS.

3. METASURFACE UNIT CELL AND ARRAY

This section describes the simulation and fabrication process of metasurface unit cells and arrays. In addition, the measurement process has been performed using a 1-18 GHz wideband horn antenna (OBH-10180) in an anechoic chamber, MJIIT.

3.1. Unit Cell of Metasurface

Designing a metasurface requires the selection of suitable parameters based on the operating frequency. Fig. 2 shows the proposed unit cells which have been modeled into 2 different sizes. The unit cell boundaries along the X and Y axes delineate an infinite array of the metasurface. The Z-axis is designated as "Open," indicating that the metasurface is exposed and unbounded in the Z direction, both from the top and bottom. Waves are incident upon the metasurface from the Z direction.

Model 1 exhibits a unit cell of size 42 mm x 42 mm, while Model 2 showcases a reduced unit cell's size of 21.4 mm x 21.44 mm. The size of the unit cells of Model 1 and Model 2 are calculated based on equations (1) and (2), respectively [17] and shown in Fig. 2(a). Notably, Model 2's dimensions are halved in comparison to the initial design, Model 1. The substrate

https://doi.org/10.31436/iiumej.v25i2.2954

used for the metasurface is FR4 with a thickness of 1.5 mm. The dimension of parameters for both models has been obtained through numerous simulations (trial and error) and summarized in Table 3 (as shown in Fig. 2(b)).

$$Length, L_1 = \frac{\lambda}{2} \quad (1)$$

$$Length, L_2 = \frac{\lambda}{4} \quad (2)$$

where L_1 and L_2 are the lengths of each side of the unit cell for models 1 and 2, respectively, and λ is the wavelength at the operating frequency.

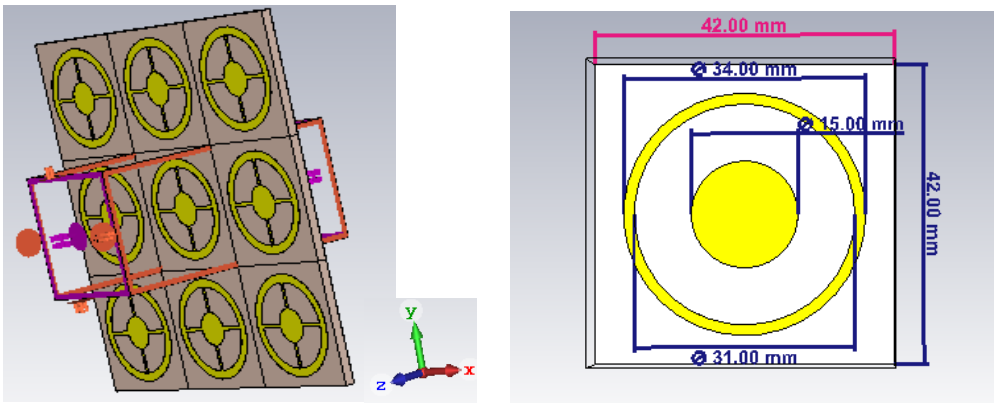


Fig. 2. The implementation of (a) Unit cells structure for Model 1 and 2, (b) Unit cell's dimension based on Table 3.

Table 3: Dimension of Unit Cell for Each Model.

Model	Dimension	Outer ring radius [mm]	Inner ring radius [mm]	Patch radius [mm]	Metal strip width [mm]
1	42mm × 42mm	17	15.5	7.5	2
2	21.4mm × 21.4mm	9	7.5	3	0.6

3.2 Array Design

The unit cell model indicates that the array size is infinite, which is impractical for fabrication. Therefore, an array size of 5 × 5 has been used in this analysis. The array, with an overall size of 107 mm × 107 mm, has been designed to accommodate both ON and OFF states. The selected size ensures that the unit cell is centrally positioned within the array. Additionally, the decision to use a 5 × 5 configuration is influenced by the observation that, as cell sizes increase, the array's performance closely aligns with that of a single unit cell. Fig. 3(a and b) depict the array in both the OFF and ON states.

https://doi.org/10.31436/iiumej.v25i2.2954

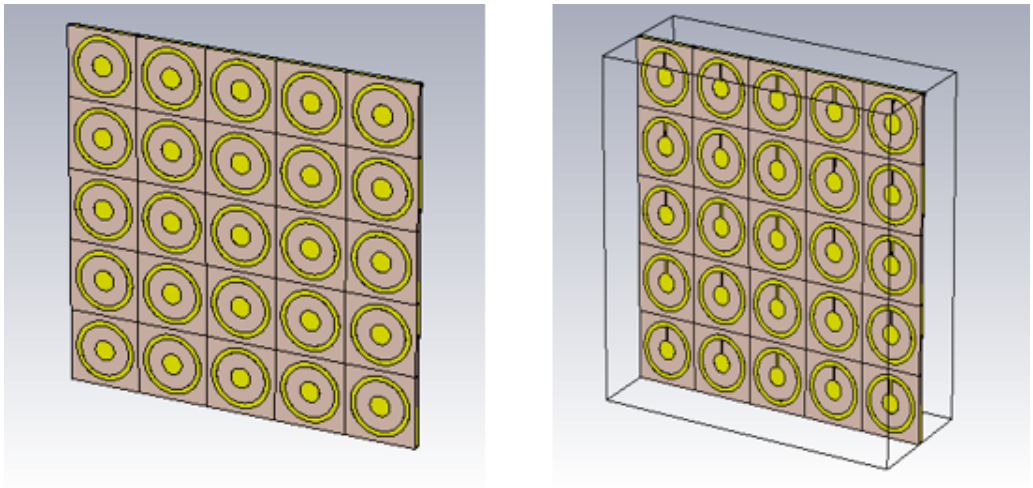


Fig. 3. 5 x 5 array for (a) OFF state (b) ON state '1000'

3.3 ON/OFF States of Metasurface

In order to demonstrate the tuning phase for the reflecting surface, the authors have placed 4 metal strips horizontally and vertically for tuning 'ON/OFF' or '1/0' states of the metal strips. The substitution of the PIN diode with a metal strip has been documented in numerous articles [18-20]. The metal strip is characterized as either an open or a short in the transmission line, as shown in Fig. 4. This configuration is more straightforward to simulate compared to the diode, resulting in faster and more accurate simulations. Additionally, the diode, being an active element, necessitates a complex bias circuit to regulate voltage and restrict current flow into the microwave circuits. The comparison between the simulated metal strip and the PIN diode, as noted in [18], reveals approximately a 6dB loss, which remains within the acceptable range.

As a result, 16 designs of metasurfaces have been investigated due to the implementation of metal strips. The combinations are 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110 and 1111. The '0' indicates the metal strip has been removed while '1' indicates the metal strip is connected. The most significant bit defines the Metal Strip 4 (MS 4) while least significant bit represents Metal Strip 1 (MS 1). Illustration of the metal strips position at the unit cell is shown in Fig. 5. The 0000 configuration is fixed to be the OFF state for both models.

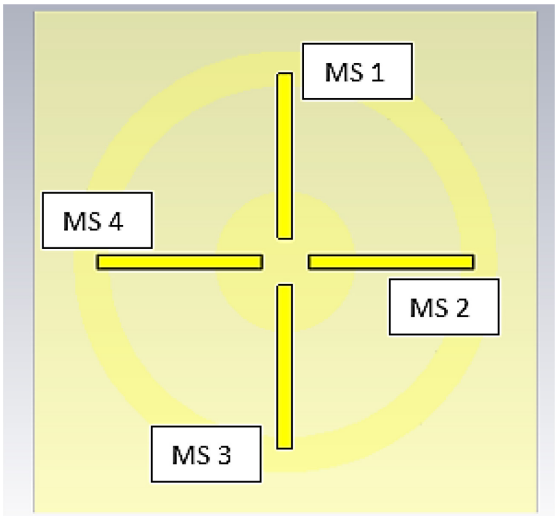


Fig. 4. Metal strip position for a unit cell on the metasurface.

https://doi.org/10.31436/iiumej.v25i2.2954

Fig. 5 and 6 depict several designs of unit cell configurations for Model 1 and Model 2, respectively.

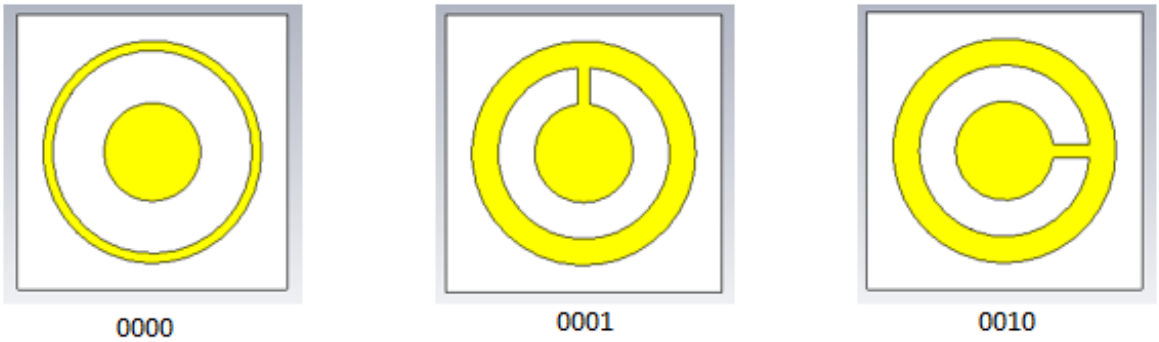


Fig. 5. Configurations of Model 1

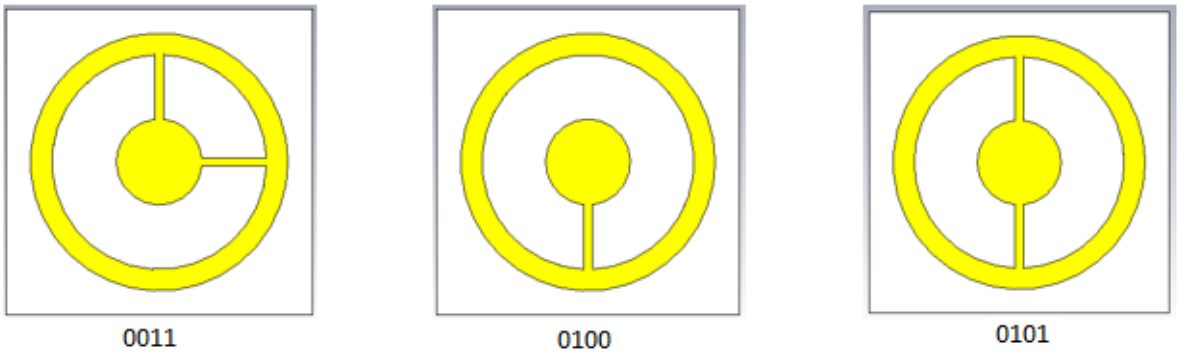


Fig. 6. Configurations of Model 2

3.4 Fabrication and Measurement of Metasurface.

A prototype of the Intelligent Reflecting Surface (IRS) in the 'OFF' state has been fabricated using FR4 material. The choice of material and design was influenced by considerations of cost and material availability. Fig. 7(a) displays the fabricated structure, which consists of a 5 x 5 array of circular patches with ring arrays. This specific configuration was selected to balance fabrication simplicity and effective functionality. The prototype aims to demonstrate the fundamental characteristics and performance of the IRS in a controlled state before further enhancements.

Meanwhile, Fig. 7(b) illustrates the measurement setup designed to assess the IRS's matching condition and overall performance. The setup employs a wideband horn antenna (OBH-10180), capable of operating within the 1-18 GHz frequency range. The reflection coefficient, a critical parameter indicating the efficiency of the reflecting surface, was measured using a vector network analyzer. The measurements were conducted in two distinct conditions: first, with the IRS placed in the path of the horn antenna's emitted waves, and second, without the IRS, serving as a baseline for comparison.

The separation between the horn antenna and the IRS was precisely set at 10 cm to ensure consistent measurement conditions. This setup allows for accurate evaluation of how the IRS modifies the incident electromagnetic waves, reflecting them in a controlled manner. The data obtained from these measurements are crucial for validating the IRS's design and operational principles, paving the way for future optimizations and applications in various communication systems.

https://doi.org/10.31436/iiumej.v25i2.2954

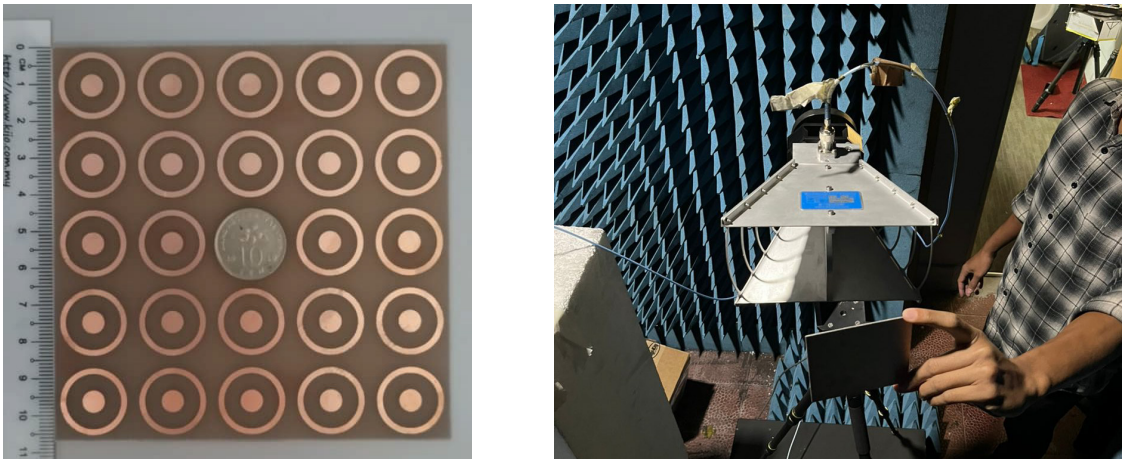


Fig. 7. Prototype (a) A fabricated of 5 x 5 array IRS in OFF state using FR4, (b) Measurement setup of IRS using a wideband horn antenna 1-18 GHz.

4. RESULTS ANALYSIS

This section discusses the simulation and measurement results, including future works for this study.

4.1. Comparison of Unit Cell and Array Analysis

Model 2, having smaller dimensions than Model 1, has been chosen for the analysis. In Fig. 8, a comparison of the magnitude of S11 between the unit cell and the 5 x 5 array of Model 2 is illustrated. The observed magnitudes of S11 at 3.5 GHz are -0.319 dB and -0.329 dB, respectively. The simulation results indicate a small difference of 1.05% between the unit cell model and the array.

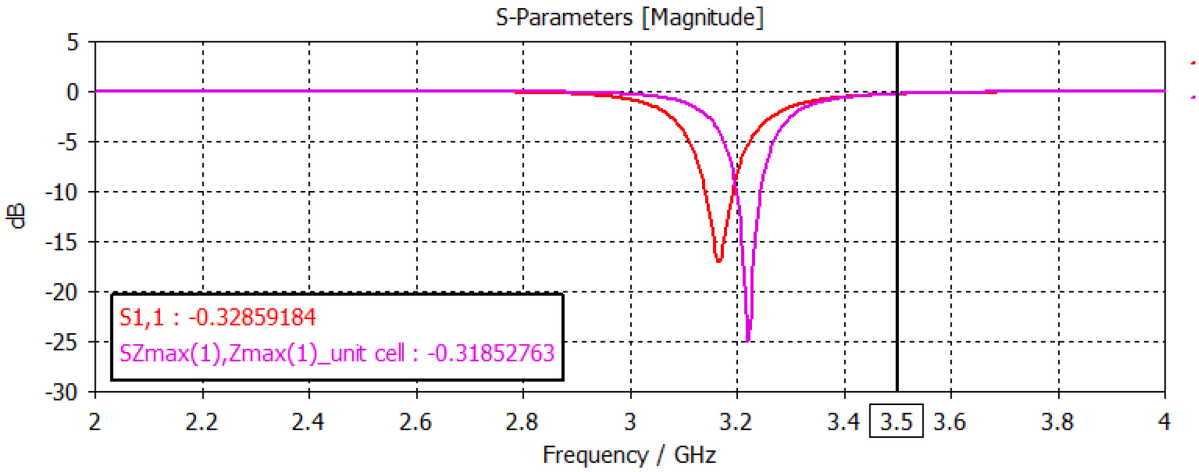


Fig. 8. S11 comparison between unit cell and 5x5 array of Model 2.

4.2. On/Off State of Metasurface

Fig. 9 and Fig. 10 display the graphs of the reflection magnitude of Model 1. The designs are compared to get the best reflection coefficient. It is desirable to achieve the magnitude of S11 as close to 0 to represent the full reflection of the signal from the IRS. The OFF state of Model 1 (0000) has the highest magnitude of reflection coefficient at 3.5 GHz with around -

https://doi.org/10.31436/iiumej.v25i2.2954

0.32 dB, which means 96.4% impinged wave is reflected back toward the desired direction. The magnitude for the 0101 configuration was around -0.46 dB, with an efficiency of 94.78%, reflecting the incoming wave. Fig. 11 and Fig. 12 depict the graphs of the reflection magnitude of Model 2. It has been noted that this model's 0000 setup likewise results in a similar reflection coefficient, approximating -0.32 dB. Several designs, such as configurations 0001, 0010, 0100, 1000, 1010, and 1110, with magnitudes higher than -0.5 dB, provide about 90%.

The best designs are determined by comparing the response for each configuration based on the reflection magnitude and its phase at 3.5 GHz. The parameter's primary concentrate is having the maximum reflection magnitude and the greatest phase difference between the ON and OFF states. By tuning the ON/OFF states, the IRS is able to independently control the phase of the reflected signals.

The comparison between the two models is based on the reflection coefficient, where both model 1 and model 2 exhibit the same magnitude of -0.32 dB. However, the performance differences between the two models become evident when they are in the ON state. Model 1 demonstrates its optimal ON state configuration as 0101, with a reflection magnitude of -0.46 dB. On the other hand, model 2's best ON state is represented by the configuration 0001, featuring a reflection magnitude response of -0.38 dB. Furthermore, the 0001 configuration of model 2 outperforms model 1 in terms of phase difference. The phase difference between the ON and OFF states of model 1 and model 2 is approximately 4º and 325º, respectively. In summary, the configurations 0000 (OFF state) and 0001 (ON state) of model 2 exhibit superior performance in the IRS system, as indicated by the results analysis.

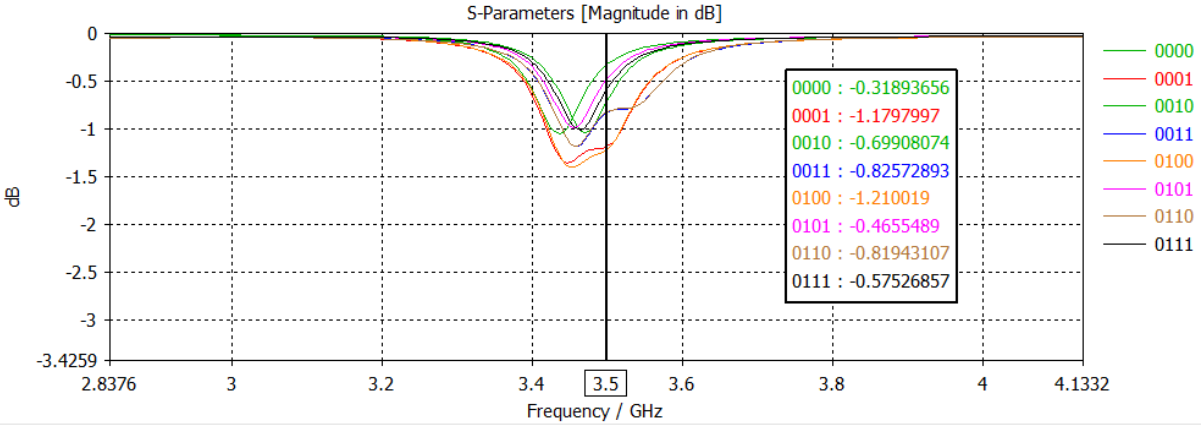


Fig. 9. Reflection magnitude response Model 1 (0000 to 0111)

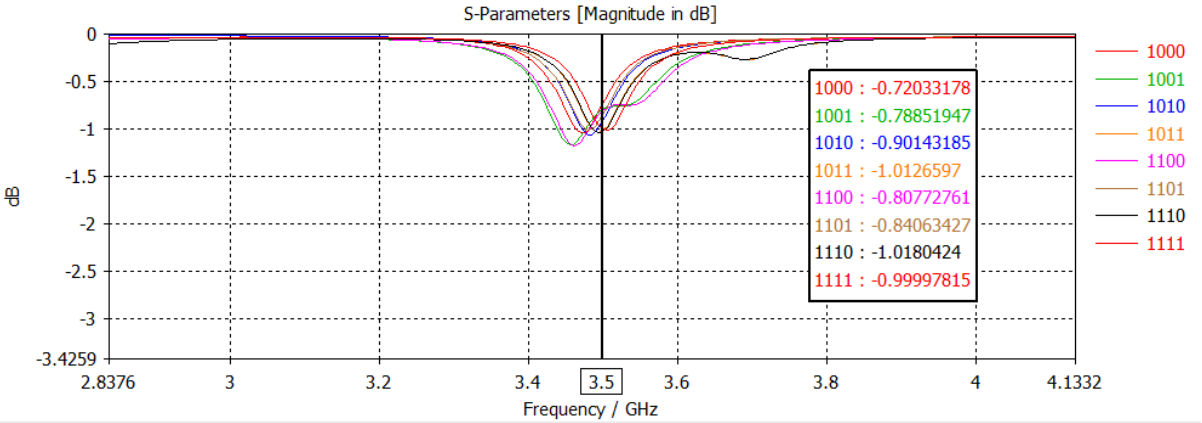


Fig. 10. Reflection magnitude response Model 1 (1000 to 1111)

https://doi.org/10.31436/iiumej.v25i2.2954

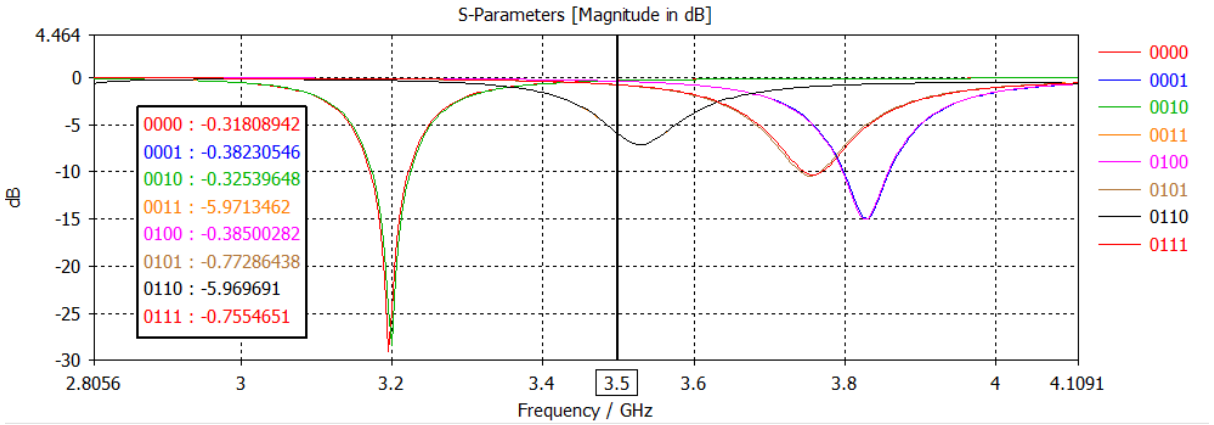


Fig. 11. Reflection magnitude response Model 2 (0000 to 0111)

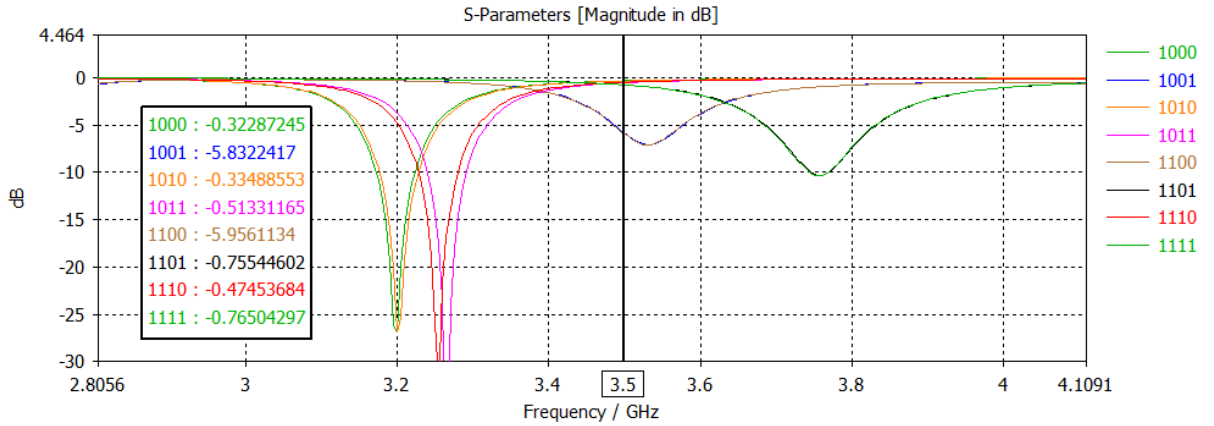


Fig. 12. Reflection magnitude response Model 2 (1000 to 1111)

Fig. 13-14 and 15-16 illustrate the phase response of Model 1 and Model 2, respectively. The phase of the configuration 0000, representing the 'OFF' state, is measured at -329.73°. To achieve a higher phase difference, comparisons were made between each configuration and the 0000 state. Configurations 0001, 0011, 0100, 0110, 1001, and 1100 exhibit a reflection range of 324°. A larger phase range correlates with improved performance of the IRS, as it allows for more precise control over the reflected signals. Table 4 provides a comprehensive summary of the performance comparison between Model 1 and Model 2, focusing on the reflection coefficient and phase response. On average, Model 1 demonstrates an S11 magnitude of -0.82 dB and a phase of 18.7°. In contrast, Model 2 shows an average S11 magnitude of -1.86 dB and a phase of -129.4°. This indicates that while Model 1 has a superior magnitude of S11, suggesting better impedance matching and lower reflection loss, Model 2 excels in achieving a larger phase difference in the reflected signal. Therefore, Model 2, with its significant phase difference, is better suited for applications requiring precise phase manipulation. In contrast, Model 1's performance in terms of reflection coefficient makes it ideal for scenarios prioritizing signal strength and stability.

https://doi.org/10.31436/iiumej.v25i2.2954

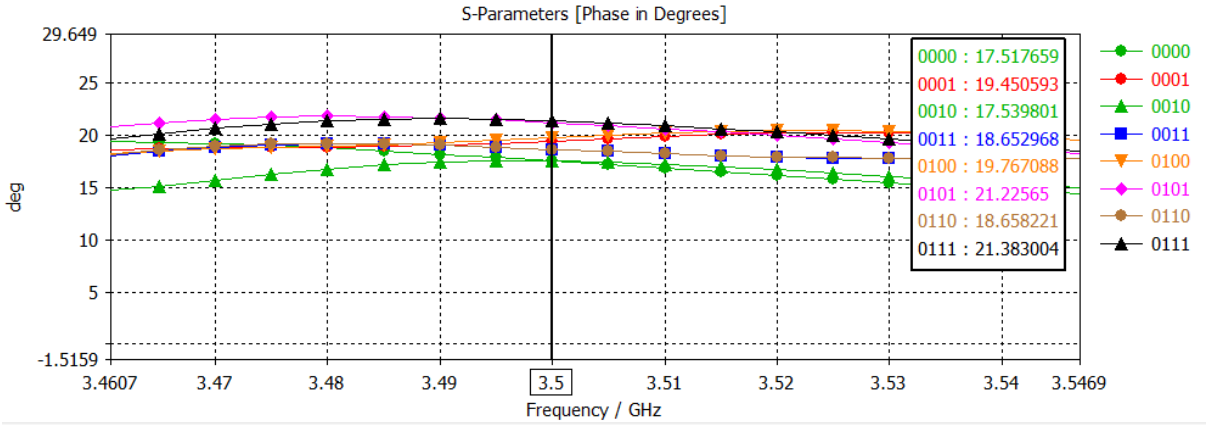


Fig. 13. Unwrap phase response Model 1 (0000 to 0111)

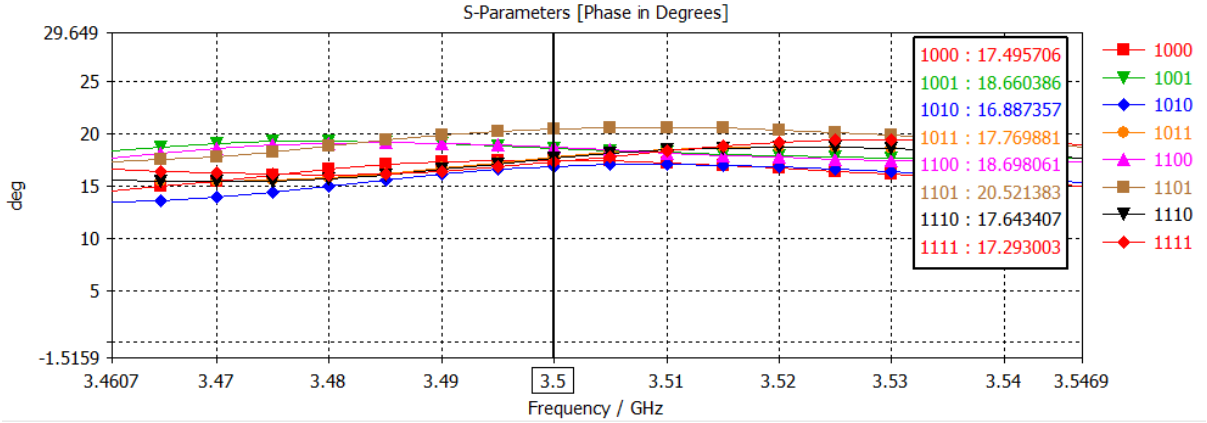


Fig. 24. Unwrap phase response Model 1 (1000 to 1111)

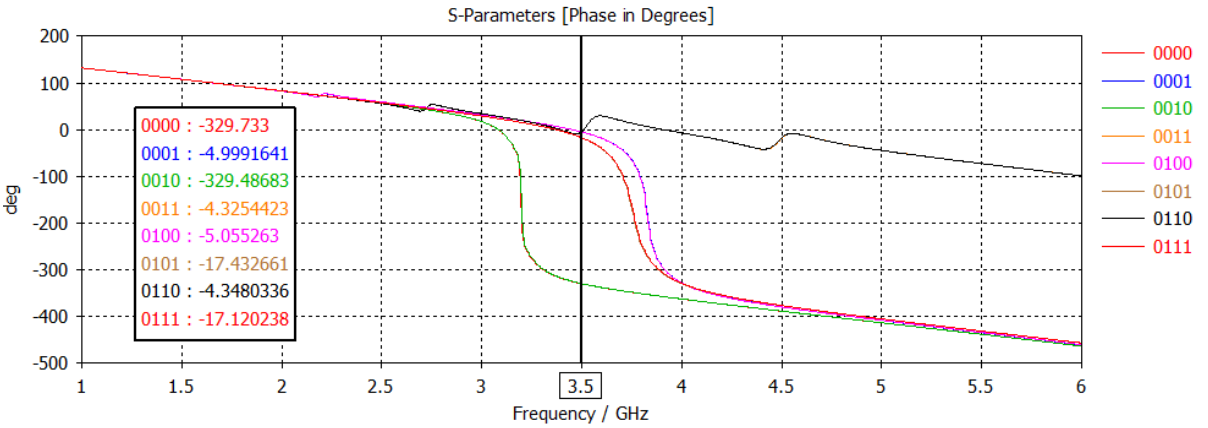


Fig. 35. Unwrap reflection phase response Model 2 (0000 to 0111)

https://doi.org/10.31436/iiumej.v25i2.2954

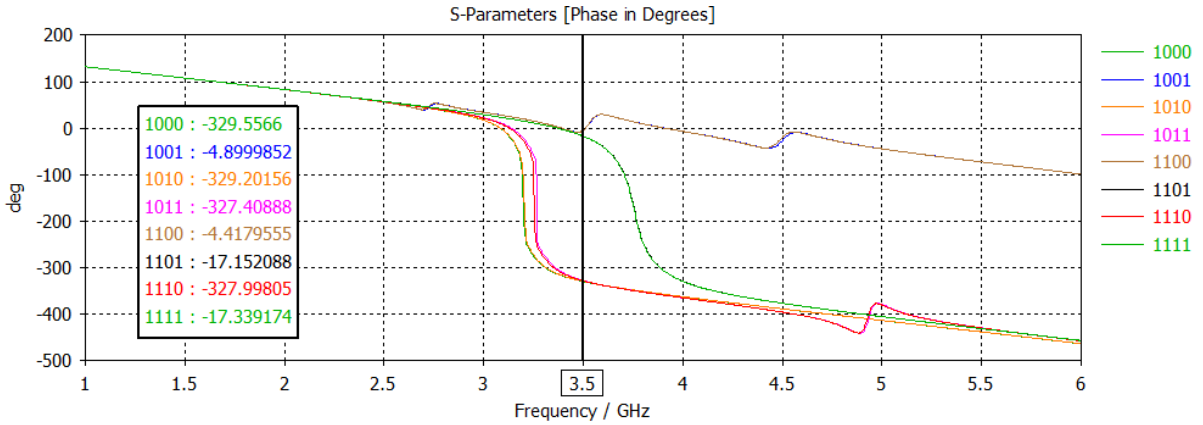


Fig. 46. Unwrap reflection phase response Model 2 (1000 to 0111)

Table 4. Comparison of Model 1 and Model 2 for the ON/OFF states.

Model 1 Configuration	Magnitude [dB]	Phase [º]	Model 2 Configuration	Magnitude [dB]	Phase [º]
0000	-0.3189	17.517	0000	-0.3181	-329.733
0001	-1.1180	19.451	0001	-0.3823	-4.999
0010	-0.6990	17.540	0010	-0.3254	-329.487
0011	-0.8257	18.653	0011	-5.9713	-4.325
0100	-1.2100	19.767	0100	-0.3850	-5.055
0101	-0.4655	21.226	0101	-0.7729	-17.433
0110	-0.8194	18.658	0110	-5.9697	-4.348
0111	-0.5753	21.383	0111	-0.7555	-17.120
1000	-0.7203	17.496	1000	-0.3229	-329.557
1001	-0.7885	18.660	1001	-5.8322	-4.900
1010	-0.9014	16.887	1010	-0.3349	-329.201
1011	-1.0127	17.770	1011	-0.5133	-327.409
1100	-0.8077	18.698	1100	-5.9561	-4.418
1101	-0.8406	20.521	1101	-0.7554	-17.152
1110	-1.0180	17.643	1110	-0.4745	-327.998
1111	-1.0000	17.293	1111	-0.7650	-17.339

In Fig. 17, the graph illustrates the phase difference concerning the number of bits for both Model 1 and Model 2. Notably, it is evident that Model 1 exhibits a constant and lower phase difference when compared to Model 2. The comparison confirms that Model 2 surpasses Model 1, as it presents a more substantial phase variation, an average of -129.4°, in contrast to 18.7° for Model 1. Therefore, Model 2 has been selected for the fabrication process in the OFF state due to its more compact design compared to Model 1, favorable S11 magnitude, and a broad range of S11 phase variations.

https://doi.org/10.31436/iiumej.v25i2.2954

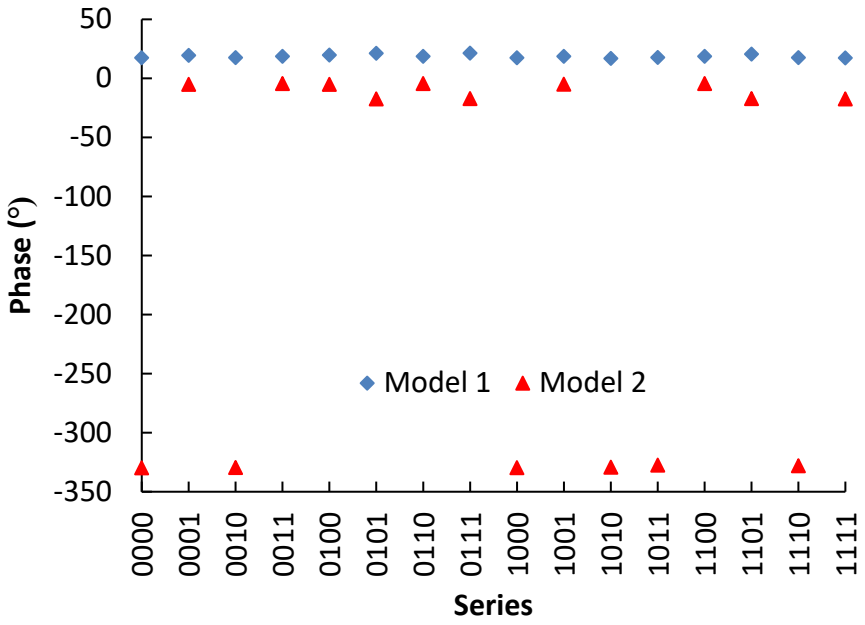


Fig. 17. Comparison of phase shift versus no. of bits between Model 1 and 2.

Table 5 compares the simulation results of Model 1 and Model 2 together with the previous work by [17] for the Off-state. The work by [17] does not provide a numerical study for the ON state. It is observed that for the OFF state (0000), Model 2 has better magnitude and phase variation of S11 at 3.5 GHz.

Table 5. The comparison of proposed work with [17].

Reference	Dimension (mm)	Magnitude of S11(dB)	Phase of S11 (°)	On/Off State
[17]	42 x 42	-1.6	-150	No
Model 1 (**this work**)	42 x 42	-0.3189	17.517	Yes
Model 2 (**this work**)	21 x 21	-0.3181	-329.733	Yes

4.3. Fabrication and Measurement

An IRS prototype for the OFF state has been fabricated and demonstrated as a reflecting surface for a horn antenna based on Fig. 7. Fig. 18 has shown that there is an increase to the S11 of the horn antenna at 3.475 GHz from -21.72 (6.73 mW) to -27.27 dB (1.87 mW). The improvement is calculated at around 72.2%. In addition, it illustrates that the resonance frequency of the horn antenna remains constant despite the presence of IRS. It proves that the proposed IRS is a good reflector that improves the matching condition of the horn antenna.

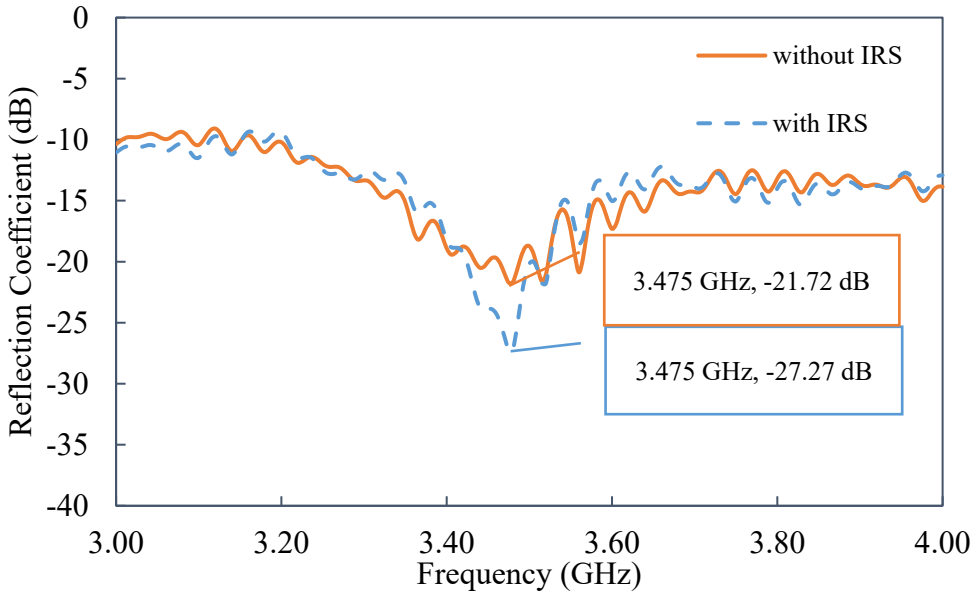


Fig. 18. The measured reflection coefficients of horn antenna with and without reflecting surface at OFF state.

5. CONCLUSION

In this research, an IRS was developed to operate at 3.5 GHz, considering both the magnitude and phase of the reflection coefficients. A circular patch with a ring metasurface was successfully designed by comparing two models with varying sizes. The unit cell dimensions for Model 1 and Model 2 were 42 mm x 42 mm and 21.4 mm x 21.4 mm, respectively. Each design featured 16 different ON/OFF configurations to manipulate the phase of reflected signals. Model 1 demonstrated superior performance with an average of -0.82 dB, compared to -1.86 dB for Model 2, across the 16 ON/OFF states. Regarding phase, Model 2 exhibited a larger variation, averaging -129.4°, in contrast to 18.7° for Model 1. However, for the OFF (0000) state, Model 2 has been selected due to smaller size, high S11 magnitude and larger phase variation for array fabrication. When the IRS was applied to a horn antenna, it resulted in a significant 72.2% improvement in the matching condition of the horn antenna. Future works include analysis of equivalent circuits for the proposed structure and extensive measurement study of reflection coefficients.

ACKNOWLEDGEMENT

The author wishes to thank Research Management Centre and International Islamic University Malaysia (IIUM) for supporting the dissemination of this research.

REFERENCES

[1] A. Araghi *et al.*, "Reconfigurable Intelligent Surface (RIS) in the Sub-6 GHz Band: Design, Implementation, and Real-World Demonstration," *IEEE Access*, vol. 10, pp. 2646–2655, 2022, doi: 10.1109/ACCESS.2022.3140278.

[2] F. C. Okogbaa *et al.*, "Design and Application of Intelligent Reflecting Surface (IRS) for Beyond 5G Wireless Networks: A Review," *Sensors*, vol. 22, no. 7, pp. 1–24, 2022, doi: 10.3390/s22072436.

[3] K. Sun *et al.*, "An overview of metamaterials and their achievements in wireless power transfer," *J. Mater. Chem. C*, vol. 6, no. 12, pp. 2925–2943, 2018, doi: 10.1039/c7tc03384b.

[4] V. Tapio, I. Hemadeh, A. Mourad, A. Shojaeifard, and M. Juntti, "Survey on reconfigurable intelligent surfaces below 10 GHz," *Eurasip J. Wirel. Commun. Netw.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13638-021-02048-5.

[5] W. Long, R. Chen, M. Moretti, W. Zhang, and J. Li, "A promising technology for 6g wireless networks: Intelligent reflecting surface," *J. Commun. Inf. Networks*, vol. 6, no. 1, pp. 1–16, 2021, doi: 10.23919/jcin.2021.9387701.

[6] M. A. S. Sejan, M. H. Rahman, B. S. Shin, J. H. Oh, Y. H. You, and H. K. Song, "Machine Learning for Intelligent-Reflecting-Surface-Based Wireless Communication towards 6G: A Review," *Sensors*, vol. 22, no. 14, pp. 1–21, 2022, doi: 10.3390/s22145405.

[7] J. A. Hodge, T. G. Spence, and A. I. Zaghloul, "A Reconfigurable Intelligent Surface Using a 2-Bit Programmable Metasurface for Communications," *2021 IEEE Int. Symp. Antennas Propag. North Am. Radio Sci. Meet. APS/URSI 2021 - Proc.*, pp. 97–98, 2021, doi: 10.1109/APS/URSI47566.2021.9703874.

[8] G. C. Trichopoulos *et al.*, "Design and Evaluation of Reconfigurable Intelligent Surfaces in Real-World Environment," *IEEE Open J. Commun. Soc.*, vol. 3, no. February, pp. 462–474, 2022, doi: 10.1109/OJCOMS.2022.3158310.

[9] K. Neema, A. Manoharan, K. J. Vinoy, and D. Das Krishna, "Reconfigurable Intelligent Surface for Highly Efficient Electromagnetic Functionality," *2022 IEEE Wirel. Antenna Microw. Symp. WAMS 2022*, pp. 2–4, 2022, doi: 10.1109/WAMS54719.2022.9847674.

[10] J. Rains, J. U. Rehman Kazim, L. Zhang, Q. H. Abbasi, M. Imran, and A. Tukmanov, "2.75-Bit Reflecting Unit Cell Design for Reconfigurable Intelligent Surfaces," *2021 IEEE Int. Symp. Antennas Propag. North Am. Radio Sci. Meet. APS/URSI 2021 - Proc.*, pp. 335–336, 2021, doi: 10.1109/APS/URSI47566.2021.9704689.

[11] A. A. S. Al - Tag, R. M. Al-mahdi, M. B. Al-hedari, B. A. Al-ariqi, and A. Al-hetar, "New Design of Intelligent Reflective Surface Based on Metamaterial Surface with Configurable Gain and Direction," pp. 1–5, 2022, doi: 10.1109/esmarta56775.2022.9935135.

[12] T. Sharma, A. Chehri, and P. Fortier, "Reconfigurable Intelligent Surfaces for 5G and beyond Wireless Communications: A Comprehensive Survey," *Energies*, vol. 14, no. 24, p. 8219, 2021, doi: 10.3390/en14248219.

[13] X. Song *et al.*, "Switchable metasurface for nearly perfect reflection, transmission, and absorption using PIN diodes," *Opt. Express*, vol. 29, no. 18, p. 29320, 2021, doi: 10.1364/oe.436261.

[14] S. Foo, "Liquid-crystal reconfigurable metasurface reflectors," *2017 IEEE Antennas Propag. Soc. Int. Symp. Proc.*, vol. 2017-Janua, pp. 2069–2070, 2017, doi: 10.1109/APUSNCURSINRSM.2017.8073077.

[15] D. Rotshild *et al.*, "Steer by Image Technology for Intelligent Reflecting Surface Based on Reconfigurable Metasurface with Photodiodes as Tunable Elements," *Crystals*, vol. 12, no. 7, 2022, doi: 10.3390/cryst12070951.

[16] M. Dajer *et al.*, "Reconfigurable intelligent surface: design the channel – a new opportunity for future wireless networks," *Digital Communications and Networks*, vol. 8, no. 2. pp. 87–104, 2022. doi: 10.1016/j.dcan.2021.11.002.

[17] B. Rana, S. S. Cho, and I. P. Hong, "Parameters and Measurement Techniques of Reconfigurable Intelligent Surfaces," *Micromachines*, vol. 13, no. 11, pp. 1–12, 2022, doi: 10.3390/mi13111841.

[18] M. F. Ismail, M. K. A. Rahim and H. A. Majid, "The Investigation of PIN diode switch on reconfigurable antenna," 2011 IEEE International RF & Microwave Conference, Seremban, Malaysia, 2011, pp. 234-237, doi: 10.1109/RFM.2011.6168737.

[19] M.R. Hamid, P. Gardner, P.S. Hall "Frequency Reconfigurable Log Periodic patch Array" Electronic Letters, Vol.46; No.25, 2010.

[20] M. F. Ismail, M. K. A. Rahim, F. Zubir, O. Ayop "Log-Periodic Patch Antenna with Tunable Frequency" Proceedings of the 5 th European Conference on Antennas and Propagation (EuCAP), pages 2165-2169, 2011.

https://doi.org/10.31436/iiumej.v25i2.2963

BRAIN TUMOR SEGMENTATION AND CLASSIFICATION USING CNN PRE-TRAINED VGG-16 MODEL IN MRI IMAGES

T. GAYATHRI [1*], K. SUNDEEP KUMAR [2]

[1]*Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, India*
[2]*Department of CSE, Geethanjali Institute of Science and Technology, Nellore, India*

**Corresponding author: gayal28@gmail.com*

(Received: 15 August 2023; Accepted: 12 February 2024; Published online: 15 July 2024)

ABSTRACT: The formation of a group of abnormal cells in the brain that penetrate the neighboring tissues is known as a brain tumor. The initial detection of brain tumors is necessary to aid doctors in treating cancer patients to increase the survival rate. Various deep learning models are discovered and developed for efficient brain tumor detection and classification. In this research, a transfer learning-based approach is proposed to resolve overfitting issues in classification. The BraTS – 2018 dataset is utilized in this research for segmentation and classification. Batch normalization is utilized in this experiment for data pre-processing and fed to a convolutional layer of CNN for extracting features from Magnetic Resonance Images (MRI). Then, an Adaptive Whale Optimization (AWO) algorithm is utilized to select effective features. This work proposes a Convolutional Neural Network (CNN) based segmentation and a transfer learning-based VGG-16 model for effective classification. The performance of the proposed CNN-VGG-16 technique is analyzed through various tumor regions like TC, ET, and WT. The proposed method attains a Dice score accuracy of 99.6%, 95.35%, and 94%, respectively, when compared to other existing algorithms like CNN, VGG-net, and ResNet.

ABSTRAK: Pembentukan gumpalan sel abnormal dalam otak yang menembusi tisu-tisu jiran adalah dikenali sebagai tumor otak. Pengesanan awal tumor otak adalah penting bagi membantu doktor merawat pesakit kanser bagi meningkatkan kadar jangka hayat. Terdapat banyak model pembelajaran mendalam berkaitan kecekapan pengesanan tumor otak dan pengelasan. Dalam kajian ini, pendekatan pembelajaran berdasarkan pindahan dicadangkan bagi mengatasi isu terlebih padan dalam pengelasan. Set data BraTS – 2018 telah digunakan dalam kajian ini bagi tujuan pensegmenan dan pengelasan. Kelompok normal digunakan dalam eksperimen ini bagi data awal proses dan disalurkan kepada jalur lingkaran CNN bagi mengekstrak ciri-ciri dari Imej Resonan Magnetik (MRI). Kemudian, algoritma Optimalisasi Mudah Suai 'Whale' (AWO) digunakan bagi memilih ciri-ciri berkesan. Kajian ini mencadangkan Lingkaran Rangkaian Neural (CNN) berdasarkan segmentasi dan model VGG-16 berdasarkan pindahan bagi pengelasan berkesan. Prestasi teknik CNN-VGG-16 yang dicadangkan diuji dengan pelbagai bahagian tumor otak seperti TC, ET dan WT. Kaedah yang dicadangkan ini beroleh ketepatan skor Dice sebanyak 99.6%, 95.35% dan 94% masing-masing jika dibanding dengan algoritma sedia ada seperti CNN, VGG-net dan ResNet.

KEYWORDS: *Adaptive Whale Optimization, Brain tumor, Convolutional Neural Network, Magnetic Resonance Images, VGG-16.*

https://doi.org/10.31436/iiumej.v25i2.2963

1. INTRODUCTION

The human brain is an important organ in the nervous system that is responsible for the activities in humans' daily lives [1]. The brain gathers signals from the organs of the body and then handles processing, manages decisions, and gives the resultant data to the muscles [2]. A brain tumor is a collection of unmanaged cancer cells which grow around the brain. Brain tumors are classified into two types, namely, primary tumors, which grow in the spinal cord or brain alone, and secondary tumors, which are also known as brain metastases that grow anywhere in the human body and spread to the brain [3]. There are various types of scan imaging systems, such as Computed Tomography (CT), Electroencephalogram (EEG), and Magnetic Resonance Imaging (MRI), which are used for providing significant information about the vicinity, dimension, and metabolism of cerebrum tumors, assisting decision [4]. These systems are combined to produce the major information about cerebrum tumors [5]. MRI is a non-intrusive system that utilizes the signals of radio recurrence to develop interior images under the influence of an extremely attractive field [6]. The MRI system gives different types of contrast images in tissues, eventually producing significant auxiliary information and analyzing the tumor segmentation within its subregions [7].

The detection of brain tumors from MRI images is a difficult task because of the requirements of a trademark [8,9]. The accuracy of brain tumor diagnosis is increased by the process of segmentation. In recent times, methods based on deep learning are frequently utilized in medical image segmentation [10]. There is a chance of acquiring many types of noise like salt & pepper, speckle, and Gaussian while obtaining the MRI images [11]. So, the technique of noise removal is important for decreasing the noise. The selection of features plays a major role in the classification stage because it decreases the evaluation time and maximizes the performance of classification [12]. The application of deep learning produces an ideal solution because it extracts more prominent features from the entire image than the manually defined features [13]. The most recently adopted techniques for segmentation depend on deep learning, needing a masked image for an expected outcome [14]. Those labels support and guide the learning process during the segmentation stage, but the process takes more time to compute [15]. From the overall existing analysis, detecting tumors in their early stages was impossible with the previous methods; the low-grade and high-grade tumors were not classified, and without segmented images, the whole image dataset was used for classification, which resulted in neglecting the extraction of objects of interest and restricted the data available. Additionally, the computational cost was high, even when done only for the classification and not segmentation. The very challenging aspect of MRI scans is that they are not like X-ray images (2D). An MRI image is made up of various 3D volumes that represent different parts of the brain. During the segmentation of the image, these 3D volumes are integrated. While integrating different channels of MRI images, some misalignments occur that result in error. To overcome these issues, the process of segmentation should depend on the CNN architecture without utilizing masked images. The extracted features contain redundant noise, which minimizes the classification performance. To remove the redundant features, the feature selection process is required in brain tumor classification. Adaptive Whale Optimization (AWO) algorithm is used to select features from the extracted images in the convolutional layer. The AWO is utilized to overcome the limitations of whale optimization algorithms, which include low convergence speed and less capacity of global optimization, and can easily fall into local optima. The AWO algorithm improves the speed of convergence, and the ability of exploration to ignore falls into the local optima. The AWO consists of three phases: prey search, encircling prey, and bubble-net attacking.

The significant contributions of the research are given below:

- The Adaptive Whale Optimization (AWO) algorithm is utilized to select the necessary features; the AWO improves the convergence speed of WOA and ignores falling into the local optimum. Using selected relevant features for classification maximizes the classification accuracy.
- The CNN and pre-trained VGG-16 model is proposed for the image segmentation and classification respectively, to reduce overfitting and the time consumed on processing smaller batch sizes.
- The performance of the segmentation process is evaluated using DSC and HD measures, and the classification method is estimated by utilizing accuracy, precision, and f-measure criteria with various tumor regions such as Tumor Core, Enhancing Tumor, and Whole Tumor.

The rest of this research paper is described as follows: the relative research on brain tumor segmentation and classification is given in Section 2. The proposed method for segmentation and classification of brain tumors is explained in Section 3. The outcomes and comparative analysis of the proposed method are given in Section 4, and the conclusion of the paper is given in Section 5.

2. LITERATURE SURVEY

Sharif et al. [16] introduced a four-phased brain tumor prediction model that included lesion enhancement, feature extraction, selection for classification, localization, and segmentation. An inception v3 model was used to extract and select features using Non-Dominated Sorting Genetic Algorithm (NSGA). Two methods were utilized; the transfer learning-based approach was used in the classification phase and a YOLO v2 inception v3 model was used in the localization phase. Another popular method named McCulloch's Kapur entropy was used for segmenting tumor regions. The research was validated on BRATS 2018, 2019, and 2020 datasets, respectively. However, this technique was not suitable for detecting tumors in their early stages, and future work intends to implement a quantum computation algorithm to check the system's efficiency.

Decuyper et al. [17] developed an automatic brain tumor classification and segmentation pipeline model using an MRI of the BRATS 2019 dataset. A 3D U-Net was used to segment the tumor region, followed by extracting the Region of Interest (ROI) from MRI scans. To predict the grade, IDH mutation, and 1p19q at a time, the extracted features were fed into CNN. A multi-class classification enabled the network to train on large datasets and resulted in high classification accuracy. Individual classes were not predicted due to the limitation of only two IDH mutant glioblastoma in the training set, so the classification resulted in combined classes. The future work intends to be a non-invasive evaluation of genetic mutations that help in the process of classification of glioma.

Rajinikanth et al. [18] implemented an architecture of the Deep Learning method for the automatic detection of brain tumors using 2D MRI slices. The pre-trained DL networks like AlexNet, VGG16, ResNet 50, and ResNet 101 were implemented with deep features followed by classification. The classification included implementing popular classifiers like DT, KNN, SVM-linear, and SVM-RBF. To improve the detection accuracy, a VGG19 network was fused in the model, resulting in better accuracy with the help of SVM-RBF. This model was validated on the BRATS and TCIA datasets, thereby achieving better results. However, the low-grade

and high-grade tumors were not classified. Future works intend to improve categorization accuracy by adjusting fully connected and drop-out layers.

Tandel et al. [19] presented five DL and ML-based brain tumor detection models for effective classification and detection of brain tumors, which were trained and tested on four sets of natural image datasets. A majority voting ensemble optimization algorithm was introduced to validate and improve the face image classification of males versus females. The five DL models included AlexNet, VGG-16, GoogleNet, ResNet18, and ResNet50. The five ML models included Decision tree, Naïve Bayes, KNN, SVM, and linear discrimination using five-fold cross-validation. The proposed ensemble model achieved better accuracy and promising results. Instead of using segmented images, the whole image dataset was used for classification, which resulted in neglecting extracting objects of interest, which is a major drawback of this work. This was overcome by proposing image segmentation techniques.

Punn and Agarwal [20] introduced an effective $3D$ Deep Neural Network model for the segmentation of brain tumors. A $3D$ inception U-Net model and multi-modalities fusion were used to extract the tumor regions and their patterns along with segmentation. Each tumor region was divided into CT, WT, and ET by using target classes such as necrosis, edema, background, enhancing tumor, and non-enhancing tumor. The class imbalance issues were solved by introducing the loss function of weighted segmentation. The introduced method was validated on BRATS 2017 and 2018 datasets. The performed multi-modality fusion was not applied to the real-time data fusion and cascading. The future work includes a wide range of biomedical applications such as image registration, disease quantification, and other application fields.

Agrawal et al. [21] introduced a 3D-UNet deep neural network for the segmentation and classification of brain tumors utilizing the Kaggle dataset. The introduced method depended on the 3D segmentation of MRI images. The images of the brain in the 3D volume were divided into 3D sub-volumes that were fed into the method of segmentation and rejoined into one 3D volume. The advantage of the method was that it represented critical features directly from multiple-modal MRI images of both healthy and tumor brain tissues. The limitation of the introduced method was that the limited data available and the computational cost was high.

Raza et al. [22] presented a hybrid deep learning method named DeepTumorNet for the classification of brain tumors using the CE-MRI dataset. The GoogleNet structure of the CNN method was utilized as a base for the development of the hybrid method. In developing a hybrid DeepTumorNet technique, the final five layers of GoogleNet were deleted, and 15 new layers were added rather than these five layers. The ReLU activation function was utilized in the feature map to maximize the model's expressiveness. The advantage of the presented method was that it computed many discriminative and descriptive data and accurate features for the classification of brain tumors. The limitation of the method was that it carried out only classification and not segmentation, and the segmented images provided higher accuracy.

Haq et al. [23] implemented a hybrid technique dependent on Deep Convolutional Neural network (DCNN) and machine learning classifiers for the segmentation and classification of brain tumors using MRI images. CNN, in the initial stage, was utilized to learn the feature map from the image space of MRI images and the region of the tumor marker. Next, faster region-based CNN was created for the tumor region localization by utilizing a Region Proposal Network (RPN). Finally, DCNN and ML classifiers were combined in series for segmentation and classification. The method attained automatic segmentation and classification of brain tumors without a user interface. However, the limitation of the method was high computational time.

https://doi.org/10.31436/iiumej.v25i2.2963

3. PROPOSED METHODOLOGY

In the proposed methodology, the CNN-based VGG-16 pre-trained method is proposed for a better classification of brain tumors. An imaging dataset of BraTS 2018 is utilized for the proposed method. Batch normalization is utilized in the pre-processing stage, and a convolutional layer of CNN is utilized for the extraction of features from MRI images. The AWO algorithm is employed to select the features. The fully connected CNN is employed for image segmentation, and the pre-trained model VGG-16 is employed for effective classification. The flowchart of the proposed methodology is represented in Fig. 1.

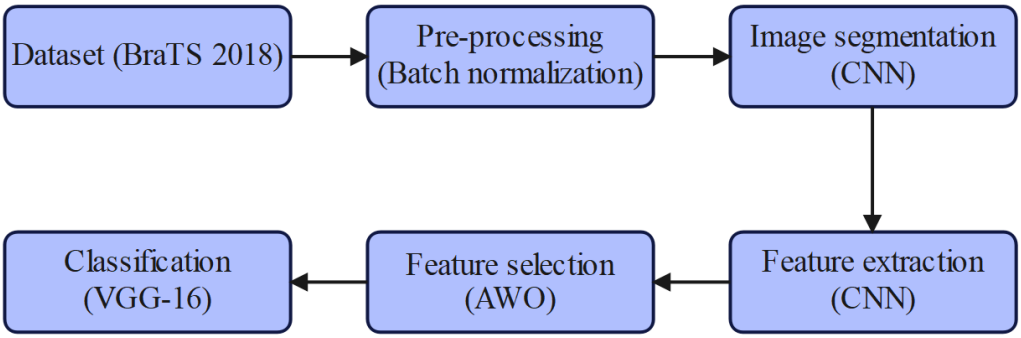


Fig. 1. Workflow of proposed methodology

3.1. Dataset

The imaging dataset utilized for this research is a publicly available dataset named Brain Tumor Segmentation (BraTS-2018) [24] to segment and classify brain tumor MRI images. The dataset has four multiple contrast MRI scans of 477 patients with high or low-grade glioma brain tumors. The four multiple contrast MRI scans generated for every patient are $T1$ (T1-weighted gradient-echo), $T1 - Gd$ (T1-weighted post-gadolinium contrast gradient-echo), $T2$ (T2-weighted gradient-echo) and T2-FLAIR or FLAIR (T2-weighted fluid-attenuated inversion recovery). The sample images of the dataset are represented in Fig. 2.

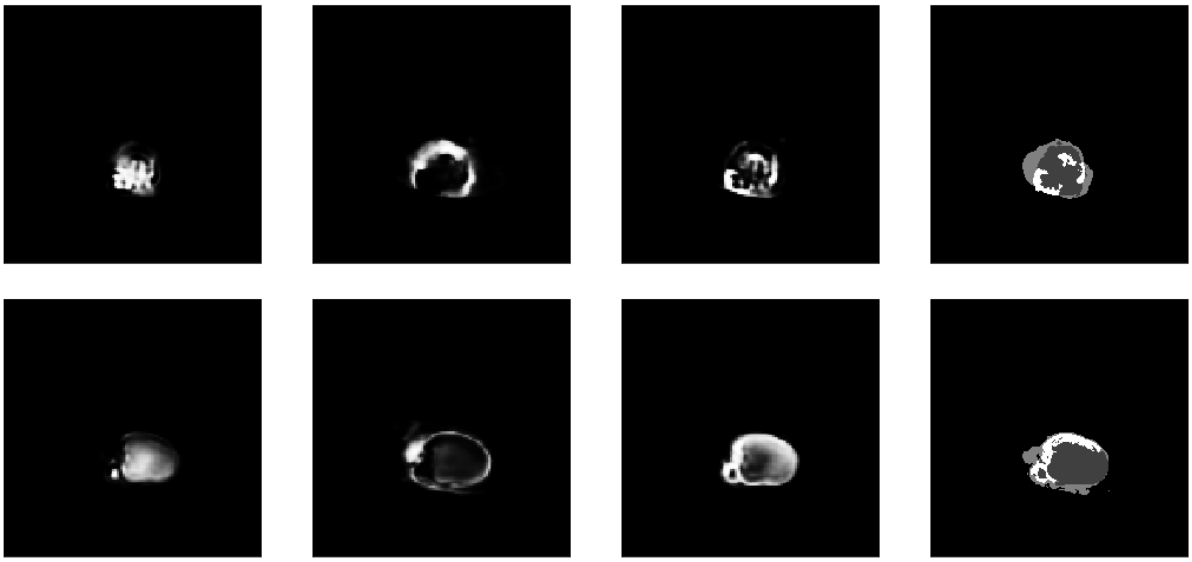


Fig. 2. Sample images of dataset

https://doi.org/10.31436/iiumej.v25i2.2963

3.2. Data Pre-processing

The batch normalization is a pre-processing method in this research; here, the term 'batch' represents a set of input data and processes performed in terms of batches. Normalization is used to change numerical data to a general scale while protecting their shapes. In other words, it is a procedure of converting data to 0 as an average value and 1 as standard deviation. Batch normalization is a process of making neural networks quicker with huge stability by including additional layers in deep neural networks. The standardizing and normalization processes are processed through a new layer by taking values of the prior layer.

The mean and standard deviation (SD) of hidden activation is measured by considering batch input along layer h as Eq. (1) and Eq. (2).

$$\mu = \frac{1}{m}\sum h_i \tag{1}$$

$$\sigma = \left[\frac{1}{m}\sum(h_i - \mu)^2\right]^{1/2} \tag{2}$$

Where, μ is the mean, σ represents SD, and m represents the number of neurons in the layer h.

By utilizing Eq. (1) and Eq. (2), hidden activations are normalized. The average value is subtracted along every input, and then an entire value is divided with the addition of standard deviation and smoothing term (ε). Smoothing makes sure to keep the numerical values stable within the process by protecting separation by zero and is measured using Eq. (3).

$$h_{i(norm)} = \frac{(h_i - \mu)}{\sigma + \varepsilon} \tag{3}$$

After the stage of pre-processing, the training set's size is maximized to overcome the overfitting problem. The pre-processed information is fed to the fully connected CNN for segmentation of brain MRI images.

3.3. Image Segmentation

In image processing, a neural network is the most powerful method to perform segmentation. MRI is the major imaging technique for capturing brain images. Automatic segmentation is a difficult task due to the variability of structure and large spatial area. So, the research implements an automatic segmentation technique based on CNN with 3×3 Kernels [25]. The Kernels of smaller size help to develop the deep architecture by utilizing a smaller number of weights in the network. Kernels are shared through whole units of similar feature maps, and the conv layer contains smaller weights for training than dense FC layers, thus making the CNN suitable for training, leading to low overfitting.

The following are concepts that are significantly deployed in the CNN segmentation:

3.3.1. Initialization

Initialization is significant for attaining quick convergence. By utilizing the initialization, the activations and gradients are managed in the control stages or else, the back-propagated gradients disappear.

3.3.2. Activation Function

The activation function is manageable for information transformed non-linearly. Rectifier linear units (ReLU) are represented by Eq. (4).

https://doi.org/10.31436/iiumej.v25i2.2963

$$f(x) = max(0, x) \quad (4)$$

The activation function is found for attaining superior results than traditional sigmoid, as well as to fasten the training. However, establishing a constant 0 damages the flow of gradient and successive weight adjustments. To overcome these limitations, utilize a Leaky Rectifier Linear unit (LeakyReLU) which implements a little slope on the function's negative part. The function is represented as Eq. (5).

$$f(x) = max(0, x) + \alpha\, min(0, x) \quad (5)$$

where, α represents the parameter of leakiness, and a fully connected layer is utilized as a softmax.

3.3.3. Pooling

Pooling combines the dimensions of close features in feature maps. The integration of possible unnecessary features makes the portrayal much more compact and invariable to little changes of the image, such as unwanted information, which also reduces the estimation load of the next phases. To merge features, it is typical to utilize max-pooling.

3.3.4. Regularization

Regularization is utilized to minimize overfitting. The dropout is utilized in a fully connected layer. In every training stage, it deletes nodes from the network with the possibility. Like this, it forces each node of fully connected layers to learn the best data description, protecting nodes from cooperating. Every node is used at the testing stage while the dropout is represented as an ensemble of various networks and bagging form, since all networks are trained with the training data portion.

3.3.5. Loss Function

Loss function is the process that needs to be decreased in the training phase. The research deploys a categorized Cross-entropy described as Eq. (6).

$$H = -\sum_{j \in voxels} \sum_{k \in classes} c_{j,k}\, log\left(\hat{c}_{j,k}\right) \quad (6)$$

Where, $\hat{c}$ represents the prediction of probabilistic (after the softmax) and c represents the target.

By utilizing the fully connected CNN model, the tumor-affected MRI brain images are segmented, and the segmented images are fed into the convolutional layer of the CNN model for extracting features.

3.4. Feature extraction

Feature extraction is a process of converting original information into numerical features which are processed while protecting data in the original dataset. Feature extraction provides superior outcomes as opposed to applying the raw data directly to machine learning. The conv layer is a fundamental layer of CNN structure which performs feature extraction, including the combination of linear and non-linear operations. In this research, CNN is used to extract features that input features of image classification, based on various neural networks. The 3D volume dimensions such as N, 155, 240, and 4 are extracted by the conv layer. The extracted features contain redundant noises as features that minimize the performance of classification. To remove the redundant features, the feature selection process is required in brain tumor

https://doi.org/10.31436/iiumej.v25i2.2963

classification. The features extracted through convolutional layers are fed into the AWO algorithm for selecting features.

3.5. Feature Selection

Adaptive Whale Optimization (AWO) algorithm is utilized to select features from images extracted from the Convolutional layer. The AWO is utilized to overcome the limitations of WOA, such as low convergence speed, reduced ability to optimize globally, and the ability to easily fall into local optima. The AWO algorithm improves the speed of convergence and increases the ability of exploration to avoid falling into local optima. The AWO consists of three phases: prey search, encircling prey, and bubble-net attacking [26].

3.5.1. Encircling prey

In encircling the prey stage, every agent processes an optimum agent at a distance that depends on A acquired by Eq. (8), and a value is minimized linearly with the iteration. Due to a linear decrease in the size of the step, the performance of the agent exploration also decreases at this phase. To maximize group diversity and avoid falling into the local optimum, the research utilizes AWO in the encircling phase. The AWO is multiplied by the current agent target prey to create a new target and modify the AWO by the agent's fitness. The improved encircling prey phase is represented in Eq. (7).

$$X_i(t+1) = w_i(t) \cdot X^*(t) - A \cdot |C \cdot X^*(t) - X_i(t)| \quad (7)$$

$$A = 2a \cdot r - a \quad (8)$$

$$C = 2 \cdot r \quad (9)$$

Where, X represents the position vector, X^* represents the position vector of the optimum solution obtained, t represents a present iteration, $A\ and\ C$ represent the coefficient of vectors, which are measured using Eqs. (8) and (9).

The components of a are linearly minimized from $2\ to\ 0$ and r represents a random number between $(0,1)$. The convergence's adjustment is decided by the operator $w_i(t)$, range of $w_i(t)\ \in [0,1]$ which is modified by the present agent's position. While the position of the agent in the solution space is matchable, the AWO is maximized.

3.5.2. Bubble-net attacking

This technique mimics the whale's behavior when detecting its prey during the bubble-net hunting phase. The whale positions itself near the prey, using a prey in the central point, resulting in being gradually unconscious in the target. This strategy involves a whale encircling the prey, primarily gauging the distance between itself and its target before veering towards it in a curved manner. The improved bubble-net attacking phase is represented in Eq. (10).

$$X(t+1) = D' \cdot e^{bl} \cdot cos(2\pi l) + X^*(t) \quad (10)$$

$$D' = |CX^*(t) - X(t)| \quad (11)$$

where, D' represents the distance of the i^{th} whale to the optimum solution obtained and D' is calculated as Eq. (11). b describes a constant that defines the shape of the curl, and l represents the random number in $[-1,1]$. That is considered when the whale moves around a prey in the outer circle, which decreases the radius of the enclosure as well.

https://doi.org/10.31436/iiumej.v25i2.2963

3.5.3. Prey search

In prey search, whales utilize a score of A to manage them if they are in the encircling stage. When $|A| > 1$, whales can not acquire effective data on prey and require simultaneous attempts to identify a track of clues of prey along random routes. The numerical representation of a changed strategy of hunting is represented in Eq. (12) and Eq. (13).

$$D = |C \cdot X^{rand}(t) - X(t)| \tag{12}$$

$$X(t+1) = X^{rand}(t) - A \cdot D \tag{13}$$

where, X^{rand} represents randomly choosing a vector of position in the whale.

3.5.4. Quasi-Oppositional Solution

For maximizing the search ability after initialization, a quasi-oppositional solution is developed. The solution is utilized for minimizing execution time and also improves the convergence capacity of WOA.

The numerical expression for the opposite solution S_0 for an arbitrary solution is given as Eq. (14).

$$S_0 = u + v - S \tag{14}$$

The mathematical expression for multiple dimension search space is given as Eq. (15).

$$S_0^i = u^i + v^i - S^i; \;\; i = 1,2, \dots, d \tag{15}$$

The numerical equation in an arbitrary solution of the quasi-opposite solution is given as Eq. (16).

$$S_{q0} = rand\left(\frac{u+v}{2}, S_0\right) \tag{16}$$

The numerical expression for multi-dimensional search space is given as Eq. (17).

$$S_{q0}^1 = rand\left(\frac{u^i+v^i}{2}, S_0^i\right) \tag{17}$$

Using Adaptive Whale Optimization, the necessary features are selected, thereby minimizing the method's executional cost and the model's performance. These selected features are fed to a pre-trained VGG-16 model for effective classification. Fig. 3 represents the process of AWO for feature selection.

https://doi.org/10.31436/iiumej.v25i2.2963

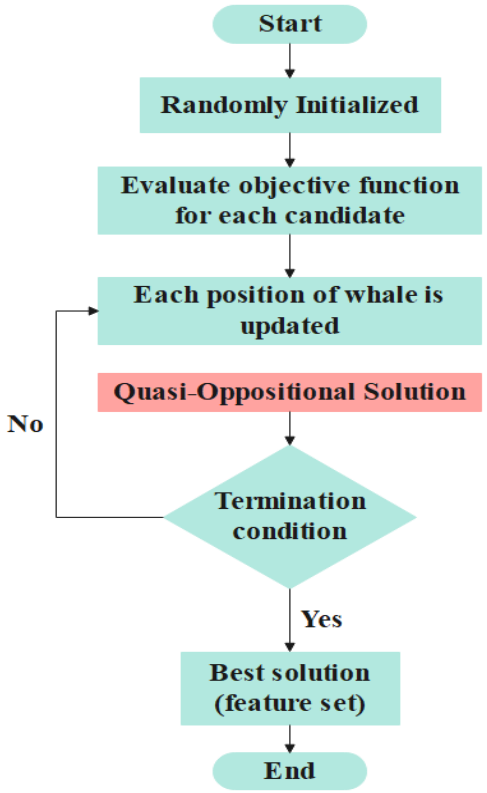


Fig. 3. Process of AWO in feature selection

3.6. Classification

The pretrained CNN structure is deployed for classification, utilizing various labeled images for model training acquired from a feature selection phase. The pretrained CNN structure utilized for the classification is VGG–16. The dense layer is an important part of the classification of brain MRI images, which depends on the results from convolution layers. This gives output from the previous layer to every neuron; every neuron generates a single outcome to the next layer (rework this sentence). Fig. 4 represents the process of VGG-16.

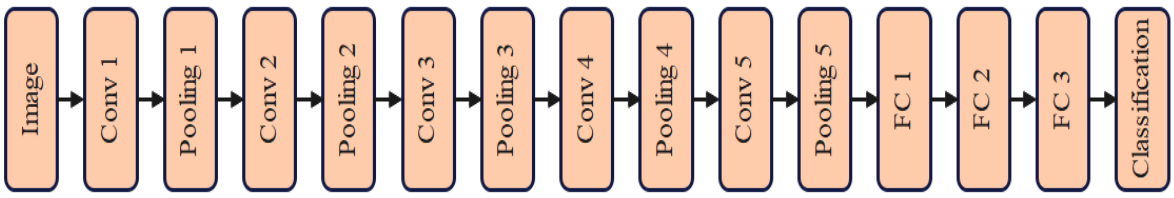


Fig. 4. Pictorial representation of the Classification process of VGG -16

The VGG–16 pre-trained CNN method is employed for classification since there are fewer images in the dataset and in order to overcome the overfitting issue. The method is fine-tuned by frosting certain conv layers. The VGG–16 pretrained CNN method contains 16 convolution layers. The method considers brain MRI images as input along the dimension of $224 \times 224 \times 3$. The method has conv layers with fixed filter size of 3×3 and 5 max-pooling layers of dimension size of 2×2 within the network. The method has the activation function of ReLU and two fully connected layers, having a softmax output layer. The VGG–16 model is a wide network that includes approximately 138 million hyperparameters. The method

https://doi.org/10.31436/iiumej.v25i2.2963

assembles multi-conv layers to develop deep neural networks that improve its capacity for learning handcrafted features, which is invisible. Fig. 2 represents the overall classification process of CNN based pre-trained VGG-16 model. By utilizing the CNN-based pre-trained method of VGG-16, types of brain tumors are classified into primary or secondary brain tumors.

4. PERFORMANCE ANALYSIS

The performance of the proposed technique is simulated in the Python environment with these system requirements: processor - Intel core i7, Operating System - Windows 10 (64 bit), and RAM - 16GB. The performance of the developed CNN-VGG-16 technique is analyzed in terms of segmentation and classification with actual features after selecting the features. The performance of the proposed CNN-VGG-16 is compared with the existing techniques like CNN, VGG-net, and ResNet. The outcomes of the segmentation and classification are discussed below as subsections.

4.1. Evaluation for Segmentation

The segmentation process using CNN with actual features is analyzed by performance metrics like Dice Score Coefficient (DSC) and Hausdorff Distance (HD) for various tumor regions such as Tumor Core (TC), Enhancing Tumor (ET), and Whole Tumor (WT). The CNN involves analyzing features extracted from the image to process tasks like segmenting tumors. VGG-16 stacked multiple convolutional and FC layers efficiently capture data and classify the tumors. Table 1 and Fig. 5 represent the performance of the segmentation process along with actual features. The proposed CNN-VGG-16 method for the segmentation process accomplishes DSC of 98%, 98.5%, and 97% for various dice coefficient values, resulting in higher performance than the existing techniques.

Table 1. Performance of segmentation with actual features

Methods	DSC (%)			HD (mm)		
	TC	ET	WT	TC	ET	WT
CNN	80	70	85	25.4	18.89	20.11
VGG-net	89	81	90	17.45	15.36	18.19
ResNet	94	95	91	16.77	13.49	15.99
CNN-VGG-16	98	98.5	97	12.37	11.58	17.54

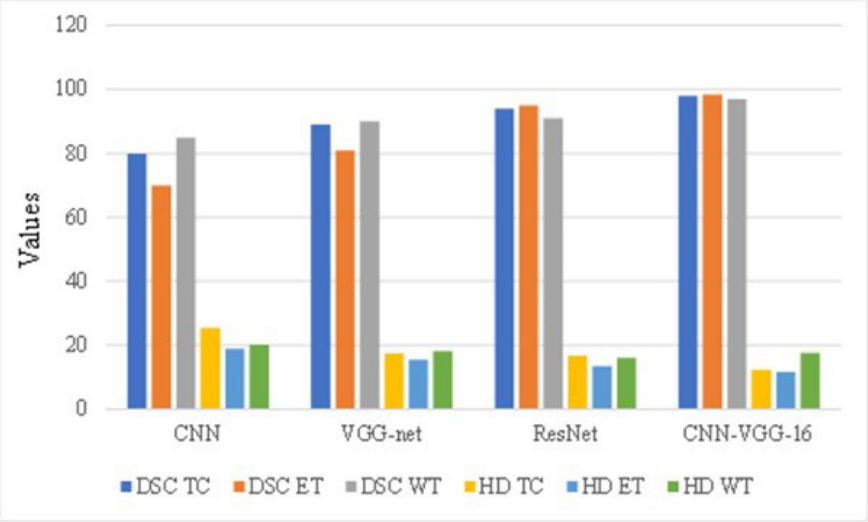


Fig. 5. Performance of segmentation with actual features

The segmentation process by CNN after feature selection is analyzed by performance metrics like DSC and HD with various tumor regions such as TC, ET and WT. Table 2 and Fig. 6 represent the performance of the segmentation process after feature selection. The proposed CNN-VGG-16 method for the segmentation process accomplishes DSC of 98%, 98.5%, and 97% for various dice coefficient values, which is superior to the existing techniques.

Table 2. Performance of segmentation after feature selection

Methods	DSC (%)			HD (mm)		
	TC	ET	WT	TC	ET	WT
CNN	82.5	72	86.6	26.57	21.81	21.99
VGG-net	92	82.9	91.2	18.78	17.24	19.59
ResNet	96.9	97	93.5	19	16.47	17.85
CNN-VGG-16	99.4	99.5	99.7	14.14	13.53	18.63

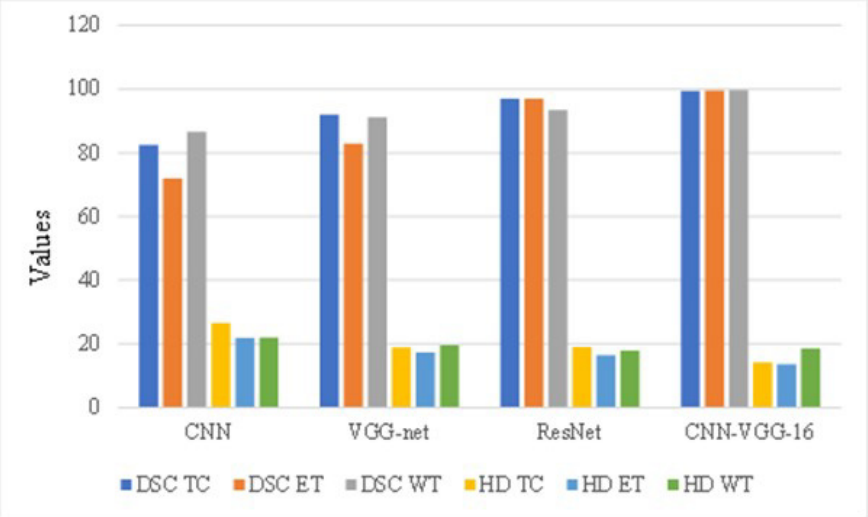


Fig. 6. Performance of segmentation after feature selection

4.2. Evaluation for Classification

The classification process by CNN-VGGnet with default features is analyzed through performance metrics like Accuracy, Precision and F-measure with various tumor regions like TC, ET, and WT. Table 3 and Fig. 7 represent the performance of the classification process with actual features. The proposed CNN-VGG-16 method for a classification process achieves an accuracy of 98%, 94%, and 91% for various dice coefficient values, being comparatively more robust than the existing techniques.

Table 3. Represents the performance of classification with default features

Methods	Accuracy (%)			Precision (%)			F-measure (%)		
	TC	ET	WT	TC	ET	WT	TC	ET	WT
CNN	85	77	82	80	69	84	90	91	94
VGG-net	92	88	88	87	78	90	94	94	96
ResNet	95	92	89	92	87	94	96	95	96
CNN-VGG-16	98	94	91	93	89	97	97	96	97

https://doi.org/10.31436/iiumej.v25i2.2963

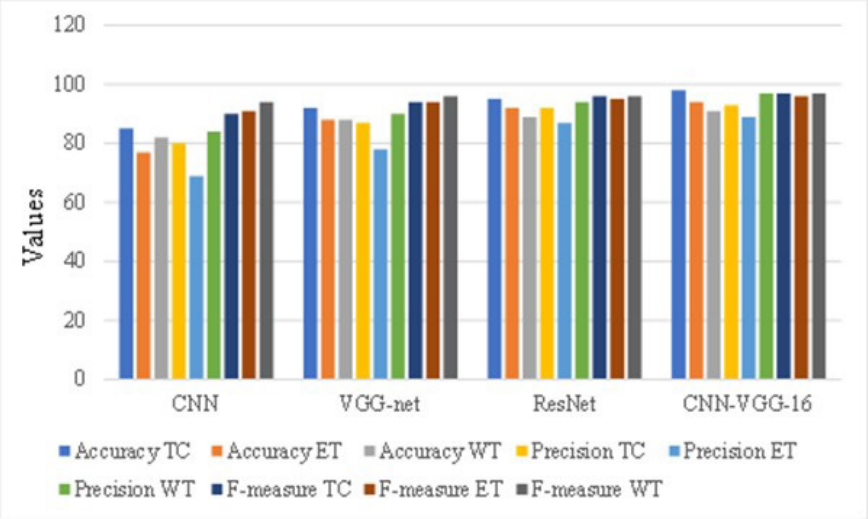


Fig. 7. Performance of classification with actual features

The classification process by the CNN-VGG-16 method after feature selection is analyzed on the basis of performance metrics like accuracy, precision, and F-measure with various tumor regions like TC, ET, and WT. Table 4 and Fig. 8 represent the performance of the classification process after feature selection. The proposed CNN-VGG-16 method outperforms the previous methods for the classification process by achieving an accuracy of 99.6%, 95.3%, and 94% for various dice coefficient values.

Table 4. Represents the performance of classification after feature selection

Methods	Accuracy (%)			Precision (%)			F-measure (%)		
	TC	ET	WT	TC	ET	WT	TC	ET	WT
CNN	87.1	79	84	81.8	70.3	85.9	92.7	92.6	96.8
VGG-net	93.8	89.1	90.2	88.4	80.5	92.7	96.4	95.4	98.2
ResNet	96.2	93.3	91.7	94.1	89.6	95.1	98.9	96.8	97.7
CNN-VGG-16	99.6	95.3	94	94.2	91.4	99	99.7	98.3	99

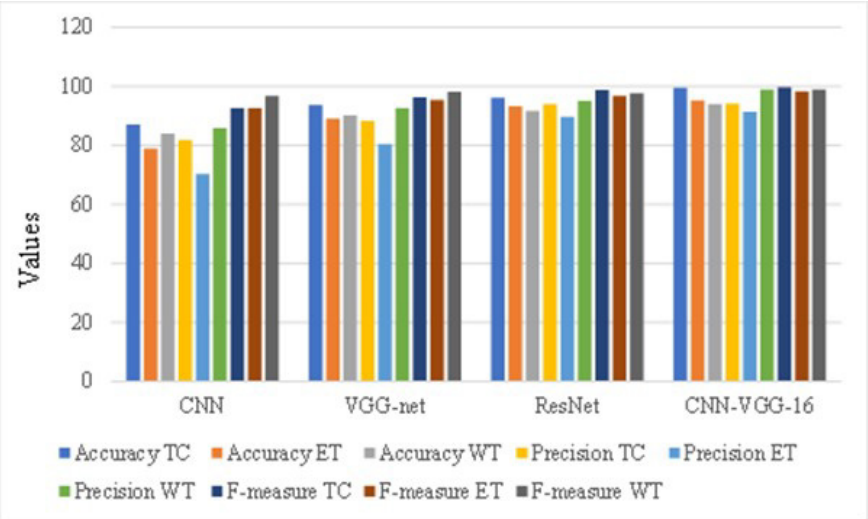


Fig. 8. Represents the performance of classification after feature selection

4.3. Comparative Analysis

The comparative analysis of the proposed technique is described in this section. The performance of the proposed CNN-VGG-16 is compared to existing models like softmax ensemble methods [16] and the SVM-RBF classifier [18] in terms of accuracy, specificity, and sensitivity. As shown in Table 5, the proposed method achieves an accuracy of 99.6%, which is comparatively higher than other existing methods, such as softmax ensemble, which attains 99.1%, and SVM-RBF, which attains 97.70%.

Table 5. Comparative Analysis with Other Methods

Author	Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
Sharif et al. [16]	Softmax ensemble	99.1	99.00	100
Rajinikanth et al. [18]	SVM-RBF	97.70	98.00	97.25
Proposed method	CNN-VGG-16	99.6	99.00	99.7

5. CONCLUSION

The early detection of brain tumors is necessary to aid doctors in treating cancer patients to increase their survival rate. In this research, the transfer learning-based approach is proposed to resolve the overfitting problem in classification. The BraTS – 2018 dataset is utilized in this research for segmentation and classification. Batch normalization is utilized in this experiment for data pre-processing, which normalizes the data in terms of batches. The CNN segments the images into specific parts, which makes classification effective, and the relevant features are extracted by the conv layer of CNN. Next, the AWO algorithm is utilized for effective feature selection, which selects relevant features for classification. After selecting relevant features, they are classified by VGG-16 with high accuracy. The performance of the proposed CNN-VGG-16 technique is analyzed on various tumor regions like TC, ET, and WT. The proposed method achieves a Dice score accuracy of 99.6%, 95.35% and 94%, outperforming the previously existing techniques like CNN, VGG-net, and ResNet. In the future, the hyperparameter tuning of the classification model will be performed to improve classification efficiency.

REFERENCES

[1] Dang K, Vo T, Ngo L, Ha H. (2022) A deep learning framework integrating MRI image preprocessing methods for brain tumor segmentation and classification. IBRO Neuroscience Reports, 13:523-532.

[2] Pitchai R, Supraja P, Victoria AH, Madhavi M. (2021) Brain tumor segmentation using deep learning and fuzzy K-means clustering for magnetic resonance images. Neural Processing Letters, 53(4):2519-2532.

[3] Ramesh S, Sasikala S, Paramanandham N. (2021) Segmentation and classification of brain tumors using modified median noise filter and deep learning approaches. Multimedia Tools and Applications, 80(8):11789-11813.

[4] Hossain A, Islam MT, Rahman T, Chowdhury MEH, Tahir A, Kiranyaz S, Mat K, Beng GK, Soliman MS. (2023) Brain Tumor Segmentation and Classification from Sensor-Based Portable Microwave Brain Imaging System Using Lightweight Deep Learning Models. Biosensors, 13(3):302.

[5] Das S, Bose S, Nayak GK, Satapathy SC, Saxena S. (2022) Brain tumor segmentation and overall survival period prediction in glioblastoma multiforme using radiomic features. Concurrency and Computation: Practice and Experience, 34(20):e6501.

[6] Atia N, Benzaoui A, Jacques S, Hamiane M, Kourd KE, Bouakaz A, Ouahabi A. (2022) Particle Swarm Optimization and Two-Way Fixed-Effects Analysis of Variance for Efficient Brain Tumor Segmentation. Cancers, 14(18):4399.

[7] Chahal PK, Pandey S. (2021) A hybrid weighted fuzzy approach for brain tumor segmentation using MR images. Neural Computing and Applications, 35(3):23877-23891.

[8] Masood M, Nazir T, Nawaz M, Mehmood A, Rashid J, Kwon H-Y, Mahmood T, Hussain A. (2021) A Novel Deep Learning Method for Recognition and Classification of Brain Tumors from MRI Images. Diagnostics, 11(5):744.

[9] Latif G, Brahim GB, Iskandar DNFA, Bashar A, Alghazo J. (2022) Glioma Tumors' Classification Using Deep-Neural-Network-Based Features with SVM Classifier. Diagnostics, 12(4):1018.

[10] Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D. (2021) A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network. Healthcare, 9(2):153.

[11] Khan AR, Khan S, Harouni M, Abbasi R, Iqbal S, Mehmood Z. (2021) Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification. Microscopy Research and Technique, 84(7):1389-1399.

[12] Aboussaleh I, Riffi J, Mahraz AM, Tairi H. (2021) Brain Tumor Segmentation Based on Deep Learning's Feature Representation. Journal of Imaging, 7(12):269.

[13] Sasank VVS, Venkateswarlu S. (2021) Brain tumor classification using modified kernel based softplus extreme learning machine. Multimedia Tools and Applications, 80(9):13513-13534.

[14] Reddy AS, Reddy PC. (2021) MRI brain tumor segmentation and prediction using modified region growing and adaptive SVM. Soft Computing, 25(5):4135-4148.

[15] Balamurugan T, Gnanamanoharan E. (2023) Brain tumor segmentation and classification using hybrid deep CNN with LuNetClassifier. Neural Computing and Applications, 35(6):4739-4753.

[16] Sharif MI, Li JP, Amin J, Sharif A. (2021) An improved framework for brain tumor analysis using MRI based on YOLOv2 and convolutional neural network. Complex & Intelligent Systems, 7(4):2023-2036.

[17] Decuyper M, Bonte S, Deblaere K, Holen RV. (2021) Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma. Computerized Medical Imaging and Graphics, 88:101831.

[18] Rajinikanth V, Raj ANJ, Thanaraj KP, Naik GR. (2020) A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection. Applied Sciences, 10(10):3429.

[19] Tandel GS, Tiwari A, Kakde OG. (2021) Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. Computers in Biology and Medicine, 135:104564.

[20] Punn NS, Agarwal S. (2021) Multi-modality encoded fusion with 3D inception U-net and decoder model for brain tumor segmentation. Multimedia Tools and Applications, 80(20):30305-30320.

[21] Agrawal P, Katal N, Hooda N. (2022) Segmentation and classification of brain tumor using 3D-UNet deep neural networks. International Journal of Cognitive Computing in Engineering, 3:199-210.

[22] Raza A, Ayub H, Khan JA, Ahmad I, Salama AS, Daradkeh YI, Javeed D, Ur Rehman A, Hamam H. (2022) A Hybrid Deep Learning-Based Approach for Brain Tumor Classification. Electronics, 11(7):1146.

[23] Haq EU, Jianjun H, Huarong X, Li K, Weng L. (2022) A hybrid approach based on deep cnn and machine learning classifiers for the tumor segmentation and classification in brain MRI. Computational and Mathematical Methods in Medicine, 2022:6446680.

[24] BraTS 2018 dataset link: https://ioft-data.engin.umich.edu/brats-2018-dataset/

[25] Srinivas C, Prakash KSN, Zakariah M, Alothaibi YA, Shaukat K, Partibane B, Awal H. (2022) Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images. Journal of Healthcare Engineering, 2022:3264367.

[26] Cao D, Xu Y, Yang Z, Dong H, Li X. (2023) An enhanced whale optimization algorithm with improved dynamic opposite learning and adaptive inertia weight strategy. Complex & Intelligent Systems, 9(1):767-795.

https://doi.org/10.31436/iiumej.v25i2.3044

MEMETIC ALLIGATOR OPTIMIZATION ALGORITHM FOR OPTIMAL THERMOREGULATORY CONTROL IN PIPING SYSTEMS

WENG-HOOI TAN, NORAMALINA ABDULLAH*, JUNITA MOHAMAD-SALEH

School of Electrical & Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal, 14300, Penang, Malaysia

**Corresponding author: eenora@usm.my*

(Received: 27 September 2023; Accepted: 4 March 2024; Published online: 15 July 2024)

ABSTRACT: Applying optimization techniques to control systems remains a challenging task. Control system technology is emerging rapidly due to the high demands for commercialization in engineering and industrial fields, but existing optimization techniques are considered weak to cope with increasingly complex control problems. More powerful optimization techniques are urgently needed to catch up with the prerequisites for optimal control. Hence, this research is dedicated to the development of an improved optimization algorithm to solve the optimal thermoregulatory control problem for piping systems in a more efficient manner. As a research outcome, the Memetic Alligator Optimization (MeAgtrO) algorithm is proposed. On top of the mathematical hunting and relocating mechanisms, MeAgtrO adds several evolutionary operators that replicate satiety awareness, mating, generational alternation, and dispersed hunting. Unlike the standard optimizer, which only emphasizes global and local search transitions, these improved variants give the ability to shuffle, swap, replace, and disperse agent information for greater flexibility. Upon application to a simulated piping system to optimally control thermoregulation processes, MeAgtrO statistically outperformed the other compared algorithms, showing 100% accuracy, 99.99% precision, and 99.99% robustness in minimizing the tracking error, response time, and equipment burden of the system. MeAgtrO has been shown to have high processing speed for optimal application control, which corresponds to its superior convergence speed to stabilize at 40% of iterations. While showing satisfactory clustering properties, MeAgtrO also demonstrated the best step response with a rise time of 0.40s, settling time of 0.99s, 0 tracking error, 0% overshoot, and 0% undershoot.

ABSTRAK: Penerapan teknik optimasi pada sistem kawalan adalah tugas mencabar. Teknologi sistem kawalan berkembang pesat disebabkan oleh permintaan tinggi bagi komersialisasi dalam bidang kejuruteraan dan industri, tetapi teknik optimasi sedia ada dianggap lemah dalam mengatasi masalah kawalan yang semakin kompleks. Teknik optimasi lebih kuat diperlukan dengan segera untuk memenuhi prasyarat kawalan optimal. Oleh itu, penyelidikan ini bertujuan bagi membangunkan algoritma optimasi yang lebih baik bagi menyelesaikan masalah kawalan termoregulator optimum pada sistem paip dengan lebih cekap. Penyelidikan ini mencadangkan algoritma Memetic Alligator Optimization (MeAgtrO). Selain dari mekanisme pemindahan dan pemburuan matematik, MeAgtrO memiliki beberapa pengendali evolusinari yang menggandakan kesedaran kenyang, mengawan, alternasi generasi, dan pemburuan tersebar. Berbeza dengan pengoptimuman standard, penekanan hanya pada peralihan global dan carian tempatan, ini membaiki varian dengan memberikan keupayaan menyusun semula, menukar, mengganti, dan menyebar maklumat ejen bagi fleksibiliti yang lebih besar. Apabila digunakan pada sistem paip bersimulasi bagi mengawal proses termoregulasi secara optimum, MeAgtrO secara statistik mengatasi algoritma lain, menunjukkan ketepatan 100%, kejituan 99.99%, dan kekuatan 99.99% dalam mengurangkan kesilapan pengesanan, masa tindak balas, dan beban sistem peralatan. MeAgtrO telah terbukti mempunyai kelajuan pemprosesan yang tinggi bagi aplikasi kawalan optimum, sejajar dengan kelajuan konvergensi yang lebih bagus bagi menstabilkan iterasi 40%. Di samping menunjukkan sifat kelompok yang memuaskan, MeAgtrO juga memiliki respons langkah terbaik kenaikan masa 0.40s, masa penyelesaian 0.99s, ralat pengesanan 0%, lebih sasaran 0%, dan kurang sasaran 0%.

KEYWORDS: *bio-inspired, metaheuristic, optimization, thermoregulation in piping system, Matlab-Simulink*

1. INTRODUCTION

In the field of appropriate thermoregulatory control in pipe systems, the constant growth of cutting-edge technology needs increasingly efficient optimizations to improve application performance. This growing demand highlights the need for more advanced optimization methodologies, motivating researchers to focus their efforts on developing strong optimization algorithms that are customized to the particular problems provided by modern technology applications. Following a thorough survey, it is clear that the immediate applications of optimization techniques in the context of optimal thermoregulatory control extend to a variety of areas, including optimal equipment design, optimal equipment diagnostics, optimal system operation, optimal decision-making under uncertainty, and optimal system planning. These applications act as critical motivators, pushing researchers to develop and expand optimization algorithms, propelling the technology to new heights, particularly in the field of thermoregulatory control in pipe systems.

In the context of thermoregulatory control, optimal equipment design entails meticulously optimizing building procedures to minimize quantity while maximizing uniformity using particular statistical criteria [1–4]. Furthermore, optimum equipment diagnostics, as a subject of control engineering, plays an important role in system monitoring, identifying failure kinds, and detecting faults. The use of optimal equipment diagnostics, particularly in the building of fault classification models, not only improves fault diagnosis accuracy but also decreases training time, allowing for faster recovery of the pipe system by replacing problematic components. The technology's application in the medical industry for automated computer-aided diagnostics complements the overall subject of effective thermoregulatory management in pipe systems. This application aims to remove human mistakes, resulting in more accurate and prompt early detection and treatment of chronic diseases, as demonstrated by computer-aided diagnosis in pathology and radiology.

Optimal system functioning within the context of thermoregulatory control ensures accurate and efficient job performance, often reducing exposure and increasing operational capacity. Simultaneously, optimum decision-making seeks to hedge and maximize average profit while minimizing risk, demanding a thorough examination of profitability analysis, operational efficiency, and realistic possibilities, especially in the context of thermoregulatory management. In contrast, ideal system design for thermoregulatory control entails thorough optimization of construction, equipment layout, material procurement, facility expansion, quality assurance, and cost-effectiveness. This optimization ensures that all components not only meet but exceed user expectations, improving the functioning and integration of subsystem components within the larger context of thermoregulatory control in pipe systems. Recent applications of system planning optimization span a wide range of sectors, including academics, medicine, economics, engineering, politics, and academia, demonstrating its flexibility and significance in optimizing thermoregulatory control systems for pipelines.

Some examples of related technologies include PID (Proportional-Integral-Derivative) controllers, which are commonly used in industrial control systems to regulate process variables like temperature, flow, and pressure; fuzzy logic controllers, which use fuzzy logic to deal with uncertain or imprecise data; neural networks, which are used for adaptive control and complex pattern recognition in systems where mathematical models are difficult to create; and so on. The existing approach has issues with efficiency in optimizing accuracy,

handling complexity and ambiguity, and limited flexibility. In very dynamic or non-linear systems, many of the current techniques, such as PID controllers, could not work as well. Complex and uncertain problems can be solved using computationally demanding fuzzy logic and neural networks, which need large amounts of data. To manage ever-increasingly complicated control systems and efficient optimization, there is an ongoing need for algorithms with greater accuracy, resilience, and convergence rates, even if these approaches are typically effective. To overcome these obstacles, this study developed an algorithm for thermoregulation in pipe systems that integrates evolutionary computation with other techniques for better control. We entitled the newly developed algorithm the Memetic Alligator Optimization (MeAgtrO) algorithm, where the term "memetic" stands for a deeper mimicry of the alligator's ecological behaviours in nature from a broader evolutionary perspective. As novelty correlates with major finding, the novelty of MeAgtrO lies in the adoption of novel decision coefficients (satiety), the addition of evolutionary breeding (mating) operators, incorporation of agent replacement (alternation of generation) strategies, the utilization of multi-reference global best solution (scatter hunting) strategies, and the modification of adaptive coefficient (enhanced search behaviours).

When compared to other algorithms, MeAgtrO showed the greatest statistical results in terms of the best, mean, and standard deviation (SD). This is one of the main reasons why MeAgtrO stands out. It performs consistently and reliably well. For control applications in industry, where accuracy and dependability are of the utmost importance, it attained extraordinary resilience, reliability, and accuracy. However, other than that, MeAgtrO demonstrated strong clustering capabilities. If the algorithm does a good job of classifying data or solutions, it should be able to optimize more efficiently and effectively. Among the algorithms that were compared, MeAgtrO had the best convergence rate. In real-time control applications, the algorithm's ability to discover the best solution rapidly is crucial, hence a quicker convergence rate is preferred. When compared to competing algorithms, MeAgtrO appears to have performed better in all important respects that matter much for this particular thermoregulation application in pipe systems. Having these qualities is crucial for optimizing control systems, especially in engineering and industrial settings where dependability and efficiency are paramount. The algorithm was chosen for this investigation because it outperformed others in these critical areas.

This research paper is outlined as follows: Section 1 introduces this research work, Section 2 explains the methodology of the developed MeAgtrO, Section 3 describes the modelling and simulation of control piping system, Section 4 analyses the results, and Section 5 concludes the study.

2. PROPOSED MeAgtrO

The Memetic Alligator Optimization (MeAgtrO) algorithm is proposed as an augmented optimization method for application purposes. It is formulated to execute a search operation that combines the hunting phase and the relocating phase. The overall equation is expressed as follows:

https://doi.org/10.31436/iiumej.v25i2.3044

$$\begin{cases} X_i^{t+1} &= X_i^t + F_{1,i}\overrightarrow{V_{hunting_i}}^t + (1 - F_{1,i})\overrightarrow{V_{relocating_i}}^t \\ \overrightarrow{V_{hunting_i}}^t &= \left[\overrightarrow{V_{ps_i}}^t + F_{2,i}\overrightarrow{V_{c_i}}^t\right] \\ \overrightarrow{V_{relocating_i}}^t &= \left[\overrightarrow{V_{t/h_i}}^t + \overrightarrow{V_{w_i}}^t\right] \\ \overrightarrow{V_{PS_i}}^t &= (1 - A)\left(X_g^t - X_i^t\right) \\ \overrightarrow{V_{C_i}}^t &= (1 + M)\left(X_g^t - \left(X_i^t + \overrightarrow{V_{ps_i}}^t\right)\right) \\ \overrightarrow{V_{T/H_i}}^t &= (1 - |C|)\left(2r_1 X_{i,lbest}^t - X_i^t\right) \\ \overrightarrow{V_{W_i}}^t &= \left(1 \pm e^B\left(L(\cos(2\pi B)) + (1 - L)\left(\sin(2\pi B)\right)\right)\right) \bullet \left(X_{i,lbest}^t - X_i^t\right) \end{cases} \quad (1)$$

where $i \in \{1, 2, \ldots, N_p\}$ denotes the index of population, N_p is the population number, $t \in \{0, 1, \ldots, t_{max} - 1\}$ denotes the index of iteration, t_{max} is the maximum number of iterations, X represents the position point or solution, X_g^t denotes the global best solution that agents (population) had visited so far along the iteration, r_1 is the random number within 0 and 1, $\overrightarrow{V_{hunting_i}}^t$ is a vector representing the hunting phase, $\overrightarrow{V_{relocating_i}}^t$ is a vector representing the relocating phase, $\overrightarrow{V_{PS_i}}^t$ imitates the purse seining behaviour, $\overrightarrow{V_{C_i}}^t$ imitates the catching behaviour, $\overrightarrow{V_{T/H_i}}^t$ imitates the traveling and homing instinct behaviour, $\overrightarrow{V_{W_i}}^t$ imitates the wave vector, $F_{1,i}$ is the hunting flag of the i^{th}agent, $(1 - F_{1,i})$ is the relocating flag opposed to $F_{1,i}$, and $F_{2,i}$ is the catching flag of the i^{th}agent.

For a more refined search mechanism, MeAgtrO proposes the following adaptive coefficients:

$$\begin{cases} A = (2r_2 - 1)\left(1 - \frac{t}{t_{max}}\right)^3 \\ M = e\left(\frac{f\left(X_{gbest}^t\right) - f(X_i^t)}{\max(f(X^t)) - \min(f(X^t))}\right) \\ C = 9(2r_3 - 1)\left(1 - \left(\frac{t}{t_{max}}\right)^{3\left(1 - \frac{t}{t_{max}}\right)}\right)^3 \\ B = -9r_4\left(1 - \frac{t}{t_{max}}\right)^3 \end{cases} \quad (2)$$

where r_2, r_3 and r_4 are the random numbers between 0 and 1.

The coefficient adaptation strategy primarily aims to achieve the best results for the global market by refining both global and local search capabilities while smoothing the transition from exploration to exploitation. Coefficient A assigns the possible purse seining distance between the agent (alligator) and the global best solution, toefficient C represents the relocating range towards or away from the home territory (individual local best solution), and coefficient B denotes the pulling force exerted on the agent or the strength of the agent against the wave vector. In MeAgtrO, $\overrightarrow{V_{ps_i}}^t$ becomes the only operator responsible for global exploration and global exploitation, whose characteristics depend on the coefficient A value. The purse seining operation needs to be ideally accelerated. In theory, when the coefficient $|A| > 0.5$, it is more inclined to global exploration, and conversely, when the coefficient

https://doi.org/10.31436/iiumej.v25i2.3044

$|A| \leq 0.5$, it is more inclined to global exploitation. Given that agents require fewer iterations to stabilize, while ensuring that they do not skip any search types throughout their execution, the maximum available value of coefficient A is gradually tapered off at a faster rate to effectively speed up the transition from exploration to exploitation. The faster pace of purse seining behavior now allocates less global exploration and more global exploitation, resulting in a more desirable trade-off between global exploration and global exploitation. By contrast, the coefficient C is adapted to have a slower declination velocity to ensure that the agents can have enough iterations to process the local search operation, giving them more opportunities to escape local optima while searching for any possible global optimal region. Depending on the situation, this approach can positively speed up convergence or negatively hinder optimization output. The coefficient B is thereby adapted to have a quicker reduction, just to synchronize to some extent to cover (balance) the negative effect caused by the coefficient C. Overall, the adaptation of these important coefficients is expected to achieve better optimization efficiency.

2.1. Satiety Awareness

The search agent decides whether to enter the hunting or relocation phase during execution. For more reliable decision-making methods instead of random selection, MeAgtrO proposes decision-making approaches based on a novel decision coefficient S, where S theoretically represents the satiety of the agents (alligators). S_i^t, on the other hand, can be denoted as the satiety coefficient of the i^{th} agent at the t^{th} iteration. MeAgtrO utilizes S_i^t to control the selection methods of the i^{th} agent in a smarter way, mainly via $F_{1,i}$ and $F_{2,i}$, which can be expressed mathematically as follows:

$$\begin{cases} F_{1,i} = \begin{cases} 1, & \text{if } S_i^t \leq r_1 \\ 0, & \text{otherwise} \end{cases} \\ F_{2,i} = \begin{cases} 1, & \text{if } S_i^t \leq r_1 \text{ and } i \in J \\ 0, & \text{otherwise} \end{cases} \\ S_i^{t+1} = e^{(-r_5 S_i^t)} \end{cases} \tag{3}$$

where:

$$\begin{cases} J \in \{i_{R\{1\}}, i_{R\{2\}}, \ldots, i_{R\{N_s\}}\} \\ N_s = \left\lfloor \text{length}\left(\text{find}(S_{\text{sorted}}^t > 0)\right)/4 \right\rfloor \\ R = \text{find}(S_{\text{sort}}^t > 0) \\ S_{\text{sort}}^t = \underset{\text{ascend}}{\text{sort}}\left(F_{1,1\to N_p}^t \bullet S_{1\to N_p}^t\right) \end{cases} \tag{4}$$

where $e^{(-r_5 S_i^t)}$ returns the exponential value, S_{sort}^t denotes satiety values sorted in ascending order which corresponds to search agents sorted from lowest satiety to highest satiety, R is a function that identifies the index of the search agent that satisfies the condition $S_{sort}^t > 0$ (by which search agents that do not execute $\overrightarrow{V_c}$ at the t^{th} iteration will be excluded), N_s is the estimated number of hunting agents allowed to be included in executing $\overrightarrow{V_c}$ at that iteration, and J represents the index of the search agent that is finally qualified to execute $\overrightarrow{V_c}$ after rounds of selection.

The initial coefficient value of S_i^t is 1. When $S_i^t \approx 1$, it indicates that the i^{th} agent is full, and when $S_i^t \approx 0$, it indicates that the i^{th} agent is starving. From mathematical perspective, S_i^t gradually decreases with iterations, simulating the fact that creatures lose their satiety over

https://doi.org/10.31436/iiumej.v25i2.3044

time if they do not feed. The decreasing degree of S_i^t can refer to Equation (3). Extending the explanation, this coefficient plays an important role in deciding whether the i^{th} agent biases the hunting phase or the relocating phase, indirectly via $F_{1,i}$. It should be noted from the formula of $F_{1,i}$ that the smaller the value of the coefficient S_i^t, the greater the chance of the i^{th} agent to select the hunting phase. This brings up another implication that the agent urgently needs to restore its satiety. Overall, this coefficient S can actually assign a perfect balance of trade-offs between the hunting phase and the relocation phase in any given optimization problem. It maintains the systematic characteristics to some extent, but at the same time inherits the minimal required randomness, which is more capable to decide which favourable phase to enter according to the current situation. The main concept, in turn, allows the agents (alligators) to regain satiety from feeding, forcing the agent to leave the hunting phase and enter the relocating phase at the next iteration. Mathematically, the i^{th} agent that executed the catching mechanism ($\overrightarrow{V_{c_i}}^t$) restores satiety, thereby resetting S_i^{t+1} to 1. But as specified, only a small fraction (usually ¼ fraction) of agents can perform the catching mechanism in one iteration, hardly imitating the fact that alligators systematically "take turns" while catching prey in sequence. To more faithfully mimic alligators as social predators that tolerate each other while hunting, F_2 rationally assigns agents with lower satiety to have a greater chance of preferentially executing catching behavior in Equation (3). The proposed decision-making or selection method using this novel decision coefficient (S) is expected to achieve better optimization performance in terms of global best results and convergence speed.

2.2. Mating

Inspired by the biological reproduction process, MeAgtrO proposes to add on an evolutionary breeding operator. For the breeding purpose, the algorithm separates population into half male ($X_{1\to N_p/2}^t$) and half female ($X_{1+N_p/2\to N_p}^t$). At the beginning of each iteration, the agents execute breeding operations, simulating the arrival of the mating season. From a mathematical point of view, the breeding operator allows for the crossover of information between agents of different genders, which adds an extra step to the algorithm to generate a bunch of talented offspring that inherit the predominant (genetic) information. Agents of different genders selected to execute the breeding operation are referred to as parents. For the next breeding step, all solutions of the offspring are mutated to a certain extent to adapt to the global best solution (i.e., the hunting ground). The proposed breeding operation can be expressed mathematically as follows:

$$\begin{cases} X_{off,gh}^t &= (\vec{r}_1 \geq 0.2) \bullet X_{cross,gh}^t + (\vec{r}_1 < 0.2) \bullet X_{mutate,gh}^t \\ X_{mutate,gh}^t &= X_{gbest}^t + \xi \left| X_{cross,gh}^t - X_{gbest}^t \right| \\ X_{cross,gh}^t &= (\vec{r}_2 \leq 0.5) \bullet X_g^t + (\vec{r}_2 > 0.5) \bullet X_h^t \\ \xi &= 1 - e^{\left(-\left(\frac{1-t}{t_{max}}\right)^3\right)} \end{cases} \quad (5)$$

where $\vec{r}$ is the random array within 0 and 1, $|\square|$ denotes the absolute function, $X_g^t \in \left\{X_{i=1}^t, X_{i=2}^t, \ldots, X_{i=N_p/2}^t\right\}$ is the g^{th} male agent at t^{th} iteration, $X_h^t \in \left\{X_{i=1+N_p/2}^t, X_{i=2+N_p/2}^t, \ldots, X_{i=N_p}^t\right\}$ is the h^{th} female agent at t^{th} iteration, and $X_{off,gh}^t$ is the final offspring solution reproduced by the g^{th} male agent and the h^{th} female agent at t^{th} iteration. Note that agents of different genders that have been selected as parents will not be selected again for the upcoming crossover processes at one iteration, which mean that the selected g^{th} male agent and h^{th} male agent will not be selected as the next parent candidates

https://doi.org/10.31436/iiumej.v25i2.3044

for consecutive crossover processes at the same iteration. This concept ensures that every agent can become a parent candidate without being excluded or duplicated. For rational selection, g is selected based on the normal cumulative distribution of $\underset{\text{ascend}}{\text{sort}}\left(f\left(X^t_{1\to N_p/2}\right)\right)$, while h is selected based on the normal cumulative distribution of $\underset{\text{ascend}}{\text{sort}}\left(f\left(X^t_{1+N_p/2\to N_p}\right)\right)$. In a more specific way of interpretation, the breeding operator includes crossover and mutation processes. Referring to Equation (5), offspring theoretically inherit about 80% of the crossover information from the parents and develop about 20% of the mutational information. During the crossover process, it generates the new offspring by exchanging the information of two agents from different assigned genders, where the dimensional information represents the genes of the agents. This crossover of dimensional information has improved the escape ability of agents from the local optima to some extent, at the same time, covering any insufficiency in exploration operations. In theory, the male and female agents are both queue in an ascending order from best-performed to poorest-performed units. The selection is said to highly depend on the current fitness of the agents, as the agents select their partner via the normal cumulative distribution approach, where upper ranked units are more likely to select the upper ranked partner from different gender in the breeding operation. The crossover process between well-behaved agents (from different genders) is thought to produce offspring that inherit promising dimensional information that contributes to local absconding skills, global proximity, and accelerated progression. The crossover process is followed by the mutation process. The mutation process replaces approximately 20% of the content in the offspring solution (post-crossover) with global best information that deviates from the fluctuation range. This slightly diverges the offspring solution from the parent solutions, bringing the offspring closer to the prey positions to facilitate participation in the next hunting phase. Mathematically, the mutation process adds fluctuations and oscillations in terms of flexibility, accelerating the optimization process by refining exploration operations.

2.3. Alternation of Generation

Offspring generated from equation (5) are randomly assigned gender (either male or female). In MeAgtrO, every offspring can potentially replace the position point of its parent of the same gender if a better solution is provided. This is thought to eliminate less efficient solutions from the population to speed up the optimization process (convergence rate), and to some extent provide potential escape routes for the population to escape from local optima. Despite the little extra computational time spent during the comparison process, this approach is highly appreciated because the newly generated offspring inheriting better information can benefit greatly in the next search operation, potentially reaching a more optimal region, thereby guiding the entire population towards a better global best solution. This ultimately strengthens the clustering properties of the algorithm. Upon formulation, the agent replacement strategy at the t^{th} iteration can be expressed as:

$$\begin{cases} X^t_g = \begin{cases} X^t_g & \text{, if } r_6 \le 0.5 \text{ and } f\left(X^t_g\right) < f\left(X^t_{off,gh}\right) \\ X^t_{off,gh} & \text{, otherwise} \end{cases} \\ X^t_h = \begin{cases} X^t_{off,gh} & \text{, if } r_6 > 0.5 \text{ and } f\left(X^t_h\right) < f\left(X^t_{off,gh}\right) \\ X^t_h & \text{, otherwise} \end{cases} \end{cases} \tag{6}$$

When a new offspring is spawned and replaces its parent, its satiety ($S^t_{i=g|h}$) restarts from a full number 1, but does not inherit the satiety of its replaced parent, as both are considered

https://doi.org/10.31436/iiumej.v25i2.3044

to be independent units. In contrast, offspring that replace its parent will inherit the local best solution ($X^{t+1}_{i=g|h,lbest}$) to mimic the "home" of the designated parent candidate passed to its descendant. This is what is called inheriting the achievements of the older generation, benefiting the younger generation regardless of living conditions. However, if there is a better solution, it mathematically updates its current position point to the local best solution during the initial fitness comparison.

2.4. Scatter Hunting

There should usually be more than one prey in the natural environment. As for a more faithful imitation, a multi-reference position seems necessary. Therefore, a multi-reference global best solution strategy is proposed in MeAgtrO, which prepares four selectable prey positions ($X_{gbest} \in \{X_{gbest(1)}, X_{gbest(2)}, X_{gbest(3)}, X_{gbest(4)}\}$) for the population. Each agent randomly defines its global best solution from these four candidates, where the probability of selecting any one is ¼. The agents update solutions from $X_{gbest(1)}$ to $X_{gbest(4)}$ in ascending order: if a solution is found better than $X_{gbest(1)}$, $X_{gbest(1)}$ is replaced by that solution, otherwise it moves on to compare that solution with $X_{gbest(2)}$, and so on. The multi-reference global best solution can theoretically disperse the agents to the respective prey positions, thereby expanding the search range of the agents, mainly for more effective and efficient global exploration in the early stage of iteration. Though it is estimated that prey positions ($X_{gbest(1\to4)}$) will gradually cluster along the iterative update. These somewhat improve the performance in terms of global best results and convergence speed.

The following pseudocode presents the entire mechanism of MeAgtrO:

Input parameters
Initialize population (agents)
Calculate the objective fitness values for all agents
$t = 0$;
$X^t_{gbest(1\to4)}$ = best agent solutions
$X^t_{1\to N_p,lbest} = X_{1\to N_p}$
WHILE ($t < t_{max}$)
  FOR $i = 1$ TO $N_p/2$
    Evolutionary breeding operation: Equation (5)
    Agent replacement strategy: Equation (6)
  END FOR
  Population satiety: Equation (3)
  FOR $i = 1$ TO N_p
    Adaptive coefficients: Equation (2)
    Search operational mechanism: Equation (1)
    Calculate the objective fitness value of i^{th} agent

https://doi.org/10.31436/iiumej.v25i2.3044

Replace $X_{i,lbest}^{t}$ if i^{th} agent has a better solution

END FOR

Replace $X_{gbest(1\rightarrow4)}^{t}$ in sequence if any agent provides a better solution

$X_{gbest(1\rightarrow4)}^{t+1} = X_{gbest(1\rightarrow4)}^{t}$

$X_{1\rightarrow N_p,lbest}^{t+1} = X_{1\rightarrow N_p,lbest}^{t}$

$t = t + 1$

END WHILE

Global best objective fitness value = $f\left(X_{gbest}^{t=t_{max}}\right)$

Output $X_{gbest}^{t=t_{max}}$

Output $f\left(X_{gbest}^{t=t_{max}}\right)$

3. SIMULATION OF THERMOREGULATION IN PIPING SYSTEMS

For simulation, the model of the piping system is customized while referring to the literature [5]. The entire layout of the simulated piping system is displayed in Figure 1, consisting of water tank, heat exchanger, thermocouple, controller, gate valve, pipe, and wire. For a clearer image, Figure 2 shows an enlarged image of these widgets.

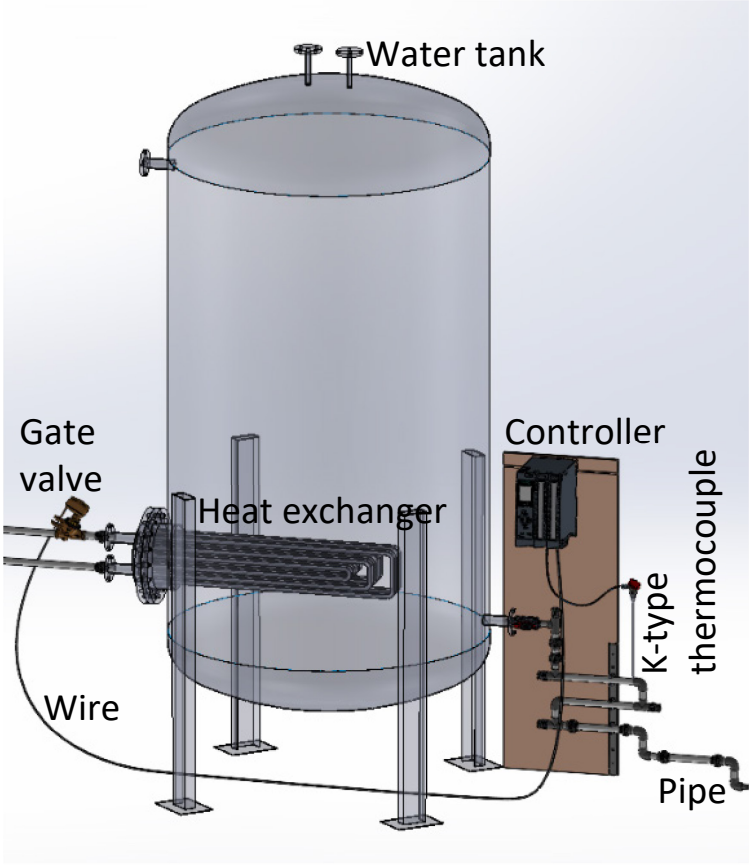


Figure 1. Layout of simulated piping system

https://doi.org/10.31436/iiumej.v25i2.3044

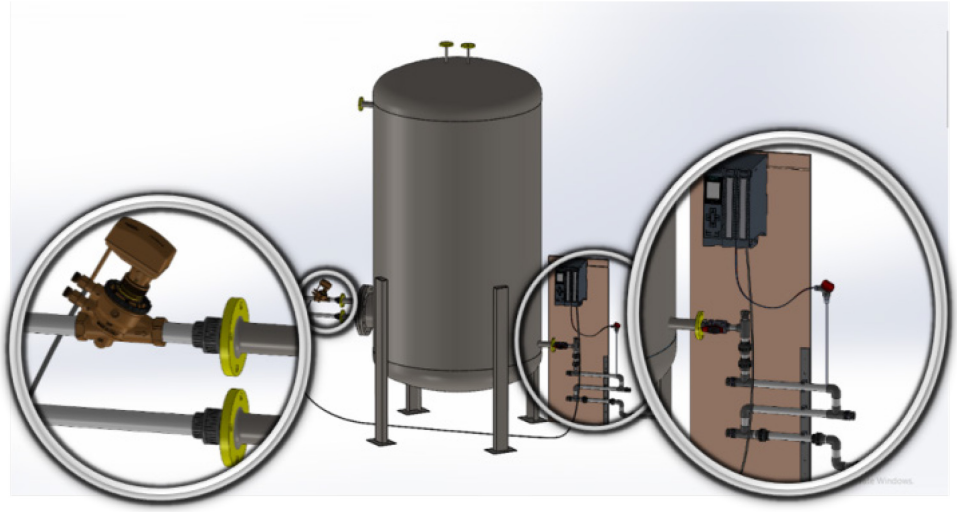


Figure 2. Enlarged layout of simulated piping system

The steam heater exchanger supplies sustainable heat to the fluid in the storage tank, while the fluid in the storage tank will flow to other equipment through the pipeline. The main objective of the simulation is to get the fluid in the pipe to reach and maintain the desired temperature. This is the so-called thermoregulation process. Considering the mathematical heat loss, the temperature of the fluid in the pipes is not exactly equal to the temperature in the tank. There are basically three types of heat loss: conduction, convection, and radiant heat loss that occur in the pipeline. These heat losses are strongly affected by the material and thickness of each composed layer in the insulated pipe. Therefore, the material selection of insulated pipeline must consider various factors, including its thermal insulation, electrical conductivity and corrosion resistance [6]. All composed material details can be reviewed from related articles and book chapters [7, 8]. For simulation, we constructed a pipeline consisting of a layer of fiberglass insulating protective cover and copper tubing. For ease of understanding, Figure 3 displays a cross-sectional view of the constructed insulated pipe, and Figure 4 illustrates the mechanism of the heat transfer process in this insulated pipe.

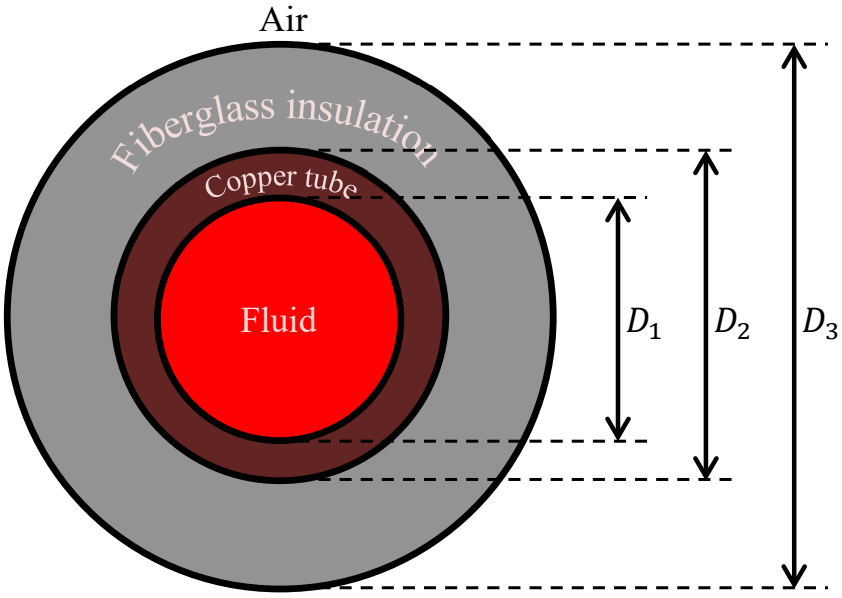


Figure 3. Cross-sectional view of insulated pipeline

https://doi.org/10.31436/iiumej.v25i2.3044

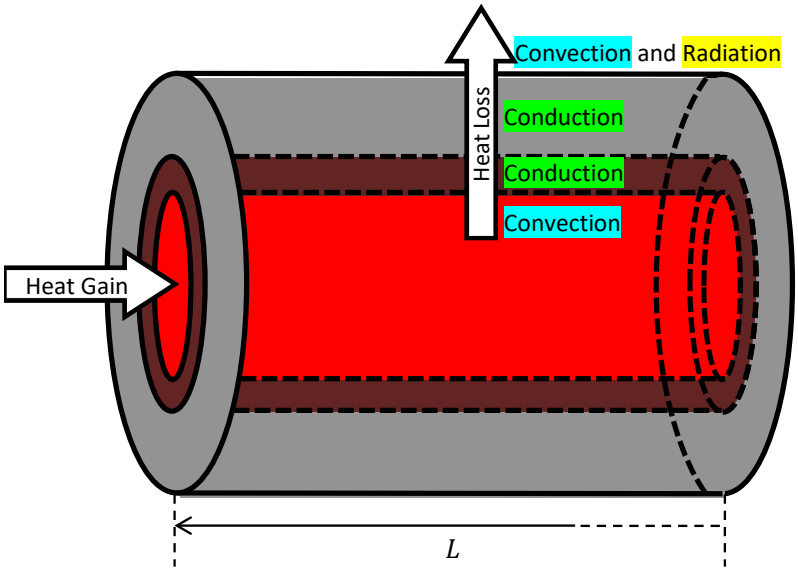


Figure 4. Mechanism of heat transfer process in insulated pipeline

From a mathematical perspective, due to net heat transfer, the operating temperature of the fluid inside the pipe in the time series domain can be expressed as:

$$T_{op}(\dot{t} + \dot{t}_{interval}) = T_{op}(\dot{t}) + \int_0^{\dot{t}_s=1} \frac{1}{\rho V c} \left[C_p T_{in}(\dot{t}) + \left(\frac{T_{op}(\dot{t}) - T_a(\dot{t})}{\frac{\ln(D_2/D_1)}{2\pi L k_{tube}} + \frac{\ln(D_3/D_2)}{2\pi L k_{insulation}} + \frac{1}{(\pi D_3 L) h_{air}} + \frac{1}{(\pi D_1 L) h_{fluid}}} \right) \right] d\dot{t} \quad (7)$$

where T represents the temperature, and hence T_{op} is the operating temperature of the fluid (unit: K) and T_{in} is the input temperature of the fluid when entering the pipe (unit: °C). $\dot{t} \in \{0, \dot{t}_{interval}, 2\dot{t}_{interval}, \dots, \dot{t}_{end}\}$ represents the index of time, where $\dot{t}_{end}$ is the simulation duration (unit: s), $\dot{t}_{interval}$ is the time interval per update (unit: s), and $\dot{t}_s$ is the timespan. Supposed that the fluid in the entire L-length pipe shares the same temperature of T_{op}, where T_{op} is initialized to T_a when $\dot{t} = 0$. During simulation, $t_{interval} = 0.01$ and $t_{end} = 10$, which allows the system to update the operating temperature every 0.01s from 0s to 10s. For convenience of explanation, the definitions and values of the remaining parameters are displayed in Table 1.

The simulated piping system designed for thermoregulation application is considered a feedback control system. To ensure that the fluid maintains the desired temperature when it reaches the desired point in the pipeline, a k-type thermocouple is placed at any corner end of the L-length pipeline, acting as a sensor to acquire temperature data at that point. The wire sends the information data obtained from the thermocouple to the controller. The controller computes the temperature difference and the percentage error, which then tunes the parameters and sends the signal out to the gate valve. After receiving the command signal from the controller, the gate valve adjusts the steam flow rate in the heat exchanger and the heat supplied to the water tank, thereby regulating the temperature of the fluid in the water tank. Note that the controller installs the specified optimization algorithm to handle the process of parameter optimization. The optimization process is repeated for a given number of iterations until the piping system reaches its optimal thermoregulation performance.

https://doi.org/10.31436/iiumej.v25i2.3044

Table 1. Detailed parameters in the pipeline

Indicator	Description	Value	Unit
Pipeline			
$\boldsymbol{L}$	Length of insulated pipeline	0.1	m
Fiberglass insulation			
$\boldsymbol{D_3}$	Outer diameter of insulation	0.1889	m
$\boldsymbol{k_{insulation}}$	Thermal conductivity of insulation	0.04	$^{W}/_{mK}$
Copper tube			
$\boldsymbol{D_1}$	Inner diameter of copper tube	0.0779	m
$\boldsymbol{D_2}$	Outer diameter of copper tube	0.0889	m
$\boldsymbol{k_{tube}}$	Thermal conductivity of copper tube	385	$^{W}/_{mK}$
Fluid			
$\boldsymbol{h_{fluid}}$	Heat transfer coefficient of fluid	100	$^{W}/_{m^2K}$
$\boldsymbol{\rho}$	Density of fluid	997	$^{kg}/_{m^3}$
$\boldsymbol{V}$	Volume of fluid	$\pi L\left(\frac{D_1}{2}\right)^2$	m^3
$\boldsymbol{c}$	Specific heat capacity of fluid	4200	$^{J}/_{kgK}$
Air			
$\boldsymbol{h_{air}}$	Heat transfer coefficient of air	10	$^{W}/_{m^2K}$
$\boldsymbol{T_a}$	Ambient temperature	296.15	K
Input			
$\boldsymbol{C_p}$	heat input coefficient	250	$^{W}/_{°C}$

In a more specific explanation, the controller employs the concept of proportional-integral-derivative (PID). It is a feedback control loop mechanism [9–11] that continuously tunes the signal to be output to the gate valve while processing the negative feedback input from the thermocouple until the ideal zero error is obtained. We define u as the control signal fed into the gate valve for regulation, where u generated by the PID concept is computed in the time domain using the following expression [12]:

$$\begin{cases} u(t) = K_p \varepsilon(t) + K_i \int \varepsilon(t)\, dt + K_d \frac{d\varepsilon(t)}{dt} \\ \varepsilon(t) = T_{set} - T_{op}(t) \end{cases} \quad (8)$$

where T_{set} denotes the desired temperature (set point) of the fluid in the pipe and T_{op} denotes the operating temperature (feedback input from thermocouple) of the fluid in the pipe, where T_{set} is constantly set to 100°C throughout the simulation. ε represents the tracking error, which defines the error difference between the set point (desired temperature, T_{set}) and the feedback input (temperature measured by the thermocouple, T_{op}). $K = [K_p, K_i, K_d]$ is a set of control variables of the optimization algorithm, where K_p is the proportional gain, K_i is the integral gain, and K_d is the derivative gain. Each coefficient gain corresponds to each variable: K_p times the magnitude of ε, K_i times the integral of ε, and K_d times the derivative of ε. After numerous preliminary tests, it was determined that $-300 \leq K_p \leq 300$ and $-50 \leq K_i, K_d \leq 50$ are most appropriate for the system's characteristics. Indeed, all gains lie within theoretically convincing ranges, with positive upper and negative lower bounds to accommodate controls of the opposite sign. K_p as a proportional gain is set to the widest possible range to allow massive amplification of the signal amplitude, thereby achieving faster processing response. In contrast, shorter ranges work well for K_d and K_i, as they require fine-tuning to control linear distortion and distal term error respectively, but

https://doi.org/10.31436/iiumej.v25i2.3044

excessively narrow ranges are prohibited, as K_d and K_i must provide all conceivable combinations. Figure 5 visually explains the mechanism in the controller when the optimization algorithm is applied. The application of the optimization algorithm to the tuning process is expected to yield better performance in a step response, primarily by increasing the speed at which the set point threshold is reached and reducing the percentage of overshoot (or undershoot) relative to the set point. Hence, the objective fitness of an optimization algorithm is given as follows:

$$\min f(K) = (1 + |\varepsilon(\dot{t}_{end})|)(R_t + S_t + O_s + U_s) \quad (9)$$

where $\varepsilon(\dot{t}_{end})$ also represents the steady-state error of the step response, R_t represents the rise time, S_t represents the settling time, O_s represents the percentage overshoot, and U_s represents the percentage undershoot. By definition, rise time is the time required for the T_{op} response to rise from 10% to 90% of the steady-state response (T_{set}), and settling time is the time required for the T_{op} response to reach and remain within 2% of the steady-state response (T_{set}). The system's damping ratio indicates whether a system is likely to overshoot or undershoot. A system that is over-damped will undershoot its target value. Conversely, an under-damped system will overshoot its target value. Figure 6 illustrates these quantities in a typical step response plot. The minimal fitness that the algorithm can achieve will determine its optimal performance in this optimization application. Overall, the smaller the fitness value an algorithm can achieve, the better the algorithm performs in this application.

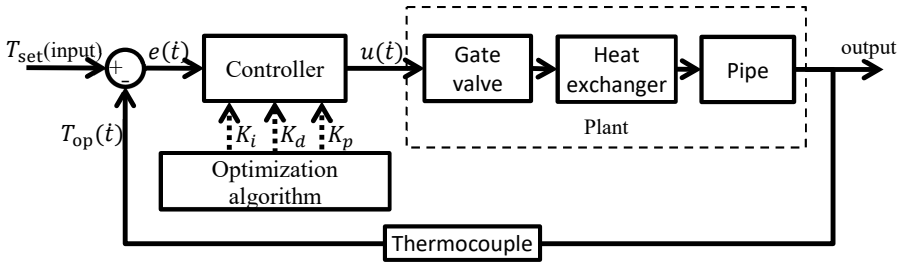


Figure 5. Layout of thermoregulation control system

During the evaluation process, the proposed MeAgtrO and other comparative algorithms (including Alligator Optimization (AgtrO) algorithm [13], Reptile Search Algorithm (RSA) [14], Pelican Optimization Algorithm (POA) [15], Aquila Optimizer (AO) [16], Arithmetic Optimization Algorithm (AOA) [17], Dingoes Optimization Algorithm (DOA) [18], Social Ski Driver (SSD) algorithm [19], and Seagull Optimization Algorithm (SOA) [20]) were installed in the controller in turn to examine and compare the performance of each algorithm in this application. The algorithms chosen for comparison with MeAgtrO were based on the following considerations: AgtrO was chosen as it is the basic algorithm of MeAgtrO. In addition, only state-of-the-art algorithms were considered, among which only algorithms proposed within the last 5 years were selected as comparative algorithms. In fact, AgtrO was proposed in 2023, RSA and POA were proposed in 2022, AO, AOA and DOA were proposed in 2021, SSD was proposed in 2020, and SOA was proposed in 2019. For fair comparison, all comparative algorithms were only selected from the same school of population-based optimization and a category of nature-inspired optimization. The fact that they are standard optimizers was also a crucial factor in their selection as candidates, which makes it easier to assess how better improved MeAgtrO variations affect the application. The parameter settings of all algorithms are set to default, as stated in Table 2, to ensure an equitable comparison in performance. 10 simulation runs were conducted for the application of each

algorithm to facilitate robust statistical analysis and result comparison. Given the problem definition, $t_{max} = 50$ and $N_p = 50$ per simulation run.

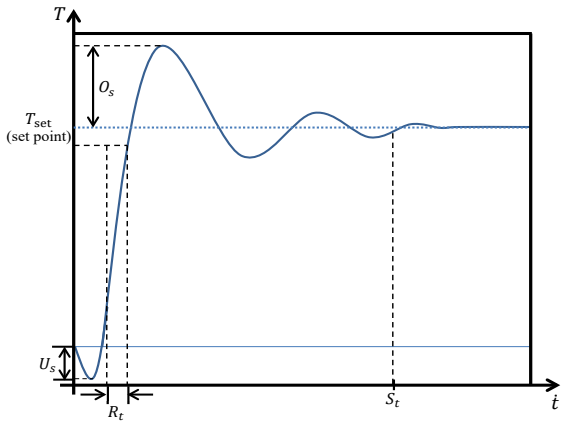


Figure 6. Typical step response

Table 2. Detailed parameters in the pipeline

Parameter	Range or value
MeAgtrO	
—	—
AgtrO	
α	$[0, 1]$
β	$[0, 1]$
δ	$[0, 20]$
γ	$[-50, 0]$
RSA	
α	0.1
β	0.005
POA	
—	—
AO	
α	0.1
δ	0.1
u	0.0265
r_0	10
ω	0.005
φ_0	$3\pi/2$
AOA	
MOP_{max}	1
MOP_{min}	0.2
α	5
μ	0.499
DOA	
P	0.5
Q	0.7
N_{init}	2
N_{end}	N_p/N_{init}
N	$\lfloor N_{init} + (N_{end} - N_{init})r \rfloor$
SSD	
a	$[0, 2]$
SOA	
F_c	$[0, 2]$
b	1

https://doi.org/10.31436/iiumej.v25i2.3044

4. RESULTS AND DISCUSSION

In this section, we compare and analyse the optimal control performance of MeAgtrO (for thermoregulation applications in piping system) against other comparative algorithms based on different collected data.

Table 3 collects the statistical f(K) results obtained by all the evaluated algorithms in the thermoregulation application. The statistics include three evaluation matrices such as the best, mean, and standard deviation (SD) results for every tested algorithm. Knowing that the objective f(K) fitness (refer Equation (9)) is specified for minimization, the superiority on the corresponding evaluation matrices will reveal the respective characteristics of the algorithms in minimizing tracking error, response time, and equipment burden. More specifically, the accuracy, reliability and robustness of the algorithm for solving this optimal control application can be determined by the statistical best, mean and SD results, respectively.

By statistically comparing the best results, it can be noticed that MeAgtrO, AgtrO, POA, POA and SOA achieved the minimum optimality. AO ranked second with excellent statistical best results but slightly worse than these 5 algorithms, followed by RSA in third, AOA in fourth, and finally SSD in fifth. Here, we can analyse that MeAgtrO, AgtrO, POA, DOA and SOA are equally outstanding in terms of accuracy, and definitely outperform the remaining 4 algorithms. By converting the statistical best results into measurable indications, we can clarify that, while the accuracy of RSA, AOA and SSD is insignificant, the accuracy of MeAgtrO, AgtrO, POA, DOA, and SOA is all 100%, which is 43.33% greater than the accuracy of AO.

MeAgtrO ranked first on statistical mean, followed by AgtrO that ranked second, then AO, SOA, DOA, SSD, POA, AOA and RSA that ranked third to ninth respectively. Upon analysis, MeAgtrO, which obtained the most superior statistical mean results, is proven to be the most precise algorithm, while RSA is proven to be the least precise. It is also worth noting that the mean result for MeAgtrO is exactly equal to its best result, indicating that MeAgtrO could reach global minimum in every simulation run. This demonstrates MeAgtrO's consistency in tracking set points (i.e., required temperatures) in optimal control applications for piping system, which saves training costs and time in mechanical processes. From the measurement point of view, the reliability performance of MeAgtrO reaches about 100%, which is 1.11% higher than AgtrO, 2.29% higher than AO, 4.17% higher than SOA, 9.34% higher than DOA, 12.83% higher than POA, 29.96% higher than SSD, etc.

With great consistency comes a great standard deviation (SD) result. Note that the optimal value of SD is zero, where an SD result of zero indicates perfect robustness of the algorithm during the optimization processes. Among all comparative algorithms, only MeAgtrO could obtain statistical SD result that is extremely close to zero value. Without a doubt, MeAgtrO achieved the outperforming SD results in optimal thermoregulatory control applications. Even the second-ranked AgtrO could only achieve an SD result that is at least 106 times behind MeAgtrO; not to mention others that are much worse than AgtrO on SD results. To demonstrate the actual performance difference in terms of measurable robustness, we declare that MeAgtrO is nearly 100% robust, 1.23% better than AgtrO, 1.59% better than AO, 13.01% better than SOA, 15.58% better than DOA, 17.46% better than POA, and 18.28% better than SSD. On the contrary, AOA and RSA have approximately 0% robustness. From mere statistics, MeAgtrO has proven its outperformance in terms of accuracy, reliability, and robustness when dealing with optimal thermoregulatory control applications. All these claims support the fact that MeAgtrO can minimize tracking errors, response times, and equipment burden with minimal training runs.

https://doi.org/10.31436/iiumej.v25i2.3044

Table 3. Complete results of adopted algorithms for thermoregulation application in simulated piping system

Algorithms	Ind.	1st run	2nd run	3rd run	4th run	5th run	6th run	7th run	8th run	9th run	10th run	Best	Mean	SD
MeAgtrO	K_d	-0.43	-0.14	-0.33	0.65	-0.28	0.24	-0.15	-0.32	0.59	0.79			
	K_p	154.41	154.91	154.80	154.50	154.34	154.67	154.55	154.71	154.13	154.32			
	K_i	0.39	-0.12	0.00	0.30	0.46	0.12	0.24	0.09	0.66	0.48			
	$f(K)$	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	**1.39E+00**	**1.39E+00**	**2.72E-07**
AgtrO	K_d	11.40	-3.60	-0.84	1.02	12.69	1.13	1.76	13.11	2.09	-11.55			
	K_p	153.91	139.33	150.95	154.55	153.92	153.89	153.87	153.93	153.83	117.14			
	K_i	0.89	0.60	0.75	0.25	0.88	0.91	0.92	0.86	1.11	0.21			
	$f(K)$	3.37E+00	2.34E+00	1.50E+00	1.39E+00	3.37E+00	1.39E+00	1.39E+00	3.37E+00	1.58E+00	3.99E+00	1.39E+00	2.37E+00	1.05E+00
RSA	K_d	0.00	0.79	0.00	0.00	0.00	0.00	0.00	0.00	-0.05	0.73			
	K_p	0.00	151.19	0.00	0.00	0.00	0.00	0.00	0.00	149.99	145.70			
	K_i	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00			
	$f(K)$	6.55E+04	2.01E+00	6.55E+04	6.55E+04	6.55E+04	6.55E+04	6.55E+04	6.55E+04	2.33E+00	3.02E+00	2.01E+00	4.59E+04	3.17E+04
POA	K_d	24.76	0.61	-0.44	-1.10	-12.68	4.97	-11.37	-13.44	0.67	31.35			
	K_p	144.38	153.52	151.83	112.02	184.74	154.73	93.19	130.00	153.93	140.97			
	K_i	27.67	0.09	0.81	1.30	1.09	-0.13	0.82	-0.02	0.60	27.84			
	$f(K)$	3.84E+01	1.40E+00	1.40E+00	6.63E+00	1.87E+01	4.26E+00	6.15E+00	8.34E+00	1.39E+00	4.01E+01	1.39E+00	1.27E+01	1.49E+01
AO	K_d	-0.30	-19.27	15.46	-2.08	8.35	-4.79	-18.06	14.38	-14.28	0.59			
	K_p	154.22	88.32	149.26	155.72	153.63	140.39	99.70	149.64	105.37	153.98			
	K_i	1.33	0.55	1.31	-0.19	0.81	0.06	0.13	1.08	0.12	0.88			
	$f(K)$	2.03E+00	5.74E+00	3.92E+00	2.08E+00	3.36E+00	2.79E+00	4.90E+00	3.66E+00	4.19E+00	1.42E+00	1.42E+00	3.41E+00	1.36E+00
AOA	K_d	0.00	3.86	0.00	-0.95	-28.96	-1.07	0.00	-0.03	0.00	15.67			
	K_p	0.00	130.47	0.00	150.40	352.48	334.48	0.00	167.79	0.00	150.15			
	K_i	0.00	32.62	0.00	0.00	-1.94	0.00	0.00	-0.30	0.00	21.07			
	$f(K)$	6.55E+04	4.45E+01	6.55E+04	2.12E+00	1.04E+02	1.00E+02	6.55E+04	8.55E+00	6.55E+04	3.65E+01	2.12E+00	2.62E+04	3.38E+04
DOA	K_d	-0.83	-0.23	-1.00	21.98	25.71	-4.56	-0.07	-38.30	-1.14	27.84			
	K_p	153.68	152.13	154.95	198.19	147.40	128.63	148.16	47.64	154.21	148.49			
	K_i	0.93	0.07	-0.16	-0.03	19.96	1.00	0.52	0.32	-0.06	0.37			
	$f(K)$	1.99E+00	1.41E+00	1.39E+00	3.04E+01	3.80E+01	3.37E+00	2.30E+00	8.81E+00	1.39E+00	7.11E+00	1.39E+00	9.62E+00	1.33E+01
SSD	K_d	18.80	-9.73	-4.70	15.33	11.42	-21.10	1.80	-7.71	4.62	9.88			
	K_p	124.92	91.70	67.85	134.50	151.41	73.59	120.72	122.43	123.06	154.22			
	K_i	20.79	10.61	10.19	20.31	17.90	9.60	12.67	12.38	0.89	18.69			
	$f(K)$	3.87E+01	1.31E+01	4.48E+01	3.34E+01	2.07E+01	4.12E+01	3.20E+00	4.02E+01	6.15E+00	3.67E+01	3.20E+00	2.78E+01	1.56E+01
SOA	K_d	0.14	12.81	-1.57	0.12	0.66	-0.93	-0.28	-1.37	0.75	-0.19			
	K_p	154.28	155.94	146.95	154.14	154.62	151.98	153.83	152.06	153.78	153.76			
	K_i	0.52	17.85	0.68	0.57	0.15	0.40	0.52	0.74	0.49	0.67			
	$f(K)$	1.39E+00	3.67E+01	2.20E+00	1.39E+00	1.39E+00	1.40E+00	1.39E+00	2.03E+00	1.39E+00	1.39E+00	1.39E+00	5.07E+00	1.11E+01

The convergence plot is displayed in Figure 7. It plots the means of the objective fitness calculated in Equation (9) for all algorithms in every iteration. Due to the high simulation load, the convergence data was collected from an average of 10 simulation runs, but this amount of simulation runs is believed to produce reliable results. Theoretically, algorithms that converge and stabilize in earlier iterations in the curve are considered to have better convergence rate. Hence, we exclude RSA and AOA that did not show any observable convergence, as their deficiencies are obvious in convergence evaluation. POA barely converged at the beginning of the iteration until the 19–20th iteration, indicating that the convergence rate is slow to begin with. While AO, ROA, DOA and SOA started to converge before the 5th iteration, they only reached stability after the 40th iteration, so it is difficult to claim that they have good convergence speed. AgtrO fared slightly worse as it reached stability in the last 10 iterations but began to converge around the 9th iteration. In contrast, the MeAgtrO curve began to converge from the beginning of iterations, gradually reached the minimum fitness, and stabilized at 20th iterations. Through analysis and comparison, it can be clearly seen that MeAgtrO has a stronger and faster convergence speed. Also, it is obvious that MeAgtrO could attain more minimal global best objective fitness. These findings support the claim that MeAgtrO has the best processing efficiency during optimization for thermoregulation applications.

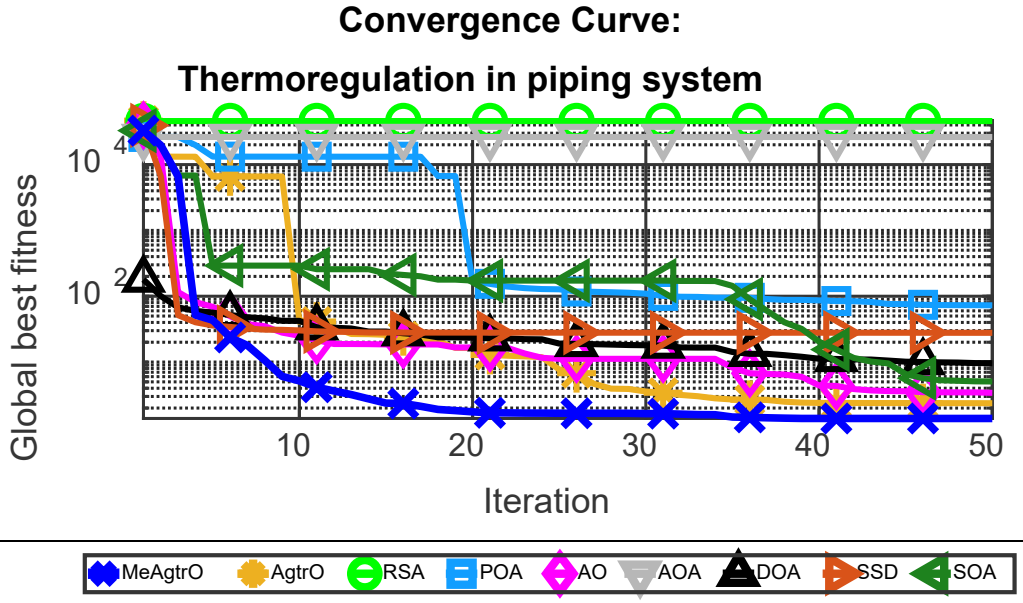


Figure 7. Convergence curves of all algorithms applied to thermoregulation in simulated piping system

It is essential to guarantee that the algorithm has generally adequate clustering properties as a promising swarm-based intelligence. The superiority of the clustering properties convinces the algorithm to a certain extent that it has better local absconding ability, global proximity, and accelerated progress. The clustering properties of the algorithm were evaluated by comparing the position point of an agent with the average position point of the population across the dimension. The degree of alignment of the two plotted linear dimension data determines the superiority of the algorithm in terms of clustering capability. In short, the more aligned the two lines in the cluster plot are, the better the clustering properties of the algorithm, and good clustering indicates the greater the potential of the algorithm to produce higher global optimality in any practical application. For our research application, Figure 8 displays distinguishable clustering plots. Obviously, AgtrO, RSA, POA, AO, AOA and DOA are not even qualified to claim to have good clustering properties. Compared with them,

MeAgtrO shows good clustering properties but is still slightly less than perfect. SSD and SOA, on the other hand, provide nearly flawless map alignment of an agent position vector on the mean position vector, revealing that they have superior clustering qualities to MeAgtrO. However, from graphical analysis alone, MeAgtrO is still competitive with SSD and SOA. The satisfactory clustering performance of MeAgtrO can support the claim that MeAgtrO has the strong ability to avoid local optima, which is beneficial for every optimization solution in any domain. Likewise, the fact that MeAgtrO outperforms AgtrO in terms of clustering properties demonstrates the contribution of the proposed variants.

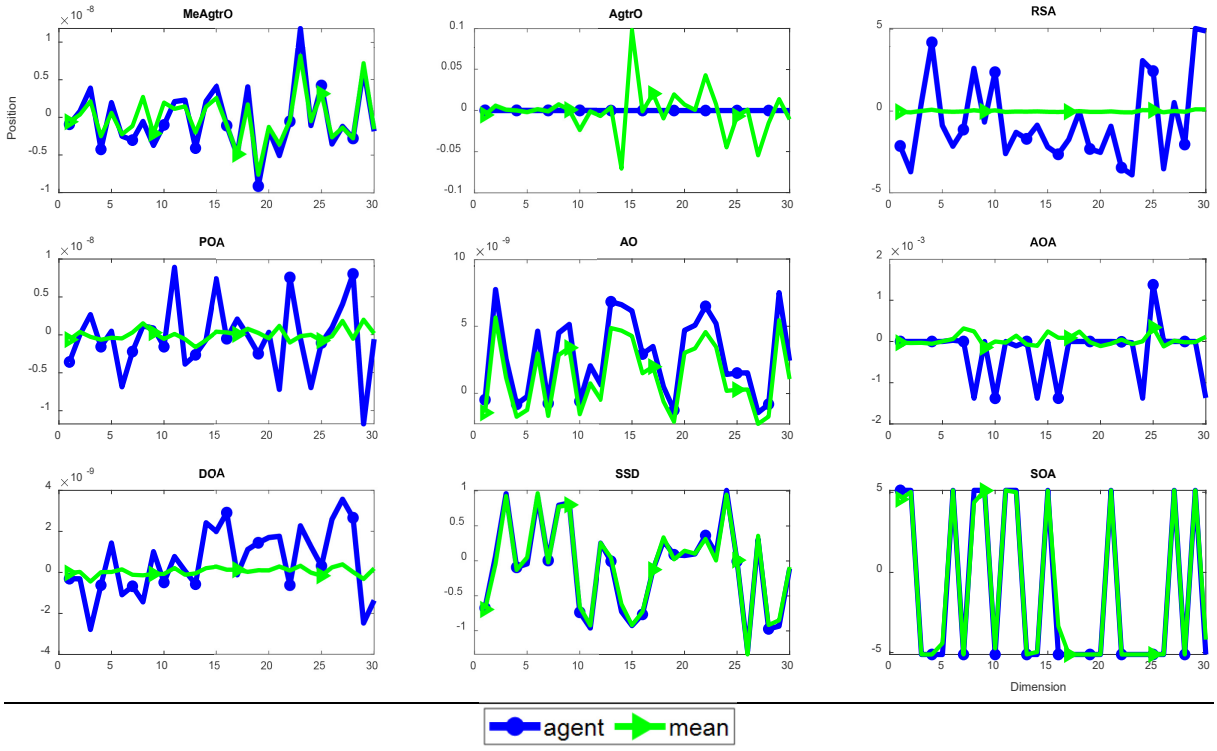


Figure 8. Clustering plots for the selected function in each benchmark set

For deeper validation, step response information obtained by various algorithms applied to thermoregulation were collected and displayed in Figure 9. Note that the set point is 100 oC. Of all the compared algorithms, only MeAgtrO, AgtrO, AO, and DOA could reach the steady state of the set point 100 oC within 10 seconds of simulation time. This shows that only these 4 algorithms show zero steady-state error in the time response. Among them, only MeAgtrO, AgtrO and AO yielded similar rise times of 0.40s, which makes them competitive with each other. However, the settling times contributed by these respective algorithms clearly differentiated their performance: MeAgtrO obtained a settling time of 0.99s, AgtrO obtained a settling time of 1.89s, AO obtained a settling time of 2.96s. Presuming that the overshoot and undershoot from these algorithms do have little affection on the final objective function value, we can hence combine all the evidence and conclude that MeAgtrO yield the best performing step response in this feedback control application. Notably, MeAgtrO provides a solution that can raise the output temperature to the steady set point in the fastest manner with negligible undershoot and overshoot damping, thereby minimizing damage to the heating system. Overall, MeAgtrO has somehow managed to tune the best, most reliable, most efficient, and most robust solutions with superior global best performance and processing efficiency that enable piping systems to achieve the most stable thermoregulation with minimal rise time, settling time, overshoot, and undershoot.

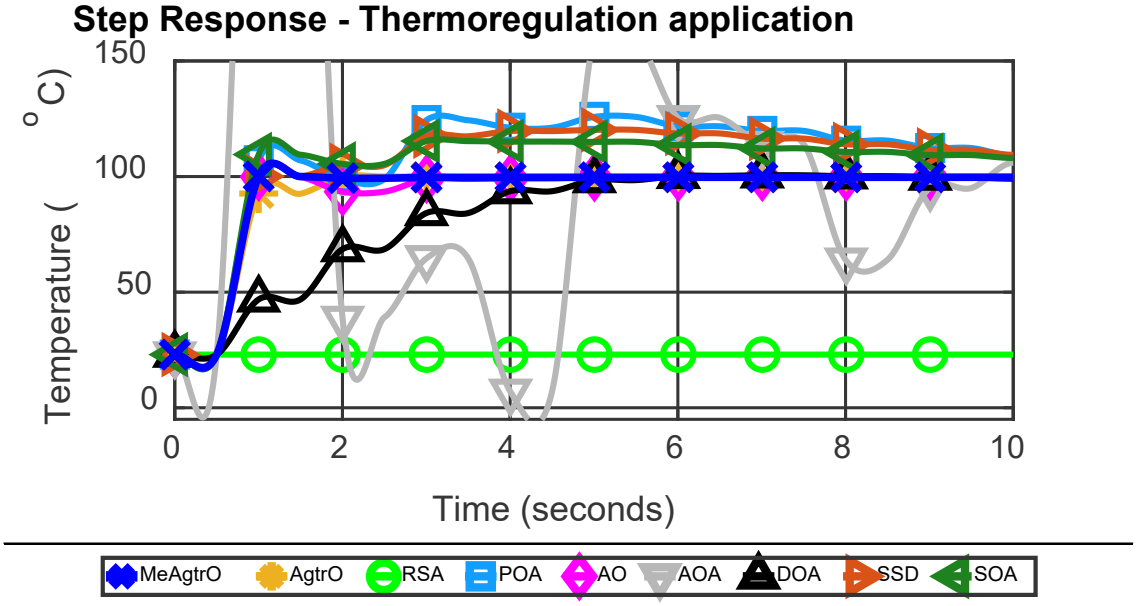


Figure 9. Step response curves of all algorithms applied to thermoregulation in simulated piping system

5. CONCLUSION

In this research work, the proposed MeAgtrO was applied to thermoregulation applications in piping systems, and its performance was compared with eight other comparative algorithms of the same classification. Upon evaluation, MeAgtrO achieved the superior statistical best, mean, and SD results, demonstrating its superior performance in terms of accuracy, reliability, and robustness in its application. MeAgtrO also obtained the best convergence rate and competitive clustering properties, which was supported by statistically plotted convergence curves and clustering maps. In addition, it is demonstrated in the step response diagram that MeAgtrO yielded the least rise time (0.40s), settling time (0.99s), overshoot percentage (0%), and undershoot percentage (0%) with zero steady-state error. All these statements demonstrate that MeAgtrO is successful in every respect as an optimizer. In fact, we can claim that MeAgtrO is highly applicable for industrial control purposes and thus is considered to maintain good performance in a wide range of engineering applications. For any potential upgrades, future emphasis can be on improving processing speed and efficiency according to current application trends.

ACKNOWLEDGEMENT

This research is supported by the Ministry of Higher Education (MOHE) Malaysia under the Fundamental Research Grant Scheme (Grant no. FRGS/1/2021/TK0/USM/02/14).

REFERENCES

[1] C. Guo (2021). "Application of Computer Technology in Optimal Design of Overall Structure of Special Machinery". *Mathematical Problems in Engineering*. vol. 2021, no. Article ID 6619485. pp. 1–9. https://doi.org/10.1155/2021/6619485.

[2] T.-H. Lee et al. (2021). "Optimal Design of a Synchronous Reluctance Motor Using a Genetic Topology Algorithm". *Processes*. vol. 9, no. 10. p. 1778. https://doi.org/10.3390/PR9101778.

[3] Y. Li and X. Deng (2020). "An efficient algorithm for Elastic I-optimal design of generalized linear models". *Canadian Journal of Statistics*. vol. 49, no. 2. pp. 438–470. https://doi.org/10.1002/CJS.11571.

[4] J. O. Agushaka and A. E. Ezugwu (2021). "Advanced arithmetic optimization algorithm for solving mechanical engineering design problems". *PLoS One*. vol. 16, no. 8. p. e0255703. https://doi.org/10.1371/JOURNAL.PONE.0255703.

[5] S. N. Sakharkar, P. Khake, and V. Kolambakar (2018). "Overview of Industrial Piping Structural Design". *International Journal of Engineering Sciences & Research Technology*. vol. 7, no. 2. pp. 45–50. https://doi.org/10.5281/ZENODO.1165605.

[6] P. Smith (2005). "THE PIPING MATERIAL ENGINEER". in *Piping Materials* Guide, Gulf Professional Publishing. pp. 1–10. https://doi.org/10.1016/B978-075067743-1/50010-7.

[7] I. C. Ezema (2019). "Materials". in *Sustainable Construction Technologies: Life-Cycle Assessment*, Butterworth-Heinemann. pp. 237–262. https://doi.org/10.1016/C2016-0-01288-3.

[8] S. Banerjee et al. (2017). "Utilization of Used Packaging Materials for Steam Pipes Insulation". Journal *of Packaging Technology and Research*. vol. 1. pp. 45–51. https://doi.org/10.1007/S41783-017-0005-2.

[9] R. P. Borase et al. (2020). "A review of PID control, tuning methods and applications". *International Journal of Dynamics and Control*. https://doi.org/10.1007/S40435-020-00665-4.

[10] S. Bhagwan, A. Kumar, and J. S. Soni (2016). "A Review on: PID Controller". International *Journal on Recent Technologies in Mechanical and Electrical Engineering (IJRMEE)*. vol. 3, no. 2. pp. 17–22.

[11] K. H. Ang, G. Chong, and Y. Li (2005). "PID control system analysis, design, and technology". *IEEE Transactions on Control Systems Technology*. vol. 13, no. 4. pp. 559–576. https://doi.org/10.1109/TCST.2005.847331.

[12] A. O'dwyer (2006). *Handbook Of Pi And Pid Controller Tuning Rules (2nd Edition)*. World Scientific.

[13] W.-H. Tan and J. Mohamad-Saleh (2023). "Alligator optimisation algorithm for solving unconstrainted optimisation problems". *International Journal of Bio-Inspired Computation*. vol. 21, no. 1. pp. 11–25. https://doi.org/10.1504/IJBIC.2023.130025.

[14] L. Abualigah et al. (2022). "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer". *Expert Systems with Applications*. vol. 191. p. 116158. https://doi.org/10.1016/J.ESWA.2021.116158.

[15] P. Trojovský and M. Dehghani (2022). "Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications". *Sensors*. vol. 22. p. 855. https://doi.org/10.3390/S22030855.

[16] L. Abualigah et al. (2021). "Aquila Optimizer: A novel meta-heuristic optimization algorithm". *Computers & Industrial Engineering*. vol. 157, no. 11. p. 107250. https://doi.org/10.1016/J.CIE.2021.107250.

[17] L. Abualigah et al. (2021). "The Arithmetic Optimization Algorithm". *Computer Methods in Applied Mechanics and Engineering*. vol. 376. p. 113609. https://doi.org/10.1016/J.CMA.2020.113609.

[18] H. Peraza-Vázquez et al. (2021). "A Bio-Inspired Method for Engineering Design Optimization Inspired by Dingoes Hunting Strategies". *Mathematical Problems in Engineering*. vol. 2021. p. 9107547. https://doi.org/10.1155/2021/9107547.

[19] A. Tharwat and T. Gabel (2020). "Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm". *Neural Computing and Applications*. vol. 32. pp. 6925–6938. https://doi.org/10.1007/S00521-019-04159-Z.

[20] G. Dhiman and V. Kumar (2019). "Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems". *Knowledge-Based Systems*. vol. 165. pp. 169–196. https://doi.org/10.1016/J.KNOSYS.2018.11.024.

https://doi.org/10.31436/iiumej.v25i2.3063

OPERATIONAL COST MODELS FOR AN EARTH STATION SYSTEM USING 2-PARALLEL AND 4- PARALLEL CONFIGURATIONS

WAN MUHAMAD SYAAMIL W ARIS[1], NADIRAH ABDUL RAHIM[1*], MAHAYAUDIN M. MANSOR[2], NUR SHAZANA ABDUL RAHMAN[3]

[1]*Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Gombak, Malaysia*
[2]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia*
[3]*FPT Software Malaysia, Kuala Lumpur, Malaysia*

**Corresponding author: nadirahabdulrahim@iium.edu.my*

(Received: 8 October 2023; Accepted: 15 March 2024; Published online: 15 July 2024)

ABSTRACT: Operational cost is important in any system. In the case of the earth station system, operational cost is very crucial. The operational cost can be divided into three costs: maintenance, failure, and replacement costs. There are many complex subsystems incorporated in the earth station system, for instance a high-power amplifier, modulators, and antennas, to name a few. In this research, only the replacement cost was considered. Moreover, there are many replacement methods that are available. These replacement methods implicitly influence both the replacement and the replacement costs. The aim of this research is to provide a new cost model based on which replacement method yields the lowest cost. Two replacement methods are involved in this research: failure-triggered and age-based. The failure-triggered and the age-based replacement methods were considered because these methods are the most used in previous research work. Furthermore, three types of cost models were also considered, and they were linear, polynomial, and exponential operational cost models. The outcomes show that the failure-triggered and age-based replacement methods of 2-parallel configuration of polynomial and linear operational cost models yielded the lowest RMSE value of 2.5. Therefore, both polynomial and linear operational cost models of the 2-parallel configuration were the most optimal operational cost models.

ABSTRAK: Kos operasi adalah penting dalam mana-mana sistem. Dalam kes sistem stesen bumi, kos operasi adalah sangat penting. Kos operasi boleh dibahagikan kepada tiga kos: kos penyelenggaraan, kegagalan dan penggantian. Terdapat banyak subsistem kompleks yang digabungkan dalam sistem stesen bumi, contohnya penguat kuasa tinggi, modulator, dan antena. Penyelidikan ini hanya membincangkan tentang kos penggantian. Selain itu, terdapat banyak kaedah penggantian sedia ada. Kaedah penggantian ini secara tersirat mempengaruhi kedua-dua kos penggantian dan penggantian. Matlamat penyelidikan ini adalah bagi menyediakan model kos baharu berdasarkan kaedah penggantian yang menghasilkan kos terendah. Dua kaedah penggantian yang terlibat dalam kajian ini: penggantian yang dicetuskan oleh kegagalan dan berdasarkan umur. Kaedah cetusan kegagalan dan penggantian berasaskan umur telah dipertimbangkan kerana kaedah ini paling banyak digunakan dalam penyelidikan terdahulu. Tambahan, tiga jenis model kos turut dipertimbangkan, iaitu model kos operasi linear, polinomial dan eksponen. Dapatan menunjukkan kaedah penggantian cetusan kegagalan dan berdasarkan umur melalui konfigurasi model 2 selari bagi kos operasi polinomial dan linear menghasilkan nilai RMSE terendah iaitu 2.5. Oleh itu, kedua-dua model kos operasi polinomial dan linear bagi konfigurasi model 2 selari adalah kos operasi paling optimum.

https://doi.org/10.31436/iiumej.v25i2.3063

KEYWORDS: *Failure-triggered, Age-based, Operational cost, Earth station system and Polynomial cost model*

1. INTRODUCTION

Operational cost is often overlooked by engineers when calculating the overall earth station system life cycle cost (LCC) because other costs, such as planning and development costs, take a higher priority in calculating the overall earth station system LCC [1-2]. Moreover, there are so many operational cost models developed in the past by researchers. However, most of the operational cost models are hard to understand because of their complicated equations [3-9]. An earth station system has many crucial subsystems that have their own maintenance priority [2]. This priority can be used to determine the subsystems that need replacement the most to minimize the earth station system operational downtime. Parallel configurations are included to study the effect of the operational cost. Hence, 2-parallel, and 4-parallel earth station system architectures are used in the design space to see the differences of the cost values on the operational cost [1]. The 2-parallel configuration means there are two redundant units present in each subsystem. Similarly, the 4-parallel configuration means there are four redundant units present in each subsystem. This research is done to support Sustainable Development Goal (SDG) 9, which is to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation [1]. In this research, the mentioned infrastructure is dedicated towards the earth station system. The RF/antenna system always experiences higher failures than other parts of the earth station system as reported in [6,10]. To prevent excessive and unnecessary costs in operating and maintaining the earth station system, this research is fixated on the impact of two different earth station subsystem replacement methods towards the earth station operational cost model analysis to help the party involved. These two replacement methods are failure-triggered and age-based [9,11]. The failure-triggered replacement comes in handy when any subsystem breaks down. Simultaneously, the age-based replacement method can be used when the subsystems are replaced under a set of intervals [9,11]. Moreover, they have their own advantages and disadvantages that can result in either higher or lower expenses in operating the earth station system [1]. The reason these two replacement methods are chosen is because they are the most practiced ones found from in former research work [9,11]. From these two replacement methods, the lowest operational cost is considered for modelling purposes. The contribution of this research is the development of a new optimal cost model that can be implemented to operate and improve earth station systems. This paper is arranged as follows: Section 2 discusses the existing literature review of this research. Section 3 describes the research methodology used to achieve the designated results. Section 4 demonstrates results and analysis. Finally, Section 5 completes this research with future recommendations.

2. LITERATURE REVIEW

Table 1 tabulates the previous work on the cost estimation method, with an eye to each work's contribution and remaining research gap. There are many operational cost models that can be found in the literature review. The problem with those models is that they are very hard to understand and require extensive operational cost data, which is scarce to obtain due to confidentiality [10]. So, the motivational factor in performing this research is to develop a simple and understandable operation cost model based on the simulation results obtained from the 2-parallel and 4-parallel configurations. From Table 1, it can be deduced that the closest research work that matches with this research can be found in Amaitik's [9] and Shazana's [1].

https://doi.org/10.31436/iiumej.v25i2.3063

Hence, these two works become the benchmark for this research. On the other hand, Table 2 shows the benchmarked cost values from the two operational cost models [1,9]. From Table 2, it can be concluded that the RMSE value of 22.82 is obtained through the percentage error calculation of both Amaitik's and Shazana's operational cost models. At the end of this research, the RMSE value can be compared to decide which model is the best one.

Table 1. Previous work on cost estimation method

Reference	Title	Contribution	Research Gap
[11]	Operational cost analysis of an earth station system using parallel configuration	Cost calculations such as replacement and operational are shown	No operational models are shown.
[1]	Sustainable framework for a geostationary satellite control earth station system using parallel configuration	Propose a new operational cost model that can be compared with existing Amaitik's [2] model. Provide earth station reliability model that can be related to the operational cost model.	No other configurations are used for the operational cost model.
[9]	Cost Modelling to support optimum selection of life extension strategy for industrial equipment in smart manufacturing	Provide cost modelling for failure triggered replacement method with 2-parallel earth station configuration. Discussing the cost driver and cost estimation relationships to calculate activity cost based on different maintenance strategies.Give detailed explanation on the cost breakdown structure.Provide detailed cost estimation methodology	The maintenance methods are too general. Various cost estimation methods are not compared
[6]	Analysis of failure frequency and failure rate of RF/Antenna subsystems for an earth station system	Clearly mentioned which subsystems in earth station system that need to be maintained. Provide data on the mean time between failure (MTBF) of the subsystems	The data only provides failure frequency of the subsystems
[2]	Maintenance cost-based importance analysis under different maintenance strategies	Discussing the priority on the subsystem maintenance to maximize the earth station system uptime. Suggest different replacement methods with detailed calculation to compare the effectiveness of said method	The calculation complexity level is high
[12]	Cost and performance modelling for earth system data management and beyond	Provide a research methodology and considerations relevant for modeling data centres.Demonstrate how to construct a coarse-grained model	Cost did not integrate fully with I/O middleware

https://doi.org/10.31436/iiumej.v25i2.3063

Table 2. Benchmark of two operational cost models [1]

Earth Station Service Year	Operational Cost Model (e10)		Percentage error (%)
	Amaitik	Syazana	Syazana/Amaitik
1	0.00	0.97	-247.00
2	7.23	3.71	48.67
3	9.07	8.39	7.47
4	17.16	13.07	23.81
5	16.57	17.75	-7.10
6	21.65	22.43	-3.58
7	25.57	27.11	-6.03
8	34.36	31.79	7.50
9	39.62	36.47	7.96
10	41.95	41.15	1.92
		RMSE Value	22.82

2.1 Failure Cost

Earth station systems usually have failures with their antennas, which are used to send and receive data signals [13]. Earth station antennas are located on the ground end of satellite links. Higher gain is needed to receive weak signals from the satellite or to transmit strong signals to the satellite. The antenna can be damaged if a lightning surge occurs during thunderstorm weather when the power cord, antenna cable, and Local Area Network (LAN) cable are still connected to the main system. The damage is not limited to the antenna; most of the main devices in the monitoring room may also be affected. These damages can disrupt the operation activity of the satellite system. To prevent this scenario, operators are advised to ensure that there is no possible connection between the devices and the system established after the operation has ended [13].

2.2 Replacement Cost

As mentioned before, there are two types of replacement methods where each result in different replacement costs for the subsystems. The first one is a failure-triggered replacement method, used if any of the subsystems break. The next one is an age-based replacement method, where there is a time interval between each replacement of the subsystems [2]. There are a lot of subsystems that need to be maintained when planning for the earth station maintenance [6]. These subsystems have their replacement cost varying based on their structure complexity and difficulty to repair.

For the first type of replacement method (failure-triggered) only the subsystems that are broken are issued a replacement. From [6], the failure frequency for the 2-parallel and 4-parallel configurations are tabulated in Table 2 and Table 3, respectively. These two tables consist of subsystems, namely HPA, Antenna, LNA, Up Converter, Down Converter, Modulator, and Demodulator, that had their downtime at least once per year. These broken subsystems were replaced as soon as they ceased to function. This automatically increases the efficiency of the earth station operation.

For the second type of replacement method, which is age-based, the management will set up the interval between each replacement of the subsystems of the earth station. This time interval can be of any length, but it will be pointless if the time is too long (for example, 1 year) because, from [6], some of the subsystems were down during the first year of its lifecycle, thus disrupting the whole earth station operation. This method however can produce a different

https://doi.org/10.31436/iiumej.v25i2.3063

result if the time interval is set up for a shorter time, but it may introduce a hike in replacement cost.

2.3 Operational Cost Models

There are three types of operational cost models configured for each replacement method for the 2-parallel and the 4-parallel earth station system configurations, respectively. Those operational cost models are linear, exponential, and polynomial [14]. The linear operational cost model is characterized by a straight-line graph and showcases a constant rate of change in operational cost as replacements are made. It offers a simplistic yet valuable perspective, allowing decision-makers to evaluate the linear relationship between replacement frequency and corresponding cost implications [14]. As mentioned in the previous section, the existing cost models available are hard to comprehend due to their complex equations [15-21].

On the other hand, the exponential operational cost model presents a more dynamic scenario, where the graph curves upwards in an accelerating fashion. This signifies that operational costs escalate exponentially as replacements occur [14]. The exponential operational cost model captures situations where the relationship between replacement frequency and operational costs is highly nonlinear, indicating that even slight increases in replacements can lead to significant cost escalations [14].

Lastly, the polynomial operational cost model, as depicted by a curved graph with varying degrees of complexity, accounts for even more intricate relationships between replacement frequency and operational costs. It enables decision-makers to analyse scenarios where the cost implications are influenced by multiple factors, potentially leading to non-uniform growth patterns [14]. By incorporating polynomial equations of different degrees, this model caters to a wide range of situations, accommodating both moderate and highly complex cost relationships [14]. These operational cost models are compared directly to their measured operational cost, which yields varying Root Mean Square Error (RMSE) value that can show which model is the most optimal. By having a lower RMSE value, a model becomes more optimal as it has higher accuracy than the one with higher RMSE [14]. The RMSE value can be derived from Eq. (1).

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}(\mathrm{x}_1^2 + \mathrm{x}_2^2 + \cdots + \mathrm{x}_\mathrm{n}^2)} \tag{1}$$

where x is the cost difference between measured cost and model cost for n, which is the number of years.

3. RESEARCH METHODOLOGY

There are several steps that need to be included to achieve the designated results. There are two parts in the methodology: the first part is the overall methodology of this research, and the second part is the replacement method selection [11]. As for the overall methodology, the explanation is in the following flowchart, as shown in Fig. 1. Fig. 1 illustrates the research methodology flowchart.

https://doi.org/10.31436/iiumej.v25i2.3063

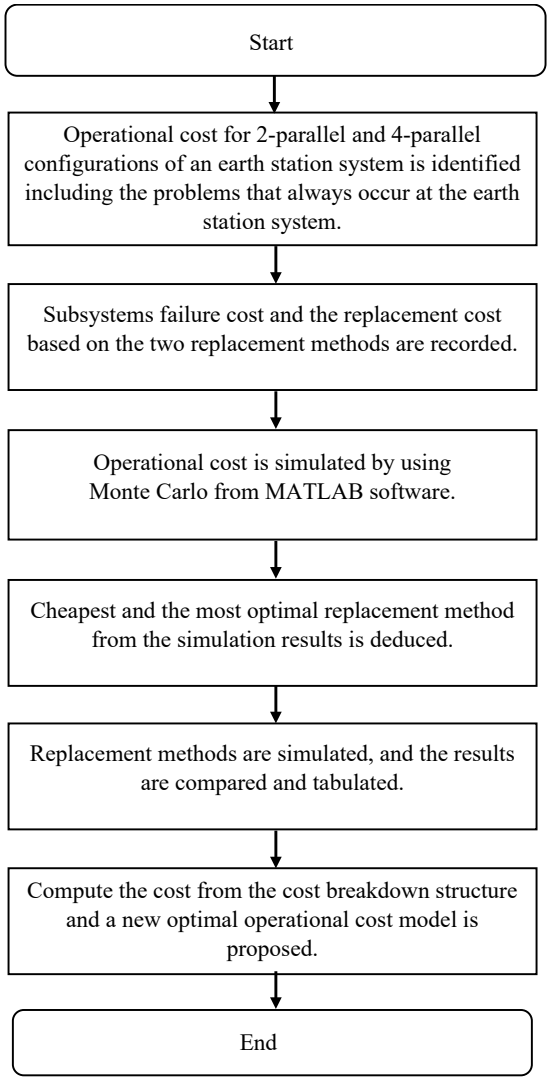


Figure 1. Research methodology flowchart.

Firstly, the operational cost for 2-parallel and 4-parallel configurations of an earth station system is identified including the problems that always occur at the earth station system. The 2-parallel and 4-parallel diagrams can be found in Fig. 2 and Fig. 3 respectively.

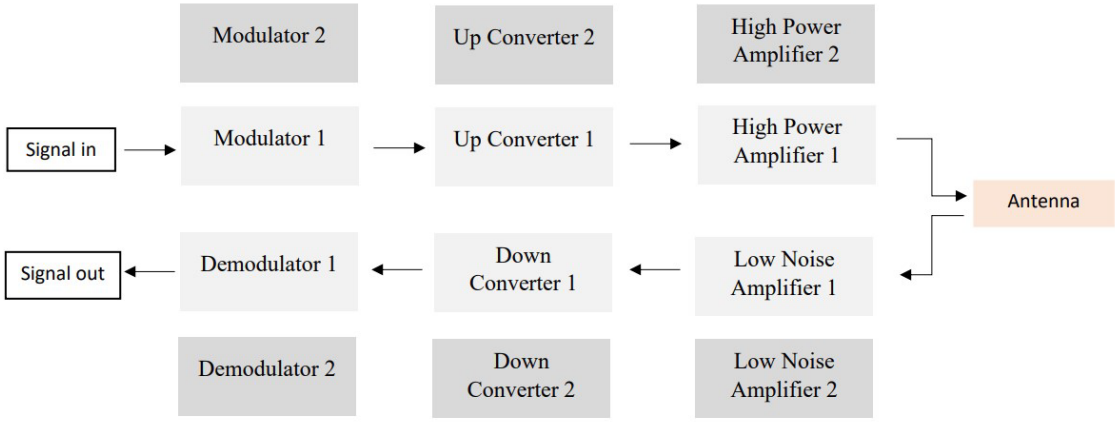


Figure 2. 2-parallel configuration.

https://doi.org/10.31436/iiumej.v25i2.3063

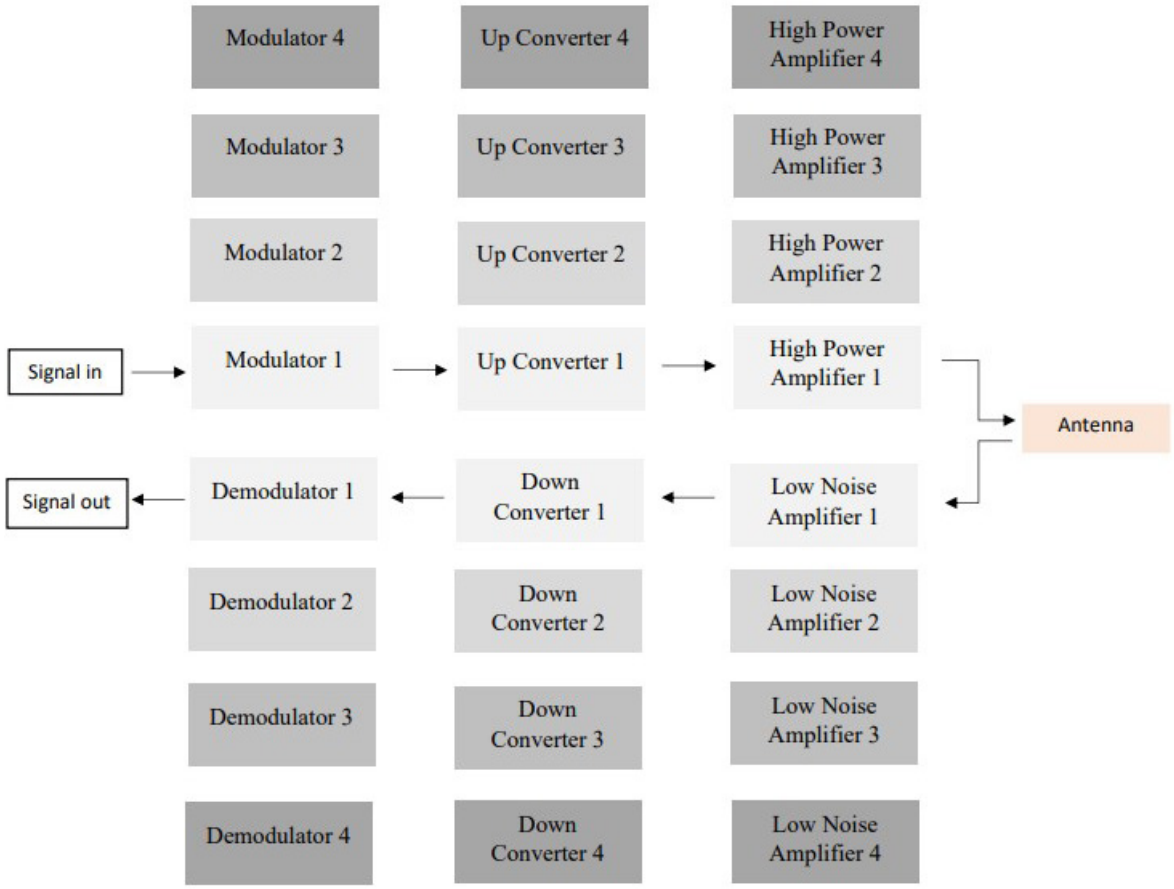


Figure 3. 4-parallel configuration.

In this case, 2-parallel and 4-parallel configurations are chosen because of Malaysia East Asia Satellite (MEASAT) recommendations [6]. MEASAT adopts these two configurations in their earth station system design [6]. Whereas 3-parallel configuration is not recommended by MEASAT because the configuration is less desirable in their earth station system design [6]. Then, the subsystems failure and the replacement costs based on the two replacement methods are recorded. Next, the operational cost is simulated by using Monte Carlo from MATLAB software [11]. After that, the cheapest and the most optimal replacement method from the simulation results is deduced. For the replacement method selection, there are two types of replacement methods: failure-triggered and age-based. Both replacement methods were simulated, and the results were compared and tabulated [11]. The last step is to compute the cost from the cost breakdown structure. The cost that is considered in this research consists of only replacement cost and based on this cost; a new optimal operational cost model is proposed.

The simulation duration was set to run for 10 years to collect sufficient operational cost data from those years. The operational cost data comprises operation and maintenance costs which were supplied by MEASAT, but the exact amounts were camouflaged. There are six parts of simulation: failure-triggered replacement linear operational cost model of the 2 and 4-parallel configurations, failure-triggered replacement exponential operational cost model of the 2 and 4-parallel configurations, failure-triggered replacement polynomial operational cost model of the 2 and 4-parallel configurations, age-based replacement linear operational cost model of the 2 and 4-parallel configurations, age-based replacement exponential operational cost model of the 2 and 4-parallel configurations, and age-based replacement polynomial operational cost model of the 2 and 4-parallel configurations. The data was then analyzed to

https://doi.org/10.31436/iiumej.v25i2.3063

find the most optimal replacement method using a certain configuration. The results are elaborated further in the next subsection.

4. RESULTS AND ANALYSIS

In this section, two types of replacement methods are shown. They are the failure-triggered and age-based replacement methods. For each replacement method, three types of operational cost models were developed: linear, exponential, and polynomial operational cost models. For each type of operational cost model, the unit is in US dollars.

4.1. Failure-triggered Replacement Linear Operational Cost Model

The first part of the cost modeling is for the failure-triggered replacement linear cost model. Fig. 4(a) shows the linear cost graph for the failure-triggered replacement of a 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The y-axis refers to the operational cost whereas the x-axis refers to the life services of the earth station system. The original measured cost can be viewed in [10]. The newly derived linear operational cost model is shown in Eq. (2).

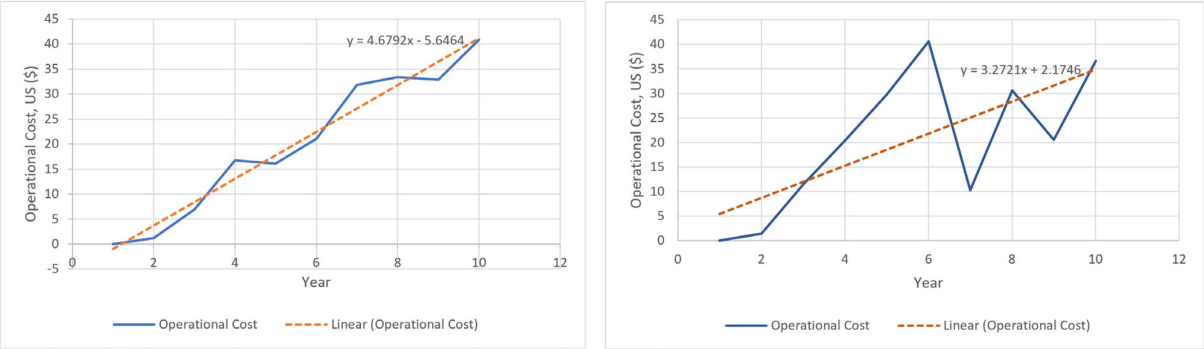
$$y = 4.6792x - 5.6464 \tag{2}$$

where x refers to the number of years and y is the linear operational cost model.

Fig. 4(b) shows the linear cost graph for the failure-triggered replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived linear operational cost model is shown in Eq. (3).

$$y = 3.2721x + 2.1746 \tag{3}$$

where x refers to the number of years and y is the linear operational cost model.



(a) (b)

Figure 4. (a) Linear model for 2-parallel configuration, (b) Linear model for 4-parallel configuration.

4.2. Failure-triggered Replacement Exponential Operational Cost Model

The second part of the cost modeling is for the failure-triggered replacement exponential operational cost model. Fig. 5(a) shows the exponential cost graph for the failure-triggered replacement of a 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived exponential cost model is shown in Eq. (4).

$$y = 0.1419e^{0.7071x} \tag{4}$$

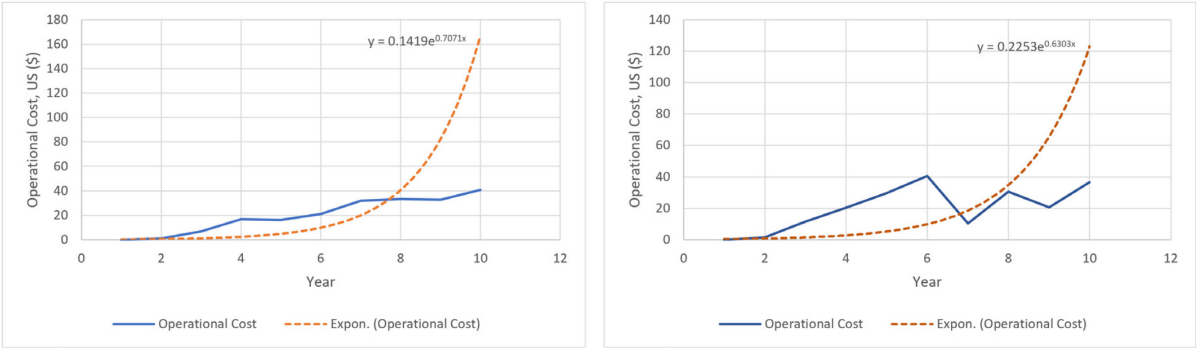
where x refers to the number of years and y is the exponential operational cost model.

https://doi.org/10.31436/iiumej.v25i2.3063

Fig. 5(b) shows the exponential cost graph for the failure-triggered replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived exponential operational cost model is demonstrated in Eq. (5).

$$y = 0.2253e^{0.6303x} \quad (5)$$

where *x* refers to the number of years and *y* is the exponential operational cost model.



(a) (b)

Figure 5. (a) Exponential model for 2-parallel configuration, (b) Exponential model for 4-parallel configuration.

4.3. Failure-triggered Replacement Polynomial Operational Cost Model

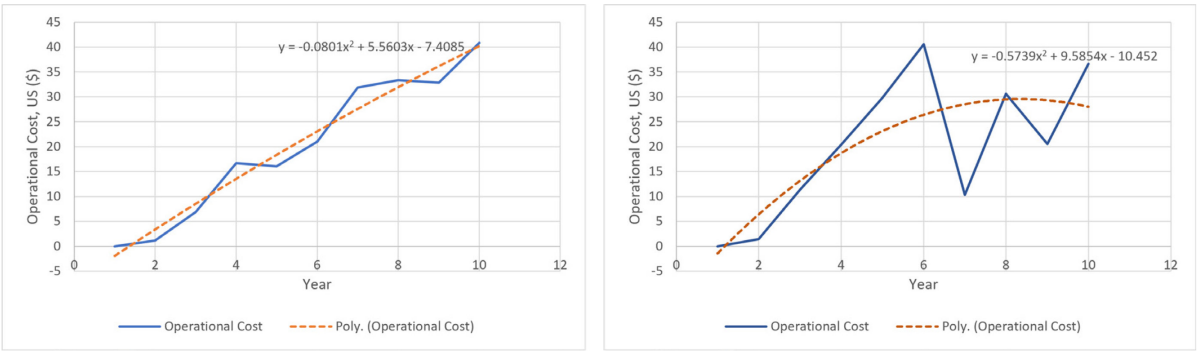
The third part of the cost modeling is for the failure-triggered replacement polynomial operational cost model. Fig. 6(a) shows the polynomial cost graph for the failure-triggered replacement of 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived polynomial operational cost model is depicted in Eq. (6) where *x* refers to the number of years and *y* is the polynomial operational cost model.

$$y = -0.0801x^2 + 5.5603x - 7.4085 \quad (6)$$

Fig. 6(b) shows the polynomial cost graph for the failure-triggered replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived polynomial operational cost model is depicted in Eq. (7).

$$y = -0.5739x^2 + 9.5854x - 10.452 \quad (7)$$

where *x* refers to the number of years and *y* is the polynomial operational cost model.



(a) (b)

Figure 6. (a) Polynomial model for 2-parallel configuration(b) Polynomial model for 4-parallel configuration

https://doi.org/10.31436/iiumej.v25i2.3063

4.4. Age-Based Replacement Linear Operational Cost Model

The fourth part of the cost modeling is for the age-based replacement linear cost model. Fig. 7(a) shows the linear cost graph for the age-based replacement of the 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived linear operational cost model is demonstrated in Eq. (8).

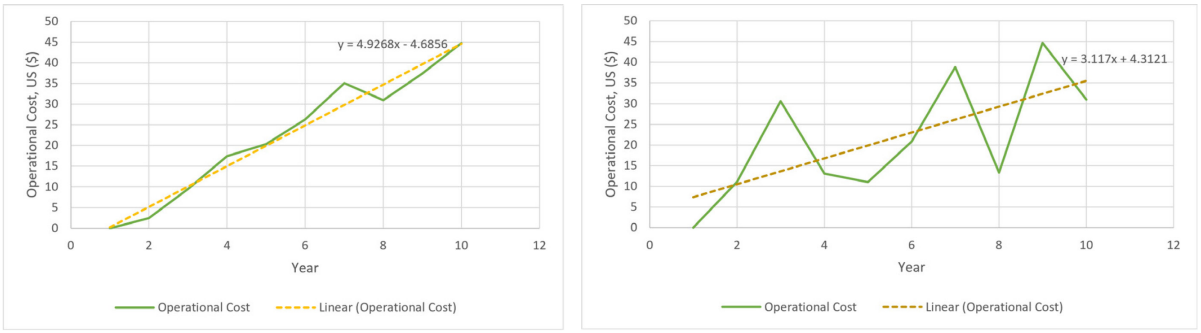
$$y = 4.9268x - 4.6856 \tag{8}$$

where *x* refers to the number of years and *y* is the linear operational cost model.

Fig. 7(b) shows the linear cost graph for the age-based replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived linear operational cost model is shown in Eq. (9).

$$y = 3.117x - 4.3121 \tag{9}$$

where *x* refers to the number of years and *y* is the linear operational cost model.



(a) (b)

Figure 7. (a) Linear model for 2-parallel configuration, (b) Linear model for 4-parallel configuration.

4.5 Age-Based Replacement Exponential Operational Cost Model

The fifth part of the cost modeling is for the age-based replacement exponential cost model. Fig. 8(a) shows the exponential cost graph for the age-based replacement of a 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived exponential operational cost model is shown in Eq. (10).

$$y = 0.2103e^{0.6699x} \tag{10}$$

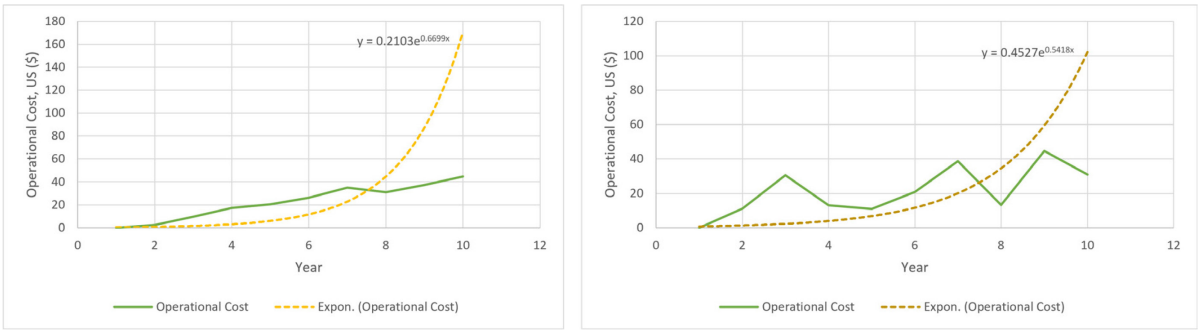
where *x* refers to the number of years and *y* is the exponential operational cost model.

Fig. 8(b) shows the exponential cost graph for the age-based replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived exponential operational cost model is demonstrated in Eq. (11).

$$y = 0.4527e^{0.5418x} \tag{11}$$

where *x* refers to the number of years and *y* is the exponential operational cost model.

https://doi.org/10.31436/iiumej.v25i2.3063



(a) (b)

Figure 8. (a) Exponential model for 2-parallel configuration(b) Exponential model for 4-parallel configuration.

4.6. Age-Based Replacement Polynomial Operational Cost Model

The last part of the cost modeling is for the age-based replacement polynomial cost model. Fig. 9(a) shows the polynomial cost graph for the age-based replacement of the 2-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived polynomial operational cost model is depicted in Eq. (12).

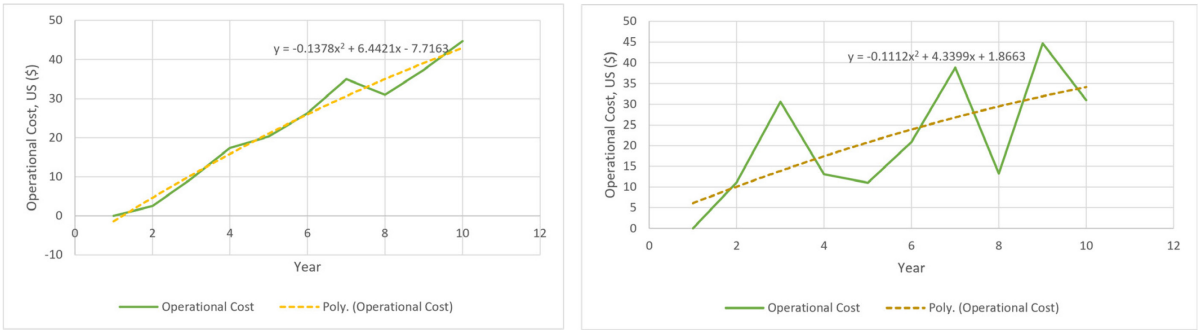
$$y = -0.1378x^2 + 6.4421x - 7.7163 \quad (12)$$

where *x* refers to the number of years and *y* is the polynomial operational cost model.

Fig. 9(b) shows the polynomial cost graph for the age-based replacement of the 4-parallel configuration in comparison to the original measured cost for 10 years of simulation duration. The newly derived polynomial operational cost model is shown in Eq. (13).

$$y = -0.1112x^2 + 4.3399x + 1.8663 \quad (13)$$

where *x* refers to the number of years and *y* is the polynomial operational cost model.



(a) (b)

Figure 9. (a) Polynomial model for 2-parallel configuration(b) Polynomial model for 4-parallel configuration.

4.7. Replacement Methods and Cost Models Comparison

Tables 3 and 4 compare all the operational cost models with their respective parallel configurations and replacement methods. The operational cost model with the lowest RMSE value is the most optimal. For the failure-triggered replacement method of the 2-parallel configuration, the polynomial operational cost model has the lowest RMSE value of 2.5, while the linear and exponential operational cost models have values of 2.6 and 24.3, respectively. Meanwhile, for the failure-triggered replacement method of the 4-parallel configuration, the

https://doi.org/10.31436/iiumej.v25i2.3063

polynomial operational cost model has the lowest RMSE value of 8.7, while the linear and exponential operational cost models have values of 9.7 and 24.3, respectively.

Proceeding with the age-based replacement method, the 2-parallel configuration of the linear operational cost model has the most optimal RMSE value of 2.5 compared to the exponential and polynomial operational cost model values of 26.6 and 2.9, respectively. For the last configuration, the age-based replacement method of the 4-parallel configuration, the most optimal operational cost models are linear and polynomial operational cost models because they have the same value of 10.1 compared to the exponential operational cost model, which has a value of 25.3. From the four types of operating cost models, it can be deduced that the polynomial and linear operational cost models for the failure-triggered and the age-based replacement methods of the 2-parallel configuration are chosen to be the most optimal operational cost model because both have the same RMSE value of 2.5.

Table 3. Cost model comparison between two replacement methods for failure 2-parallel and failure 4-parallel configuration

Earth Station Service Year	**Operational Cost Model (e^{10})**					
	Failure Triggered (2-Parallel)			Failure Triggered (4-Parallel)		
	Linear	Exponential	Polynomial	Linear	Exponential	Polynomial
1	-1.0	0.3	-1.9	5.4	0.4	-1.4
2	3.7	0.1	3.4	8.7	0.2	6.4
3	8.4	1.6×10^{-2}	8.6	12.0	0.1	13.1
4	13.1	4.0×10^{-3}	13.6	15.3	2.3×10^{-2}	18.7
5	17.7	1.0×10^{-3}	18.4	18.5	9.1×10^{-3}	23.1
6	22.4	3.0×10^{-3}	23.1	21.8	3.6×10^{-3}	26.4
7	27.1	1.0×10^{-3}	27.6	25.1	1.4×10^{-3}	28.5
8	31.8	0	31.9	28.4	5.0×10^{-4}	29.5
9	36.5	0	36.1	31.6	2.0×10^{-4}	29.3
10	41.4	0	40.2	34.9	1.0×10^{-4}	28.0
RMSE Value	**2.6**	**24.3**	**2.5**	**9.7**	**24.3**	**8.7**

Table 4. Cost model comparison between two replacement methods for 2-parallel and 4-parallel configuration

Earth Station Service Year	**Operational Cost Model (e^{10})**					
	Age based (2-Parallel)			Age based (4-Parallel)		
	Linear	Exponential	Polynomial	Linear	Exponential	Polynomial
1	0.2	0.4	-1.4	7.4	0.7	6.1
2	5.2	0.1	4.6	10.5	0.4	10.1
3	10.1	4.4×10^{-2}	10.5	13.7	0.3	13.9
4	15.0	1.5×10^{-3}	15.8	16.8	0.2	17.4
5	19.9	5.4×10^{-3}	21.0	19.9	0.1	20.8
6	24.9	1.9×10^{-3}	26.0	23.0	0.1	20.8
7	29.8	7.0×10^{-3}	30.6	26.1	5×10^{-2}	26.8
8	34.7	2.0×10^{-4}	35.0	29.2	3.2×10^{-2}	29.5
9	39.7	1.0×10^{-4}	29.1	32.4	2.1×10^{-2}	31.9
10	44.6	0	42.9	35.5	1.4×10^{-2}	34.1
RMSE Value	**2.5**	**26.6**	**2.9**	**10.1**	**25.3**	**10.1**

5. CONCLUSION AND FUTURE WORK

In summary, the aims of this research were accomplished. The first one was to analyze the differences between failure-triggered replacement and age-based replacement methods using 2-parallel and 4-parallel earth station system configurations, and the second one was to propose

the most optimal cost model based on the two types of replacement methods. From the analysis, it can be concluded that the polynomial operational cost model for the failure-triggered and the linear operational cost model for the age-based replacement methods of the 2-parallel configuration of the earth station system were the most optimal cost models because both yielded the lowest operational cost and have among the lowest RMSE value compared to the other types of cost models of the 4-parallel configuration. The justification for this finding is that MEASAT prefers to adopt a 2-parallel configuration in their earth station system design because it is way cheaper than the 4-parallel configuration. In terms of reliability value, as mentioned in [1], the wear and tear caused failure rates to increase over the years. Hence, when it reaches the end of its life cycle for the 2-parallel configuration, the reliability value of 0.24, as mentioned in [1], is still acceptable.

However, there is still room for improvement in this research. For example, there are still other types of replacement methods and earth station configurations that can be explored and used to run the simulations. One of the other replacement methods, which is a hybrid between failure-triggered and age-based replacement methods, needs to be further explored. This hybrid method may produce a good result in terms of producing the lowest operational cost and a replacement cost curvature, thus giving the most optimal operational cost to be implemented in an earth station system design.

ACKNOWLEDGEMENT

The authors would like to thank MEASAT Satellite Systems Sdn Bhd for providing data access. This research is funded by the Asian Airforce of Aerospace Research and Development (AOARD) under the grant scheme numbers FA2386-23-1-4073 and SPI23-179-0179.

REFERENCES

[1] Rahman, N.S.A. and Nadirah Abdul Rahim (2023). Sustainable framework for a geostationary satellite control earth station system using parallel configuration. Indonesian Journal of Electrical Engineering and Computer Science, 30(3): 1498-1508.

[2] Chen, L., et al. (2022). Maintenance cost-based importance analysis under different maintenance strategies. Reliability Engineering & System Safety, 222: 108-435.

[3] S. K. Sharma, S. Chatzinotas, and B. Ottersten (2006). Integrating packet-level FEC with data carousels for reliable content delivery in satellite broadcast/multicast systems. International Journal of Satellite Communications and Networking, 269:119–127.

[4] J. Bouwmeester, A. Menicucci, and E. K. A. Gill (2022). Improving CubeSat reliability: subsystem redundancy or improved testing? Reliability Engineering and System Safety, 220:1-18.

[5] N. A. Rahim and N. Nordin (2020). Reliability model and proposed maintainability activities of earth station system. International Journal of Electrical, Electronics and Data Communication, 5:20-24.

[6] N. A. Rahim, N. A. Rahman, S. Abdoli, S. Rao, and M. I. Mokhtar (2022). Analysis of failure frequency and failure rate of RF/antenna subsystems for an earth station system. NeuroQuantology, 20(10):8654–8665.

[7] N. S. A. Rahman and N. A. Rahim (2022). Analysis of reliability prediction and maintainability activities of an earth station system using parallel configurations. In WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications, 1647-1655.

[8] N. B. A. Rahim and T. L. J. Ferris (2020). A method to establish a trade-space of system requirements and life cycle cost. IEEE Systems Journal, 14(1):1257–1264.

[9] Amaitik, N., et al. (2022). Cost Modelling to Support Optimum Selection of Life Extension Strategy for Industrial Equipment in Smart Manufacturing. Circular Economy and Sustainability, 2:1425-1444.

[10] MEASAT, “TTC failure equipment record and repair cost 2017-2019 - with description,”

MEASAT, 2021.
[11] Wan Syaamil Muhammad W Aris and Nadirah Abdul Rahim (2023). Operational cost analysis of an earth station system using parallel configuration. 2023 9th International Conference of Computer and Communication Engineering (ICCCE‘ 23), 388-393.
[12] Lüttgau, J. and J. Kunkel (2018). Cost and Performance Modeling for Earth System Data Management and Beyond: ISC High Performance 2018 International Workshops, Frankfurt/Main, Germany, June 28, 2018, Revised Selected Papers, 23-35.
[13] Ali, F., S. Rahim, and M. Jusoh (2021). Amateur satellite ground station: Troubleshooting and lesson learned. Journal of Physics: Conference Series, 1768: p. 012013.
[14] Modeling the IR curve: the Exponential Polynomial Model ("EPM"), the true extension of Nielson-Siegel (2018),1-28.
[15] Khzouz, M., Gkanas, E. I., Shao, J., Sher, F., Beherskyi, D., El-Kharouf, A., & Qubeissi, M. A. (2020). Life cycle costing analysis: Tools and applications for determining hydrogen production cost for fuel cell vehicle technology. Energies, 13(15), 3783.
[16] Van Den Brand, J., Chen, L., Peng, R., Kyng, R., Liu, Y. P., Gutenberg, M. P., & Sidford, A. (2023). A deterministic almost-linear time algorithm for minimum-cost flow. In 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS), 503-514.
[17] Coffrin, C., Knueven, B., Holzer, J., & Vuffray, M. (2021). The impacts of convex piecewise linear cost formulations on AC optimal power flow. Electric Power Systems Research, 199, 107191.
[18] Perera, S. C., & Sethi, S. P. (2023). A survey of stochastic inventory models with fixed costs: Optimality of (s, S) and (s, S)-type policies—Discrete-time case. Production and Operations Management, 32(1), 131-153.
[19] Gao, Y., Liu, Y., & Peng, R. (2023). Fully dynamic electrical flows: Sparse maxflow faster than Goldberg–Rao. SIAM Journal on Computing, (0), FOCS21-85.
[20] Bernstein, A., Gutenberg, M. P., & Saranurak, T. (2022,). Deterministic decremental sssp and approximate min-cost flow in almost-linear time. In 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), 1000-1008.
[21] Cruz-Mejía, O., & Letchford, A. N. (2023). A survey on exact algorithms for the maximum and minimum-cost flow problems. Networks.

https://doi.org/10.31436/iiumej.v25i2.3119

ATTITUDE UAV STABILITY CONTROL USING LINEAR QUADRATIC REGULATOR-NEURAL NETWORK (LQR-NN)

OKTAF AGNI DHEWA[1], FATCHUL ARIFIN[2*], ARDY SETO PRIAMBODO[1], ANGGUN WINURSITO[2], YASIR MOHD. MUSTAFAH[3]

[1]Dept. of Electrical and Electronic Engineering, Vocational Faculty, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
[2]Dept. of Electronics and Informatics Engineering, Faculty of Engineering, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
[3]Dept. of Mechatronics Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Selangor, Malaysia

**Corresponding author: fatchul@uny.ac.id*

(Received: 20 November 2023; Accepted: 11 May 2024; Published online: 15 July 2024)

ABSTRACT: The stability of an Unmanned Aerial Vehicle (UAV) attitude is crucial in aviation to mitigate the risk of accidents and ensure mission success. This study aims to optimize and adaptively control the flight attitude stability of a flying wing-type UAV amidst environmental variations. This is achieved through the utilization of Linear Quadratic Regulator-Neural Network (LQR-NN) control, wherein the Neural Network predicts the optimal K gain value by fine-tuning Q and R parameters to minimize system errors. An online learning neural network adjusts the K value based on real-time error feedback, enhancing system performance. Experimental results demonstrate improved stability metrics: for roll angle stability, a rise time of 0.4682 seconds, settling time of 1.3819 seconds, overshoot of 0.298%, and Steady State Error (SSE) of 0.133 degrees; for pitch angle stability, a rise time of 0.2309 seconds, settling time of 0.7091 seconds, overshoot of 0.1224%, and Steady State Error (SSE) of 0.0239 degrees. The LQR-NN approach effectively reduces overshoot compared to traditional Linear Quadratic Regulator (LQR) control, thereby minimizing oscillations. Furthermore, LQR-NN can minimize the Steady State Error (SSE) to 0.074 degrees for roll rotation motion and 0.035 degrees for pitch rotation motion.

ABSTRAK: Kestabilan perubahan Pesawat Tanpa Pemandu (UAV) adalah penting dalam penerbangan bagi mengurangkan risiko kemalangan dan memastikan kejayaan misi. Kajian ini bertujuan mengoptimum dan menstabilkan perubahan kawalan adaptif penerbangan UAV jenis sayap terbang di tengah-tengah variasi persekitaran. Ini dicapai melalui penggunaan kawalan Rangkaian Linear Kuadratik Pengatur-Neural (LQR-NN), di mana Rangkaian Neural meramal nilai perolehan K optimum dengan meneliti parameter Q dan R bagi mengurangkan ralat sistem. Rangkaian neural pembelajaran dalam talian melaraskan nilai K berdasarkan maklum balas ralat masa nyata, ini meningkatkan prestasi sistem. Dapatan kajian eksperimen menunjukkan metrik kestabilan lebih baik: bagi kestabilan sudut gulungan, masa kenaikan sebanyak 0.4682 saat, masa kestabilan 1.3819 saat, lajakan 0.298% dan Ralat Keadaan Mantap (SSE) 0.133 darjah; bagi kestabilan sudut pic, masa kenaikan 0.2309 saat, masa penetapan 0.7091 saat, lajakan 0.1224%, dan Ralat Keadaan Mantap (SSE) 0.0239 darjah. Pendekatan LQR-NN berkesan mengurangkan lajakan berbanding kawalan tradisi Pengatur Kuadratik Linear (LQR), dengan itu mengurangkan ayunan. Tambahan, LQR-NN dapat mengurangkan Ralat Keadaan Mantap (SSE), sebanyak 0.074 darjah bagi gerakan putaran guling dan 0.035 darjah bagi gerakan putaran anggul.

KEYWORDS: *Unmanned Aerial Vehicle, Optimal Control, Adaptive, Linear Quadratic Regulator, Neural Network*

https://doi.org/10.31436/iiumej.v25i2.3119

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs), more commonly known as drones, have become a very important and transformational technology in various aspects of human life and the industrial sector [1]. The current use of UAVs has various positive implications that reflect significant impacts in various fields, such as monitoring and surveillance, photography, delivery of goods, precision agriculture, infrastructure monitoring, security, and so on [2].

One type of UAV that is the focus of development in carrying out this task is the flying wing [3]. A flying wing is an aircraft design that does not have a conventional fuselage or tail and consists almost entirely of wings [4]. This type of UAV has superior aerodynamic efficiency, which provides the ability to fly longer distances compared to other designs [5]. In addition, they have good maneuverability, resistance to different environmental conditions, and the ability to carry various types of sensor payloads and cargo [6].

Of course, in carrying out these sector missions, the flying wing UAV is supported by an autonomous system that requires good stability control [7]. Stability control is the main factor in the success of an aircraft flight. Without good stability, an aircraft cannot maintain the correct height, direction, attitude, and position, making it difficult to operate or even risk an accident [8].

One of the effective control systems for maintaining UAV flight stability is the Linear Quadratic Regulator (LQR). Recent research indicates that employing LQR control in managing UAV attitude enables swift error correction, ensuring the maintenance of desired steady-state values [9,10]. Utilizing LQR control allows the UAV to minimize system errors with stability responses across all three orientation angles (roll, pitch, and yaw) in less than 1 second [11]. Fundamentally, LQR is a control method grounded in optimal control theory [12], utilized to design optimal controllers for linear systems, particularly within the context of dynamic systems represented by linear differential equations. LQR aims to generate control signals that minimize specified performance criteria [13]. Nonetheless, LQR encounters challenges in adapting to dynamic environmental changes, necessitating the development or incorporation of methods to optimize aircraft stability in constantly evolving scenarios [14].

Hence, there is a demand for a control system that exhibits robust characteristics and adapts to handle the intricacies of tasks and prevailing environmental conditions. The adaptability of LQR can be enhanced by integrating a Neural Network, enabling the adjustment of LQR feedback gain values to suit prevailing conditions, thereby optimizing its performance in that environment [15]. Implementing the LQR control system with a Neural Network on a quadcopter-type UAV has yielded fairly satisfactory outcomes. Neural networks have significantly enhanced the aircraft's response to rotational and translational motion, nearing the desired specifications [16]. Building upon this, this research focuses on LQR with Neural Network integration, referred to as LQR-NN. LQR-NN will be applied to a flying wing-type UAV model. Given the heightened susceptibility to dynamic changes in flying wing-type UAV flights, precise accommodation, and high performance are anticipated to mitigate errors.

2. UAV FLYING WING MODELING

Designing a control system requires kinematics and dynamic modeling of the aircraft. Modeling begins by examining the forces acting on the UAV flying wing aircraft [17]. The forces are visualized in aircraft body motion, as in Figure 1 [18].

https://doi.org/10.31436/iiumej.v25i2.3119

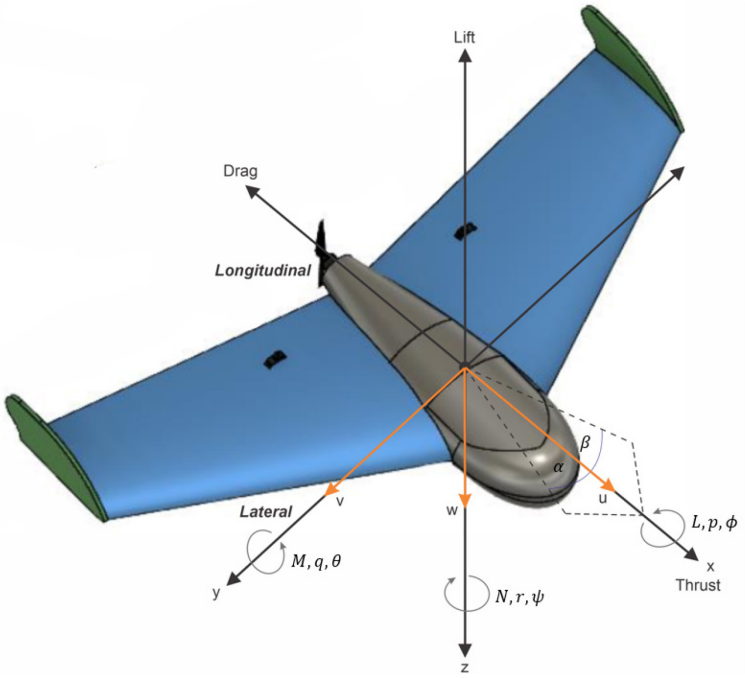


Figure 1. Flying wing UAV motion axis

During flight, an airplane can rotate along three axes (x, y, and z) concerning its center of mass. The aircraft's position control system typically manages its angular orientation by adjusting angles. Various aerodynamic forces come into play when an object is airborne, including lift, thrust, weight, and drag. These forces are crucial in determining the aircraft's behavior and maneuverability in the air.

Aircraft do not consistently maintain a strictly horizontal path while advancing [19]. This phenomenon arises from either air deflection from the front along the x-axis or from the side along the y-axis. Air deflection from the front results in the angle of attack ($\alpha - alpha$), whereas deflection from the side leads to the aircraft's sideslip angle ($\beta - beta$). The equations for these two angles deflection are as follows:

$$\alpha = tan^{-1}\frac{w}{u} \tag{1}$$

$$\beta = tan^{-1}\left(\frac{v}{\sqrt{u^2+w^2}}\right) \tag{2}$$

The system uses the North East Down (NED) frame for the Earth and aircraft body frame. So, the body frame and the NED frame can be related using a rotation matrix. The formula of a rotation matrix is defined in Eq. (1).

$$\begin{bmatrix} cos\psi cos\theta & -sin\psi cos\phi + cos\psi sin\theta sin\phi & sin\psi sin\phi + cos\psi cos\phi sin\theta \\ sin\psi cos\theta & cos\psi cos\phi + sin\phi sin\theta sin\psi & -cos\psi sin\phi + sin\theta sin\psi cos\phi \\ -sin\theta & cos\theta sin\phi & cos\theta cos\phi \end{bmatrix} \tag{3}$$

The correlation between frames generates rotational angles utilized for aircraft orientation. The rotation angles employed in the system are known as Euler angles. The Euler angles use

https://doi.org/10.31436/iiumej.v25i2.3119

roll (ϕ), pitch (θ), and yaw (ψ) angles. Each orientation angle has a derivative of angular velocity as a form of controlled state.

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix} = \begin{bmatrix} 1 & 0 & -sin\theta \\ 0 & cos\phi & cos\theta sin\phi \\ 0 & -sin\phi & cos\theta cos\phi \end{bmatrix} \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \tag{4}$$

p is the roll rate or the angular velocity of the y-axis, q is the pitch rate or the angular velocity of the x-axis, and p is the yaw rate or the angular velocity of the z-axis. Where,

$$\dot{\phi} = \frac{d\phi}{dt} \tag{5}$$

$$\dot{\theta} = \frac{d\theta}{dt} \tag{6}$$

$$\dot{\psi} = \frac{d\psi}{dt} \tag{7}$$

Furthermore, the aircraft's translation velocity along each axis (x, y, and z) is interconnected. The translation velocity along each axis is dictated by alterations in the plane frame's translation relative to the ground frame.

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} cos\psi cos\theta & -sin\psi cos\phi + cos\psi sin\theta sin\phi & sin\psi sin\phi + cos\psi cos\phi sin\theta \\ sin\psi cos\theta & cos\psi cos\phi + sin\phi sin\theta sin\psi & -cos\psi sin\phi + sin\theta sin\psi cos\phi \\ -sin\theta & cos\theta sin\phi & cos\theta cos\phi \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \tag{8}$$

where u is the translational velocity of the x-axis, v is the translational velocity of the y-axis, w is the translational velocity of the z-axis.

Both kinematic and dynamic models can characterize the aircraft system. These models were employed to formulate equations for the forces, moments, and orientation angles present within the system, which are visualized in the aircraft's two motions. The aircraft's motion is classified into translational and rotational components [20].

The determination of an aircraft's translational movement is carried out using the Newton-Euler law, which is based on Newton's second law.

$$\sum F = m.a \tag{9}$$

where m is mass (kg), and a is acceleration ($\frac{m}{s^2}$).

$$\sum F = F + F_{gravity} \tag{10}$$

$$\sum F = m\frac{d}{dt}v_T = m\frac{d}{dt}v_T + (\omega x v_T) \tag{11}$$

where v_T is the translational velocity ($\frac{m}{s}$) and ω is angular velocity.

The linear and total angular vector velocities can be calculated with Eq. (12) and (13),

$$v_T = iu + jv + kw \tag{12}$$

$$\omega = ip + jq + kr \tag{13}$$

where i, j, k are vector directions. The derivate of the linear velocity can be calculated with Eq. (14).

https://doi.org/10.31436/iiumej.v25i2.3119

$$\frac{d}{dt}v_T = i\dot{u} + j\dot{v} + k\dot{w} \tag{14}$$

Thus, the states that occur on the plane can be written in the following matrix [21],

$$\omega x v_T = \begin{bmatrix} i & j & k \\ p & q & r \\ u & v & w \end{bmatrix} = (qw - vr) + (ur - pw) + (pv - uq) \tag{15}$$

By involving linear and angular velocity, the total forces acting on the system can be expressed,

$$\sum F = \{i\dot{u} + j\dot{v} + k\dot{w} + ((qw - vr) + (ur - pw) + (pv - uq))\} \tag{16}$$

$$\sum F = \{i(\dot{u} + qw - vr) + j(\dot{v} + ur - pw) + k(\dot{w} + pv - uq)\} \tag{17}$$

$$\sum F = \{i\sum F_x + j\sum F_y + k\sum F_z\} \tag{18}$$

So, based on Eq. (17) and (18), the forces of each axis can be calculated with this formula,

$$\sum F_x = \dot{u} + qw - vr \tag{19}$$

$$\sum F_y = \dot{v} + ur - pw \tag{20}$$

$$\sum F_z = \dot{w} + pv - uq \tag{21}$$

Every axis of the plane is affected by the force of gravity. Thus, Eq. (19) to (21) will transform into Eq. (22) to (24), where g is the gravitational acceleration.

$$\frac{F_x - mg.sin\theta}{m} = \dot{u} + qw - vr \tag{22}$$

$$\frac{F_y + mg.cos\theta sin\phi}{m} = \dot{v} + ur - pw \tag{23}$$

$$\frac{F_z + mg.cos\theta cos\phi}{m} = \dot{w} + pv - uq \tag{24}$$

Because $F_x = X$, $F_y = Y$, and $F_z = Z$, the forces equation on each plane axis can be written as,

$$\frac{X}{m} - g.sin\theta = \dot{u} + qw - vr \tag{25}$$

$$\frac{Y}{m} + g.cos\theta sin\phi = \dot{v} + ur - pw \tag{26}$$

$$\frac{Z}{m} + g.cos\theta cos\phi = \dot{w} + pv - uq \tag{27}$$

Furthermore, the rotational motion of an aircraft is defined through the angular momentum acting on the system. The momentum has the following equation,

$$H = I\omega \tag{28}$$

where H is angle momentum ($\frac{\text{kg.m}^2\text{.rad}}{\text{s}}$) and I is inertia moment (kg.m^2). Eq (29) can describe the inertia moment,

$$I = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix} \tag{29}$$

The torque that occurs in the system is

$$M = I\frac{d}{dt}(\omega + \omega x\omega) + \omega xH \tag{30}$$

when $\omega x\omega = 0$, then Eq. (30) can be rewritten as in Eq. (31) and (32) [22].

https://doi.org/10.31436/iiumej.v25i2.3119

$$\frac{d}{dt}\omega = ip + jq + kr \tag{31}$$

$$\omega xH = \begin{vmatrix} i & j & k \\ p & q & r \\ h_x & h_y & h_z \end{vmatrix} \tag{32}$$

The evaluation of the determinant with unit vectors i, j, and k, along with angular momentum projections, can be expressed as a cross-product calculation in Eq. (33).

$$\omega xH = \left(qh_z - rh_y\right)i + (rh_x - ph_z)j + \left(ph_y - qh_x\right)k \tag{33}$$

The moment of inertia that occurs on the three axes of the aircraft is described in Eq. (32), with h_x, h_y, and h_z are defined in Eq. (35).

$$H = I\omega = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \tag{34}$$

$$\begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix} = \begin{bmatrix} I_{xx}p - I_{xy}q - I_{xz}r \\ -I_{xy}p + I_{yy}q - I_{yz}r \\ -I_{xz}p - I_{yz}q + I_{zz}r \end{bmatrix} \tag{35}$$

The aircraft has a symmetrical shape about the XZ axis, so $I_{xy} = I_{yz} = 0$. So, Eq. (34) transform to Eq. (36).

$$\begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix} = \begin{bmatrix} I_{xx}p - I_{xz}r \\ I_{yy}q \\ -I_{xz}p + I_{zz}r \end{bmatrix} \tag{36}$$

Substituting Eq. (36) into Eq. (33) yields:

$$\omega xH = \left(q(-I_{xz}p + I_{zz}r) - r(I_{yy}q)\right)i + \left(r(I_{xx}p - I_{xy}q) - p(-I_{xz}p + I_{zz}r)\right)j + \left(p(I_{yy}q) - q(I_{xx}p - I_{xy}q)\right)k \tag{37}$$

$$\omega xH = \left(-I_{xz}pq + I_{zz}qr) - I_{yy}qr\right)i + (I_{xx}pr - I_{xz}r^2 + I_{xz}p^2 - I_{zz}pr)j + \left(I_{yy}pq - I_{xx}pq + I_{xz}qr\right)k \tag{38}$$

The derivate of angle momentum can be calculated,

$$I\frac{d}{dt}\omega = \begin{bmatrix} I_{xx}\dot{p} - I_{xz}\dot{r} \\ I_{yy}\dot{q} \\ -I_{xz}\dot{p} + I_{zz}\dot{r} \end{bmatrix} \tag{39}$$

According to Eq. (30), the torque in each axis rotation can be expressed by adding Eq.(38) and (39)[23],

$$M_x = L = I_{xx}\dot{p} - I_{xz}(\dot{r} + pq) + (I_{zz} - I_{yy})qr \tag{40}$$

$$M_y = M = I_{yy}\dot{q} - I_{xz}(p^2 + q^2) + (I_{xx} - I_{zz})pr \tag{41}$$

$$M_z = N = I_{zz}\dot{r} - I_{xz}\dot{p} + \left(I_{xx} - I_{yy}\right)pq + I_{xz}qr \tag{42}$$

This study's stability observations were conducted solely on rotational motions around the x and y axes. Consequently, the force affecting motion along the z-axis can be disregarded. Therefore, the torque equations in Eq. (40) to (42) can be simplified to:

$$M_x = L = I_{xx}\dot{p} + (I_{zz} - I_{yy})qr \tag{43}$$

$$M_y = M = I_{yy}\dot{q} + (I_{xx} - I_{zz})pr \tag{44}$$

Eq. (43) represents the rotational movement of the aircraft about the x-axis, incorporating x-axis inertia (I_{xx}), roll angular acceleration ($\dot{p}$), pitch rate (q), and yaw rate (r). This can be expressed as a transfer function in Eq. (45).

$$\dot{p} = \frac{L}{I_{xx}} - \frac{(I_{zz} - I_{yy})qr}{I_{xx}} \tag{45}$$

Similarly, in Eq. (44), which depicts the rotational motion of the aircraft around the y-axis, incorporating y-axis inertia (I_{yy}), pitch angular acceleration ($\dot{q}$), roll rate (p), and yaw rate (r). This can be formulated into a transfer function in Eq. (46).

$$\dot{q} = \frac{M}{I_{yy}} - \frac{(I_{xx} - I_{zz})pr}{I_{yy}} \tag{46}$$

3. LQR-NN Control Design

A linear quadratic regulator is a control method that is called linear because the controller's model and shape are linear. It is also called quadratic because the cost function is quadratic. Meanwhile, it is called a Regulator because the reference is not a function of time. Figure 2 depicts the block diagram of the LQR control system.

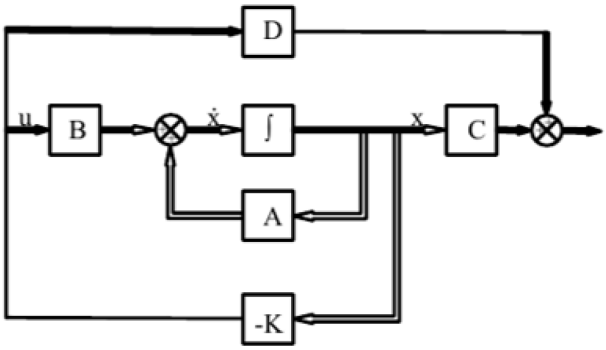


Figure 2. Full state feedback LQR

The LQR control system works on a linear system with calculations shown in Eq. (47) to (50).

$$\dot{\boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{Bu} \tag{47}$$

$$\boldsymbol{y} = \boldsymbol{Cx} + \boldsymbol{Du} \tag{48}$$

$$\boldsymbol{u} = -\boldsymbol{Kx} \tag{49}$$

$$\dot{\boldsymbol{x}} = (\boldsymbol{A} - \boldsymbol{BK})\boldsymbol{u} \tag{50}$$

In designing the control system to be optimal, energy minimization (cost function/quadratic function) is carried out, which is defined through the performance index in the interval [t0, ∞] in Eq. (51)[24]. This function involves Q and R matrices to make the system matrix non-negative.

$$\boldsymbol{J} = \int_{t_0}^{\infty} (\boldsymbol{x}^T \boldsymbol{Qx} + \boldsymbol{u}^T \boldsymbol{Ru}) dt \tag{51}$$

K determines the index performance to reach the minimum value. Finding the optimal value of K requires solving a quadratic regulator using the Riccati equation. The Riccati equation is a first-order form of a differential equation, which is an unknown quadratic function. This solution adds a differential equation with a constant P matrix [25]. This equation is shown in Eq. (52).

https://doi.org/10.31436/iiumej.v25i2.3119

$$\boldsymbol{J} = \int_{t_0}^{\infty} \frac{d}{dt}(\boldsymbol{x}^T \boldsymbol{P}\boldsymbol{x})\, dt = \boldsymbol{x}^T(0)\boldsymbol{P}\boldsymbol{x}(0) \tag{52}$$

From the reduction carried out by the Riccati equation, the determination of the K value is finalized using the Hamilton – Jacobian – Bellman procedure, which is calculated via Eq. (53) and produces Eq. (54) as the final solution.

$$\boldsymbol{K} = \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} \tag{53}$$

$$\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{Q} - \boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} \tag{54}$$

Based on the explanation of the equations and characteristics of the LQR control system, it requires a system model in the form of a linear approach. Therefore, referring to the model equation in the previous subsection, the linearization equation of the aircraft model can be written as follows [26],

$$\begin{bmatrix} U \\ \dot{U} \\ V \\ \dot{V} \\ W \\ \dot{W} \\ P \\ \dot{P} \\ Q \\ \dot{Q} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -W \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & W & 0 & 0 \\ 0 & 0 & 0 & Q & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -V & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{(I_{yy}-I_{zz})}{I_{xx}} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{(I_{zz}-I_{xx})}{I_{yy}} & 0 & 0 \end{bmatrix} \begin{bmatrix} X_E \\ U \\ Y_E \\ V \\ Z_E \\ W \\ \phi \\ P \\ \theta \\ Q \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{m} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{m} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{m} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{I_{xx}} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{I_{yy}} \end{bmatrix} \begin{bmatrix} F_x \\ F_y \\ F_z \\ L \\ M \end{bmatrix} \tag{55}$$

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_E \\ U \\ Y_E \\ V \\ Z_E \\ W \\ \phi \\ P \\ \theta \\ Q \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} F_x \\ F_y \\ F_z \\ L \\ M \end{bmatrix} \tag{56}$$

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

In this research, the control focus is on the aircraft's attitude motion, which includes controlling roll and pitch rotation motion. Therefore, Eq. (55) and (56) are reduced to a new state space from Eq. (57) to (58).

$$\begin{bmatrix} P \\ \dot{P} \\ Q \\ \dot{Q} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{(I_{yy}-I_{zz})}{I_{xx}} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{(I_{zz}-I_{xx})}{I_{yy}} & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi \\ P \\ \theta \\ Q \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{I_{xx}} & 0 \\ 0 & 0 \\ 0 & \frac{1}{I_{yy}} \end{bmatrix} \begin{bmatrix} L \\ M \end{bmatrix} \tag{57}$$

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$$

https://doi.org/10.31436/iiumej.v25i2.3119

$$\begin{bmatrix} y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \phi \\ P \\ \theta \\ Q \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} L \\ M \end{bmatrix} \quad (58)$$

$$\boldsymbol{y} = \boldsymbol{C} \boldsymbol{x} + \boldsymbol{D} \boldsymbol{u}$$

Then, to handle changes in system behavior toward the LQR, the control system will be assisted by a neural network. The neural network can adjust the K LQR gain value to changes in system dynamics both internally and externally so that the aircraft can provide high-stability flights. Therefore, the design of the anatomical structure of the aircraft control system is shown in the block diagram in Figure 3.

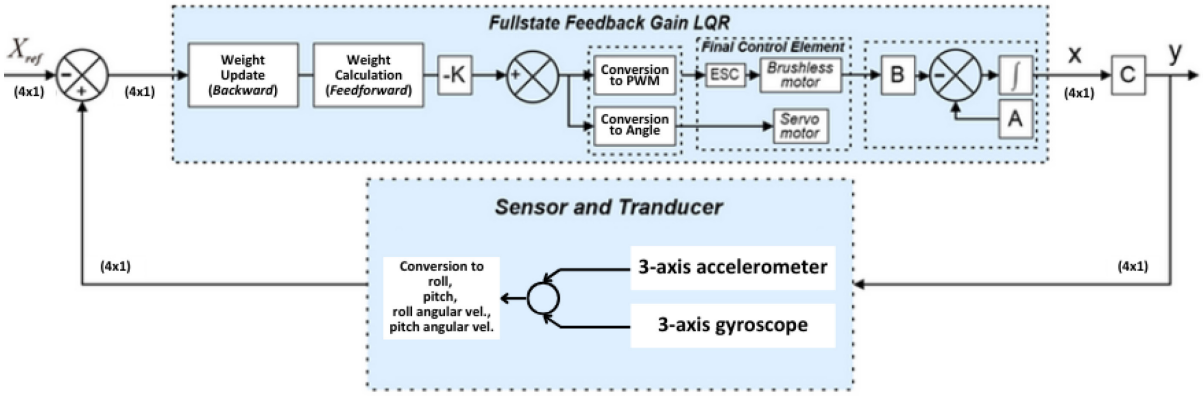


Figure 3. Control system design of flying wing UAV

In this system, there are system input values, which can be predetermined or variable values, which can be changed through control with a remote controller. The reference input value will be compared with the sensor reading value. The difference between the sensor reading and the set point value is called the error value. This error value is used as a reference for the controller to control the system so that it matches the desired reference value or set point. The output from the final control element usually also contains interference from outside the system. The output of the final control element, which is different from the input disturbance, will be considered a running system process. The system will analyze the output of the process by reading through the sensor used. The analysis is done to improve system control if the control is not as desired [27].

The control system on the UAV flying wing for attitude control uses full-state feedback control with four states from the sensor input. The input results are then entered into the full-state feedback system by multiplying the input results by the constant K in a matrix (4x1). The K constant value in this system was previously processed using an artificial neural network system. This artificial neural network system will change the K value tuned using the LQR system with constant Q and R values on each axis and then operate with Riccati calculations. Entering this artificial neural network is the same as entering the system on a flying wing UAV. The u value is the process input calculated by multiplying the feedback gain with the system's state to be controlled. This control input signal will then be used to control the speed of each rotor. Each rotor is a final control element in the UAV flying wing system.

The neural network architecture that works on the LQR control system is shown in Figure 4 [28],

https://doi.org/10.31436/iiumej.v25i2.3119

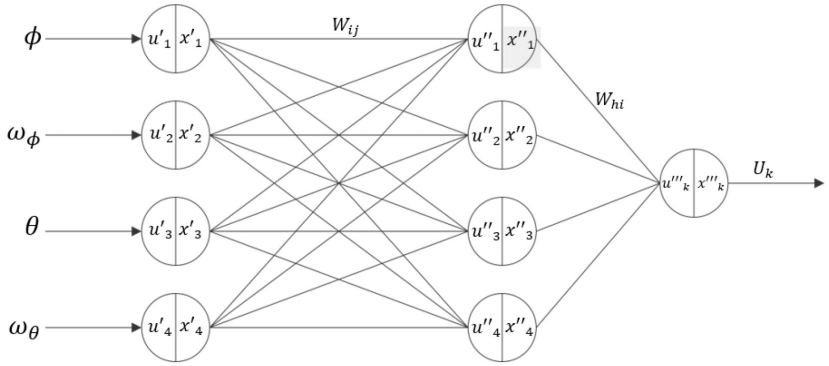


Figure 4. Neural network architecture

3.1 Input Layer

The input layer is input to the neural network system. Enter this system using four entries. Entering the neural network using input as in Figure 4 is intended so that the input to the neural network is the same as the input in the UAV flying wing control system itself.

3.2 Hidden Layer

In the hidden layer of the artificial neural network, six nodes are used. The input from the hidden layer in the artificial neural network is the sum of the multiplication of the Wij weights with the activation value of the node. Where i is the sequence of input nodes, and j is the sequence of hidden nodes.

3.3 Output Layer

In the output layer, there is one node that obtains a value from the sum of the multiplication of weights $Wh1$ and the output from the hidden node x 2. Activation at each node in the three layers uses the sigmoid activation function. The activation function aims to keep the values of each node in each layer from being excessive and still correlated with each other at each layer. This neural network system uses initial weight initialization Wij with Eq. (59).

$$W_{ij} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 1 & 1 \end{bmatrix} \quad (59)$$

Meanwhile, the neural network weight values from the hidden layer to the output layer use the initial value of the constant K resulting from tuning using the LQR method. Form the weight Whi using Eq. (60).

$$W_{hi} = [\mathrm{K}_{11}\ \mathrm{K}_{21}\ \mathrm{K}_{12}\ \mathrm{K}_{22}\ \mathrm{K}_{13}\ \mathrm{K}_{23}\ \mathrm{K}_{14}\ \mathrm{K}_{24}] \quad (60)$$

The weight value Whi in this system is a representation of the feedback constant value used in the flying wing UAV control system. The relationship between artificial neural networks and full-state feedback control here is in tuning the previous K value, which has been determined by the LQR method. The K value is entered as the initial weight value in the weight Whi.

https://doi.org/10.31436/iiumej.v25i2.3119

3.4 Back-propagation Algorithm

The artificial neural network system in this control design uses system error values originating from system output in real-time. The system response is analyzed by calculating the system output error value in the form of four states, which are represented by certain values. The error value in this system is in the form of a sensor reading value in the system, which is then compared with the system reference state value. The main objective of the artificial neural network in this system is to minimize the mean value of the system error squared (J). This system uses the error function Eq. (41). The value is the system output in the form of sensor readings. The value $\boldsymbol{xref}$ is a system reference value in the form of a system input or desire value. In the research $\boldsymbol{xref}$ the signal input of controllability state roll and pitch rotation is set at zero value. The value $y(n)$ is the value of the system output at each system state. The difference between these two values is a system error value. The n value is the number of states observed in this system, namely, four system states.

$$j = E_p = \sum_n \frac{1}{2} \left(y_{(n)} - X_{ref(n)}\right)^2 \tag{61}$$

The learning rate function with a weight change function of 0.01 is used. The weight value Whi will always change depending on the input to the artificial neural network system [29]. These weight values are then used to determine the output in the full-state feedback control system.

Determines the weight value from the hidden layer to the output layer.

To find the weight value from the node Whi hidden layer to the output layer, use Eq. (62) to (67).

$$w_{h1}(k+1) = w_{h1}(k) - \eta \frac{\partial j}{\partial w_{h1}} \tag{62}$$

where n is the learning rate with a value of 0.01. The delta J value is obtained by Eq. (62).

$$\frac{\partial j}{\partial w_{h1}} = \frac{\partial j}{\partial y} . \frac{\partial y}{\partial u} . \frac{\partial u}{\partial w_{h1}} \tag{63}$$

Where y is the response output from the UAV flying wing control system. The u_k Value is the output of the results of the artificial neural network calculation. The value Whi is the weight value from the hidden layer to the output layer of the artificial neural network. From Eq. (63), Eq. (64) is translated to Eq. (66) [30].

$$\frac{\partial j}{\partial y} = x_h^2(k) \tag{64}$$

$$\frac{\partial y}{\partial u} = \frac{\partial y}{\partial t} / \frac{\partial u}{\partial t} = \frac{\dot{y}}{\dot{u}} \approx \frac{\Delta y}{\Delta u} = \frac{y(k)-y(k-1)}{u(k)-u(k-1)} \tag{65}$$

$$\frac{\partial u}{\partial h_1} = \sum_n \frac{1}{2} \left(y_{(n)} - X_{ref(n)}\right)^2 \tag{66}$$

The value $xh1$ is the activation value of the hidden layer, and the y value is the output of the full-state feedback control system. Translating Eq. (62) to Eq. (66) yields Eq. (67), which can be used to obtain the new weight value $xh1$.

$$\frac{\partial j}{\partial w_{h1}} = x_h^2(k) . \frac{y(k)-y(k-1)}{u(k)-u(k-1)} . \sum_n \frac{1}{2} (y_{(n)} - X_{ref(n)})^2 \tag{67}$$

Determine the weight value from the insert layer to the hidden layer. To find the weight value from node i of the insert layer to node j of the hidden layer using Eq. (68).

$$w_{ij}(k+1) = w_{ij}(k) - \eta \frac{\partial j}{\partial w_{ij}} \tag{68}$$

where n is the learning rate with a value of 0.01. The value of delta J is obtained by Eq. (69).

$$\frac{\partial j}{\partial w_{ij}} = \frac{\partial j}{\partial y} \cdot \frac{\partial y}{\partial u} \cdot \frac{\partial u}{\partial w_{ij}} \tag{69}$$

The value Wij is the weight value from the hidden layer to the output layer of the artificial neural network. From Eq. (49), it is explained from Eq. (63) to Eq. (65). The x value is the value resulting from the activation of the hidden layer, so we can obtain Eq. (70) to get the new weight value Wij.

$$\frac{\partial j}{\partial w_{ij}} = W_{h1} X_i^1(k) . \frac{y(k)-y(k-1)}{u(k)-u(k-1)} \sum_n \frac{1}{2} \left(y_{(n)} - X_{ref(n)}\right)^2 \tag{70}$$

4. STRUCTURE AND PARAMETERS OF UAV FLYING WING

The aircraft used in this study have the dimensions shown in Table 1, while Figure 5 illustrates the size and position of the actuator.

Table 1. UAV flying wing dimension

Items	Dimension
Wingspan	1100 mm
Fuselage	500 mm
Winglet	200 mm
Motor Diameter	35 mm
Weight	1300 gram

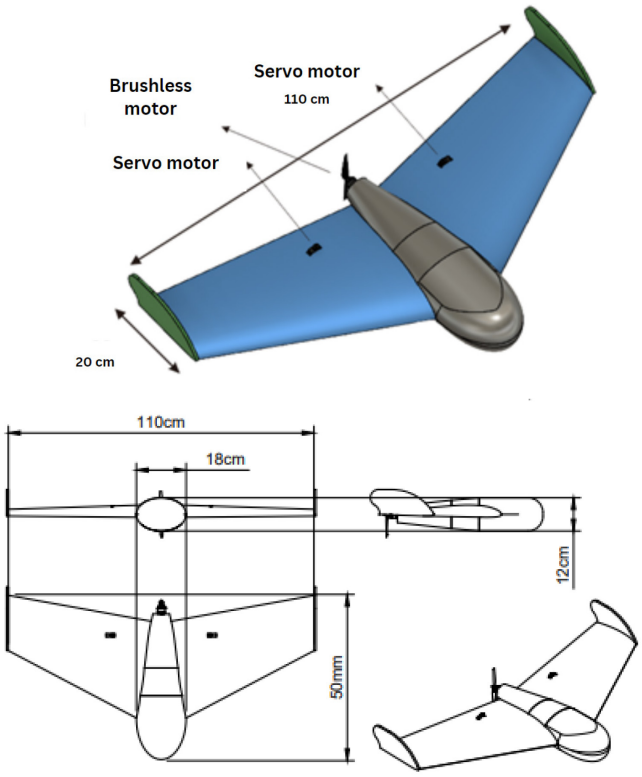


Figure 5. UAV flying wing structure

https://doi.org/10.31436/iiumej.v25i2.3119

The aircraft has dimensions of around 110 cm for the wing length, and the aircraft's flying weight is around 1.3 kg. The body of the aircraft is made from carbon fiber using a molding method to produce a body that is precise, strong, and light. The wings of the aircraft are made from high-density hard foam and coated with fiberglass. The airfoil design on the wing is chosen for the flat bottom type because wings with a flat bottom tend to provide better longitudinal stability. The center of pressure on a flat profile is closer to the balance point. This profile tends to have higher pitch stability, that is, the ability to keep the angle of attack relatively stable. So it can produce low flight speeds.

In proving the control system design, the important thing that is needed is the calculation parameters of the system on the aircraft. One important parameter is the inertia calculation for each axis. The inertia calculations used in this research are the x, y, and z-axis inertia, where each inertia is calculated according to the data in Table 2.

Table 2. Inertia parameters

Items	Length (m)	Width (m)	Height (m)	Mass (kg)	Shape
Center box (I_{KT})	9.18	0.44	0.08	1.182	briquette
Right-wing Part 1 (I_{SKN1})	0.115	0.27	0.03	0.0725	briquette
Right-wing part 2 (I_{SKN2})	0.115	0.2365	0.03	0.0675	briquette
Right-wing Part 3 (I_{SKN3})	0.115	0.2018	0.03	0.0625	briquette
Right-wing Part 4 (I_{SKN4})	0.115	0.1685	0.03	0.0675	briquette
Left-wing Part 1 (I_{SKR1})	0.115	0.27	0.03	0.0725	briquette
Left-wing part 2 (I_{SKR2})	0.115	0.2365	0.03	0.0675	briquette
Left-wing part 3 (I_{SKR3})	0.115	0.2018	0.03	0.0625	briquette
Left-wing part 4 (I_{SKR4})	0.115	0.1685	0.03	0.0675	briquette

x-axes inertia calculation (I_{xx})

$$I_{xx} = (I_{KTx}) + (I_{SKR1x}) + (I_{SKR2x}) + (I_{SKR3x}) + (I_{SKR4x}) + (I_{SKN1X}) + (I_{SKN2X}) + (I_{SKN3X}) + (I_{SKN4X})$$

$$I_{xx} = 1{,}97\ x\ 10^{-2} + 4{,}46\ x\ 10^{-4} + 3{,}19\ x\ 10^{-4} + 2{,}17\ x\ 10^{-4} + 1{,}40\ x\ 10^{-4} + +4{,}46\ x\ 10^{-4} + 3{,}19\ x\ 10^{-4} + 2{,}17\ x\ 10^{-4} + 1{,}40\ x\ 10^{-4}$$

$$\boldsymbol{I_{xx} = 2{,}194\ x\ 10^{-2}\ \text{Kg}.m^2}$$

y-axes inertia calculation (I_{yy})

$$I_{yy} = (I_{KTy}) + (I_{SKR1y}) + (I_{SKR2y}) + (I_{SKR3y}) + (I_{SKR4y}) + (I_{SKN1y}) + (I_{SKN2y}) + (I_{SKN3y}) + (I_{SKN4y})$$

$$I_{yy} = 3{,}82\ x\ 10^{-3} + 8{,}53\ x\ 10^{-5} + 7{,}94\ x\ 10^{-5} + 7{,}36\ x\ 10^{-5} + 6{,}77\ x\ 10^{-5} + 8{,}53\ x\ 10^{-5} + 7{,}94\ x\ 10^{-5} + 7{,}36\ x\ 10^{-5} + 6{,}77\ x\ 10^{-5}$$

$$\boldsymbol{I_{yy} = 4{,}434\ x\ 10^{-3}\ \text{Kg}.m^2}$$

z-axes inertia calculation (I_{zz})

$$I_{zz} = (I_{KTz}) + (I_{SKR1z}) + (I_{SKR2z}) + (I_{SKR3z}) + (I_{SKR4z}) + (I_{SKN1z}) + (I_{SKN2z}) + (I_{SKN3z}) + (I_{SKN4z})$$

$$I_{zz} = 2{,}23\ x\ 10^{-2} + 5{,}20\ x\ 10^{-4} + 3{,}89\ x\ 10^{-4} + 2{,}81\ x\ 10^{-4} + 1{,}99\ x\ 10^{-4} + 5{,}20\ x\ 10^{-4} + 3{,}89\ x\ 10^{-4} + 2{,}81\ x\ 10^{-4} + 1{,}99\ x\ 10^{-4}$$

$$\boldsymbol{I_{zz} = 2{,}50\ x\ 10^{-2}\ \text{Kg}.m^2}$$

https://doi.org/10.31436/iiumej.v25i2.3119

5. RESULT AND DISCUSSIONS

In this research, the implementation of the LQR-NN control system was carried out through simulation using MATLAB. The application is preceded by searching for the K LQR gain value from the best response obtained. The calculation of the feedback gain K value for each state involved is sought by tuning the value of the Q element. From the experiments that have been carried out, the best Q value from the results of roll and pitch anti-rotation motion control is shown in Table 3, where the data shows the best conversion of Q and R values to obtain the most optimal K gain value [31].

Table 3. Conversion metrics **Q** values to gain **K** values

Q	R	Gain K
$\begin{bmatrix} 1.0245 & 0 & 0 & 0 \\ 0 & 0.00049 & 0 & 0 \\ 0 & 0 & 1.033 & 0 \\ 0 & 0 & 0 & 0.004 \end{bmatrix}$	1	$\begin{bmatrix} 1.0124 & 0.2119 & 0 & 0 \\ 0 & 0 & 1.0149 & 0.1140 \end{bmatrix}$

The best K gain value can control the aircraft's attitude, both roll and pitch angles, with the signals shown in Figures 6 and 7. Figure 6 is for roll stability, and Figure 7 is for pitch stability.

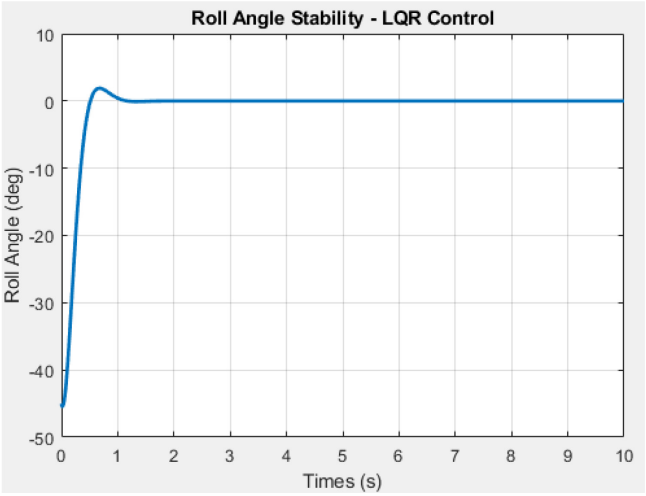


Figure 6. Roll angle stability using LQR

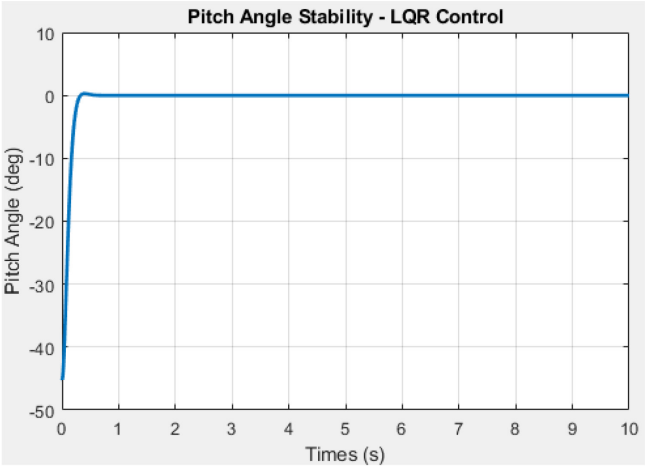


Figure 7. Pitch angle stability using LQR

https://doi.org/10.31436/iiumej.v25i2.3119

The two aircraft flight movements shown in Figures 6 and 7 have stability characteristics, which are explained through the transient response in Table 4.

Table 4. Transient response of roll and pitch stability using LQR control

Components	Roll	Pitch
Rise time	0.4682s	0.2455s
Overshoot	4.217%	0.6311%
Settling time	1.149s	0.7091s
Steady-state error (SSE)	-0.007deg	-0.0001deg

According to Table 5, the pitch angle demonstrates superior stability compared to the roll angle. With a rise time of 0.2227 seconds and a settling time of 0.44 seconds, the pitch response is faster than that of the roll. Additionally, the pitch displays minimal overshoot, measuring at 0.6311% or under 1%. Furthermore, its steady-state error is -0.0001 degrees, smaller than the roll's error of -0.0007 degrees. This discrepancy can be attributed to the aircraft's predominantly wing-based construction, making it susceptible to lateral disturbances such as side interference or slip. Moreover, larger torque occurrences slow down the control system's ability to mitigate disturbances effectively.

Apart from this, the characteristics of the roll and pitch angles show good stability. Tests carried out on the control of both states prove that the LQR control is able to maintain the stability of the roll and pitch torque against the given disturbances. Apart from that, the test shows that the greater the weighting of the Q element, the greater the K gain value will be. The greater the K gain value, the greater the torque on the aircraft system, making the system more responsive compared to a small K gain value. Of the two controlled states, roll and pitch motion will not provide the same character to the system. These two states have complementary roles in stabilizing. It can be seen by giving the K gain value to the angular speed. The greater the K value given to this state, the more the response that occurs in the system will be suppressed or dampened. Angular velocity is the first derivative of the roll orientation angle, which will change if there is a change in angular position relative to the time of change, so the faster the change, the greater the resulting angular velocity [9].

From the best K gain value obtained for each control state, K is used as a reference value in the change value of the artificial neural network. The artificial neural network model has one hidden layer, where the training process is carried out via online learning based on the errors that occur. The error value will be fed back using the backpropagation method, where the weight that changes is the gain value K with a learning rate of 0.01. Setting the learning rate value at this value so as not to force the system to accept signal changes that are too large, which can cause the aircraft to experience turbulence, resulting in a flight stall.

In this initial comparison test, the system will be subjected to a 45-degree deviation as the initial value in the step function. This approach is employed to assess the robust characteristics of the aircraft when stabilizing with relatively large deviations. The outcomes of controlling the stability of the aircraft's attitude using LQR-NN control are depicted in Figures 8 and 9.

https://doi.org/10.31436/iiumej.v25i2.3119

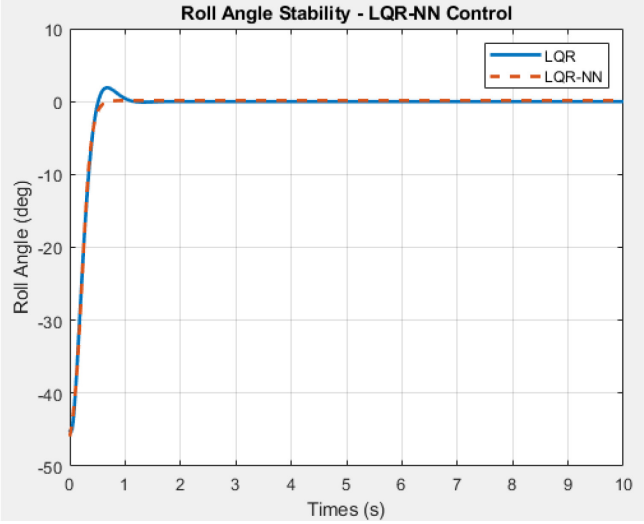


Figure 8. The comparison of roll angle stability between using LQR and LQR-NN

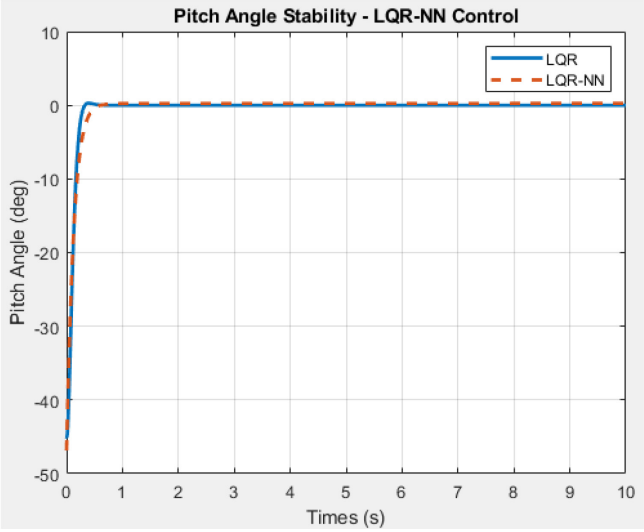
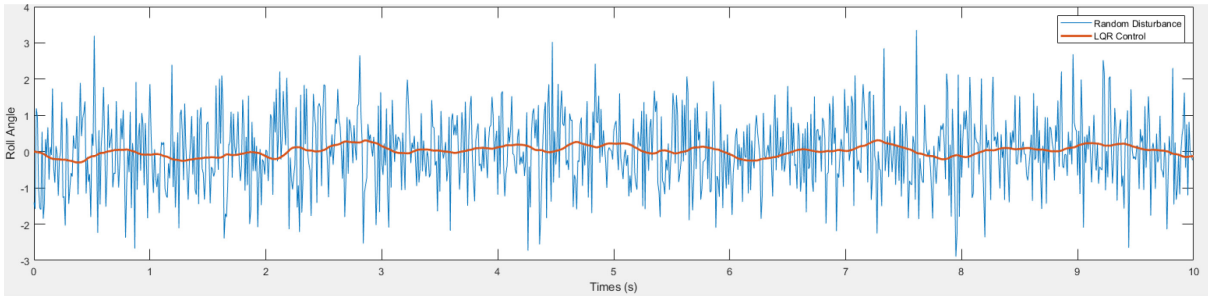


Figure 9. The comparison of pitch angle stability between using LQR and LQR-NN

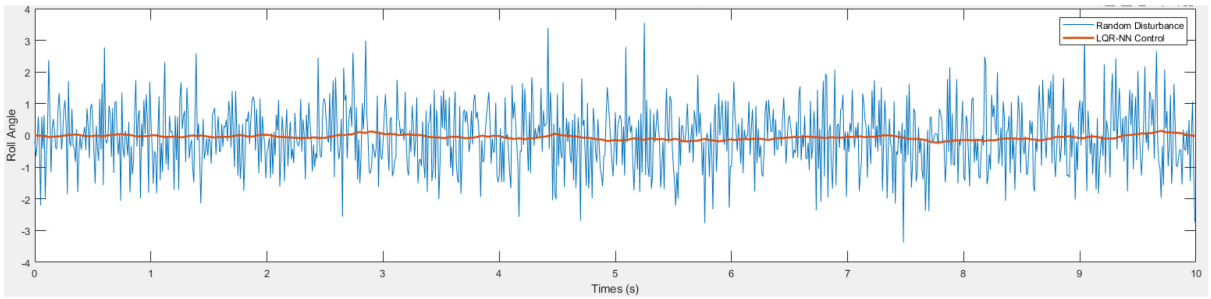
The results of this control were obtained by setting the net on the neurons to four nets with an epoch of 82. The experiments revealed that setting the epoch value above 82 causes oscillations in the system. In contrast, values below 82 have minimal impact on the neural network, thus ensuring stability across all controlled states without altering their characteristics [32]. LQR-NN control of roll and pitch angles provides the characteristics shown in Table 5.

Table 5. The comparison of the transient response of roll and pitch stability between using LQR and LQR-NN control

Components	LQR		LQR-NN		Minimum Requirements of Aircraft
	Roll	Pitch	Roll	Pitch	
Rise time	0.4682s	0.2455s	0.4682s	0.2309s	< 1s
Overshoot	4.217%	0.6311%	0.298%	0.1224%	< 3%
Settling time	1.149s	0.7091s	1.3819s	0.7091s	< 3s

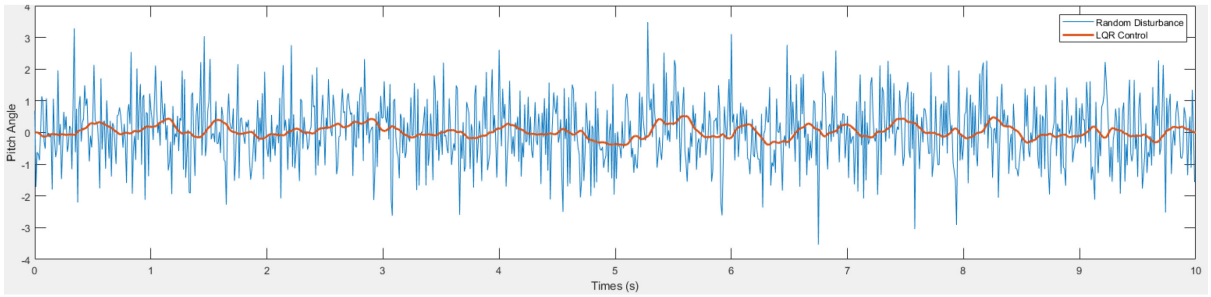


(a)

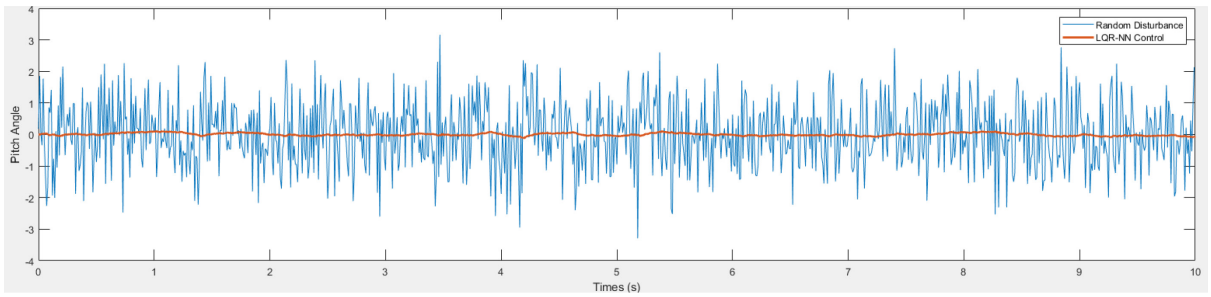


(b)

Figure 10. (a) Roll angle stability control using LQR in handling random disturbances (b) Roll angle stability control using LQR-NN in handling random disturbances



(a)



(b)

Figure 11. (a) Roll angle stability control using LQR in handling random disturbances (b) Roll angle stability control using LQR-NN in handling random disturbances

Based on Table 6, roll and pitch flight attitude with LQR-NN control has better equilibrium characteristics than LQR. The LQR-NN control can minimize overshoot during pitch and roll rotational motion to less than 1%, whereas the LQR control only meets the minimum required pitch motion criteria. Significant changes occurred in the overshoot value, which was further reduced by neural network predictions. This is important in control because the greater the

overshoot, the greater the possibility of producing multiple overshoots, which can make the flight unstable and even cause the plane to crash [33].

Furthermore, the management of steady-state error (SSE) by the LQR-NN control system surpasses that of the LQR control system. LQR-NN effectively reduces SSE to a lesser degree. This is evidenced by the graphical simulation results depicted in Figures 10 and 11.

In this observation, the system experienced random disturbances for 10 seconds. The simulation results demonstrate that LQR-NN effectively maintains SSE at 0.074 degrees for roll rotation motion and 0.035 degrees for pitch rotation motion. In contrast, LQR control maintains SSE at 0.113 degrees for roll rotation motion and 0.150 degrees for pitch rotation motion during flight. LQR-NN exhibits greater adaptability to the occurring disturbances. LQR-NN has been deemed to meet the criteria for a highly proficient control system because it satisfies all the minimum specification requirements for system response stability, particularly in the flight attitude of the UAV flying wing aircraft.

6. CONCLUSION

LQR-NN control has been successfully modeled and applied to control the stability of the roll and pitch flight attitude of UAV flying-wing aircraft. The net settings given to the hidden layer neural network are four nets with an epoch of 82. The test results show that roll angle stability has characteristics of a rise time of 0.4682 seconds, settling time of 1.3819 seconds, overshoot of 0.298%, and SSE of 0.133 degrees, while pitch angle stability has a rise time of 0.2309 seconds, settling time of 0.7091 seconds, overshoot of 0.1224% and SSE of 0.0239 degrees. LQR-NN can reduce overshoot that occurs in the LQR controller so that it can minimize oscillations that will occur. Additionally, LQR-NN effectively mitigates the steady-state error (SSE), reducing it to 0.074 degrees for roll rotation motion and 0.035 degrees for pitch rotation motion.

REFERENCES

[1] Mátyás P, Nagy M. (2019). Brief history of uav development. Repüléstudományi Közlemények, 31(1): 155-6. https://doi.org/10.32560/rk.2019.1.13

[2] Andres AM, Yilei H, Yuhan J. (2023). A review of unmanned aerial vehicle applications in construction management: 2016–2021. Standards. 3(2): 95–109. https://doi.org/10.3390/standards3020009

[3] Tri KP, Oktaf AD, Tri S. (2020). Model of linear quadratic regulator (lqr) control system in waypoint flight mission of flying wing uav. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 12(4): 43–49. https://jtec.utem.edu.my/jtec/article/view/5696

[4] Stavros K, Chris B, Pavlos K, Pericles P, Kyros Y. (2023). Parametric investigation of canards on a flying wing uav using the taguchi method. Aerospace, 10(3): 264. https://doi.org/10.3390/aerospace10030264

[5] Seyhun D. (2023). Aerodynamic performance comparison of airfoils in flying wing uav. International Journal of Innovative Engineering Applications, 7(1): 123–127. https://doi.org/10.46460/ijiea.1169652

[6] Panjani D, Indira P, Chandu A. (2022). Design of folded wing mechanism for unmanned aerial vehicle (uav). Materials Today: Proceedings, 62: 4117–4125. https://doi.org/10.1016/j.matpr.2022.04.660

[7] Yankui W, Xiangxi T, Tao L. (2020). Lateral stability and control of a flying wing configuration aircraft. Journal of Physics: Conference Series, 1509(1): 012022. https://doi.org/10.1088/1742-6596/1509/1/012022

[8] Jose DH, Camilo E, Juan PA, Gustavo S, Juliana AN, Jorge IG. (2023). Two-way coupled aero-structural optimization of stable flying wings, Aerospace. 10(4): 346. https://doi.org/10.3390/aerospace10040346

[9] Oktaf AD, Andi D, Tri KP. (2017). Model of linear quadratic regulator (lqr) control method in hovering state of quadrotor. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(3): 135–143. https://jtec.utem.edu.my/jtec/article/view/1589

[10] Oktaf AD, Faisal FR. (2022). Peningkatan kestabilan quadrotor menggunakan kendali linear quadratic regulator dengan kompensasi integrator dalam mempertahankan posisi. Buletin Ilmiah Sarjana Teknik Elektro. 4(2): 62–75. https://doi.org/10.12928/biste.v4i2.6808

[11] Riccardo D, Piero T, Salvatore FG, Alessandro M. (2021). Recent advances in unmanned aerial vehicle forest remote sensing - a systematic review, part i: a general framework. Forests. 12(3): 327. https://doi.org/10.3390/f12030327

[12] Jingjing B, Afshin M, Maryam F, Mehran M. (2019). Lqr through the lens of first order methods: discrete-time case. https://doi.org/10.48550/ARXIV.1907.08921

[13] Aisha SE, Seref NE. (2022). Robust lqr and lqr-pi control strategies based on adaptive weighting matrix selection for a uav position and attitude tracking control. Alexandria Engineering Journal, 61(8): 6275–6292. https://doi.org/10.1016/j.aej.2021.11.057

[14] Jinsong Z, Yan L, Lin L. (2023). LQR-based adaptive optimal control for aircraft engine. Proceedings of 2023 Chinese Intelligent Systems Conference. CISC 2023. Lecture Notes in Electrical Engineering, vol 1090. Springer, Singapore. https://doi.org/10.1007/978-981-99-6882-4_28

[15] Faisal FR, Tri KP. (2019). Penalaan mandiri full state feedback lqr dengan jst tiruan pada kendali quadrotor. Indonesian Journal of Electronics and Instrumentation Systems (IJEIS). 9(1): 21-32. https://doi.org/10.22146/ijeis.37212

[16] Chenxi S, Tao L, Kui Y. (2013). Balance control of two-wheeled self-balancing robot based on Linear Quadratic Regulator and Neural Network. 2013 Fourth Int. Conf. Intell. Control Inf. Process. 1: 862–867. https://doi.org/10.1109/ICICIP.2013.6568193

[17] Huynh VN, Dinh PN, Nguyen TMN, Nguyen TD, Nguyen PL, Phung ST, Le THL, Dang XB. (2021). A lqr-based neural-network controller for fast stabilizing rotary inverted pendulum. 2021 International Conference on System Science and Engineering (ICSSE), 19–22. https://doi.org/10.1109/ICSSE52999.2021.9537940

[18] John BH. (2007). Modelling Simulation and Control of Fixed-wing UAV: CyberSwan. Institutt for teknisk kybernetikk.

[19] Tri KP, Abdul MF. (2021). Modeling and Simulation of The UX-6 Fixed-Wing Unmanned Aerial Vehicle. J Control Autom Electr Syst 32, 1344–1355. https://doi.org/10.1007/s40313-021-00754-5

[20] Burak E. (2019). Fault tolerant flight control applications for a fixed wing uav using linear and nonlinear approaches. https://doi.org/10.13140/RG.2.2.25180.03205

[21] Jie C, Jianxin L. (2021). Mathematical modeling and fault tolerant control of uav with wing layout. Journal of Physics: Conference Series, 1846(1): 012045. https://doi.org/10.1088/1742-6596/1846/1/012045

[22] Ruijie S, Zhou Z, Xiaoping Z. (2022). Stability control of a fixed full-wing layout uav under manipulation constraints. Aerospace Science and Technology, 120: 107263. https://doi.org/10.1016/j.ast.2021.107263

[23] Aman S, Gabriel FL, Jan S. (2023). Identifying aerodynamics of small fixed-wing drones using inertial measurements for model-based navigation. NAVIGATION: Journal of the Institute of Navigation, 70(4): navi.611. https://doi.org/10.33012/navi.611

[24] Yuqiong S, Song W, Anastasiia K, Yujing H. (2019). Attitude control of flying wing uav based on advanced adrc. IOP Conference Series: Materials Science and Engineering, 677(5): 052075. https://doi.org/10.1088/1757-899X/677/5/052075

[25] Shi Q, Cui H, Li F, Liu Y, Ju W, Sun Y. (2017). A hybrid dynamic demand control strategy for power system frequency regulation. CSEE J Power Energy Syst, 3(2):176–185. https://doi.org/10.17775/CSEEJPES.2017.0022

[26] Jianglin L, Dezong Z. (2023). Finding the lqr weights to ensure the associated riccati equations admit a common solution. IEEE Transactions on Automatic Control, 68(10): 6393–6400. https://doi.org/10.1109/TAC.2023.3234237

[27] Gembong ES, Wijaya K, Amroy CLG. (2019). Linear quadratic regulator controller (lqr) for ar. drone's safe landing. 2019 International Conference on Sustainable Information Engineering and Technology (SIET), 228–233. https://doi.org/10.1109/SIET48054.2019.8986078

[28] Mahmud S. (2021). Modelling and simulation of small scale fixedwing autonomous aerial vehicles. PhD Thesis. Sheffield Hallan University, Business, Technology and Engineering.

[29] Faisal FR, Tri KP. (2019). Penalaan mandiri full state feedback dengan lqr dan jst pada kendali quadrotor. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems), 9(1): 21. https://doi.org/10.22146/ijeis.37212

[30] Bailun J, Boyang L, Weifeng Z, Li-Yu L, Chih-Keng C, Chih-Yung C. (2022). Neural network based model predictive control for a quadrotor uav. Aerospace, 9(8): 460. https://doi.org/10.3390/aerospace9080460

[31] Yunlong G, Guixin Z, Tong Z. (2022). Based on backpropagation neural network and adaptive linear active disturbance rejection control for attitude of a quadrotor carrying a load. Applied Sciences, 12(24): 12698. https://doi.org/10.3390/app122412698

[32] Aris T, Suroto M, Munadi M, Joga DS. (2023). Application of driving behavior control system using artificial neural network to improve driving comfort by adjusting air-to-fuel Ratio. IIUM Engineering Journal, 24(2): 337–353. https://doi.org/10.31436/iiumej.v24i2.2781

[33] Oktaf AD, Tri KP, Aris N, Yasir MM. (2022). Enhancement of stability on autonomous waypoint mission of quadrotor using lqr integrator control. IIUM Engineering Journal, 23(1): 129–158. https://doi.org/10.31436/iiumej.v23i1.1803

https://doi.org/10.31436/iiumej.v25i2.3129

BUILDING A GIS MAP FOR FORECASTING THE MIR INDEX IN AN GIANG

THANH-NGHI DOAN[1,3*], KHANH TRAN THIEN NGUYEN[2,3]

[1]*Faculty of Information Technology, An Giang University, An Giang, Vietnam*
[2]*Department of Science and Postgraduate Management, An Giang University, An Giang, Vietnam*
[3]*Vietnam National University, Ho Chi Minh City*

**Corresponding author: dtnghi@agu.edu.vn*

(Received: 29 November 2023; Accepted: 8 February 2024; Published online: 15 July 2024)

ABSTRACT: The MIR aquatic plant signal is capable of predicting specific pollution sources of water, contributing significantly to the effective management of surface water resources in An Giang province. The use of aquatic plants in water pollution treatment brings about positive effects through natural self-purification processes as they consume organic and nutrient substances such as N and P. Therefore, it is crucial to develop a tool for monitoring and supervising aquatic plant species. This paper investigates the application of GIS technology to build a GIS map representing the current status of vegetation cover in An Giang province. The background layers of the GIS database, along with detailed attribute layers regarding species composition, dominant species, and vegetation area, will serve as the basis for managing, utilizing, conserving, and restoring vegetation cover in the research area. Additionally, a predictive model for MIR indices has been constructed using machine learning methods. The results indicate that the model has a coefficient of determination (R2) of 91.7% for the dependent variable MIR compared to the independent variables. Subsequently, these results are visually displayed on a GIS map at 18 monitoring points within An Giang province, enabling users to easily observe, compare, evaluate, and propose suitable solutions for surface water quality management.

ABSTRAK: Isyarat tumbuhan akuatik MIR mampu meramalkan sumber pencemaran air secara spesifik, iaitu penyumbang penting kepada pengurusan berkesan permukaan sumber air di wilayah An Giang. Penggunaan tumbuhan akuatik dalam rawatan pencemaran membawa kepada kesan positif melalui proses rawatan kendiri secara semula jadi kerana ia mengandungi bahan organik dan nutrien seperti N dan P. Oleh itu, sangat penting bagi membangunkan alat pemantauan dan pengawasan spesies tumbuhan akuatik. Kajian ini mengkaji aplikasi teknologi GIS bagi membangunkan peta GIS mewakili status terkini keseluruhan tumbuhan di wilayah An Giang. Lapisan latar belakang pangkalan data GIS bersama lapisan sifat-sifat terperinci berkenaan spesies komposit, spesies dominan, dan kawasan tumbuh-tumbuhan, dapat menyediakan asas kepada pengurusan, penggunaan, pemuliharaan, dan pemulihan tumbuh-tumbuhan meliputi kawasan kajian. Tambahan, model ramalan MIR dibangunkan menggunakan kaedah pembelajaran mesin. Dapatan kajian menunjukkan model ini mempunyai pekali penentu (R2) sebanyak 91.7% bagi pembolehubah MIR bersandar berbanding pembolehubah tak bersandar. Menyebabkan dapatan ini secara visual dapat dilihat pada peta GIS menggunakan 18 titik pantauan dalam wilayah An Giang province, membolehkan pengguna mudah melihat, membandingkan, menilai, dan mencadangkan solusi sesuai bagi pengurusan kualiti permukaan air.

KEYWORDS: *An Giang Province, GIS map, MIR index, Machine learning methods*

https://doi.org/10.31436/iiumej.v25i2.3129

1. INTRODUCTION

An Giang, situated within the southwest region of Vietnam's Mekong Delta, is a province characterized by an intricate network of rivers, canals, and channels, ensuring a consistent supply of surface water throughout the year. This hydrological setup confers numerous advantages in terms of biodiversity and sustenance for aquaculture endeavors [1]. The province's aquaculture sector has demonstrated substantial advancements in both quantity and quality, playing a pivotal role in propelling the province's socioeconomic progress. Nonetheless, the rapid expansion of aquaculture has triggered a range of concerning environmental repercussions, notably impacting the water bodies that receive aquaculture-related wastewater. Presently, the majority of wastewater stemming from fish farms across An Giang province is being discharged directly into the natural surroundings without undergoing any form of treatment. Only a small fraction of business-operated farming areas have invested in wastewater treatment systems, constituting roughly 10% of the total farming zones [2], [3], and [4]. In light of this situation, it is imperative to undertake an evaluation of the extent of pollution caused by aquaculture wastewater within the province and to devise suitable strategies to address this concern. Furthermore, in the context of integration and economic development, An Giang is facing increasing environmental pressure, particularly water pollution, which has affected the supply of fresh water for agricultural irrigation, daily life, and industrial purposes [5].

The signal of aquatic plant development, predicting the specific pollution source of water, contributes to effective and cost-efficient solutions for surface water management in An Giang province. Furthermore, the use of various plant species for water pollution treatment in aquatic ecosystems yields positive results through natural self-cleaning processes and the cooperative consumption of organic matter, nutrients, nitrogen (N), and phosphorus (P) in the water for growth, thereby purifying the water [6]. The plant biomass after harvesting can be utilized as animal feed or organic fertilizer for fields, closing the loop in the treatment process. Therefore, implementing surface water treatment technologies in natural or ecological conditions is one suitable solution in the context of climate change and the economy in An Giang province. In [7], the authors demonstrated the possibility of distinguishing macrophytes based on their ecological preferences for water nutrient concentrations, offering a potential foundation for an efficient freshwater assessment system. The Macrophyte Nutrient Index emerged as the indicator displaying the strongest correlation with diverse nutrient forms. Notably, calibrating the macrophyte method to align with local biogeographical conditions was found to significantly enhance the effectiveness of the assessment approach.

Environmental management plays a crucial role in protecting our environment and shaping strategic approaches for its preservation. In today's society, addressing issues related to the management of land, water, and air quality is an unavoidable reality. To assist decision-makers in developing effective environmental management strategies, it is essential to understand the intricacies and characteristics of environmental challenges. Spatial information and data are the fundamental building blocks of environmental management knowledge. Among the many technological solutions available, Geographic Information Systems (GIS) stands out as a leading tool for storing, analyzing, and visualizing geographically referenced information, providing a highly efficient and accurate way to tackle these challenges [8]. Currently, in the Mekong Delta region, there are some surveys on biodiversity and the distribution of various biological groups, primarily vegetation maps [9]. GIS technology has been applied to create GIS maps that depict the current state of vegetation cover in the surveyed areas. This includes background information layers in the GIS database and detailed attribute layers containing data on species composition, dominant species, and vegetation area. These maps serve as a

foundation for the management, utilization, conservation, and restoration of vegetation in the research area [10]. Therefore, the development of GIS maps and predictive models is essential to contribute to the management and treatment of clean water resources and the conservation of aquatic vegetation in the canals and waterways of An Giang province, especially in the context of climate change, using information technology databases.

Within the analysis of aquatic systems, modeling water quality parameters holds significant importance. Traditional modeling methods rely on datasets with substantial amounts of unknown or unspecified input data, often involving time-consuming processes. The integration of artificial intelligence introduces a flexible mathematical framework capable of discerning non-linear and intricate relationships between input and output data. In [11], the authors suggested different Artificial Neural Network approaches, such as Multi-Layer Perceptron Neural Network and Radial Basis Function Neural Network, to validate the efficacy of these methods in estimating water quality parameters. In [12], the authors demonstrated the use of grid search for optimizing and fine-tuning parameters in four classification models for predicting Water Quality Class and four regression models for predicting Water Quality Index. In [13], the authors leveraged machine learning algorithms to construct a model capable of forecasting both the water quality index and its corresponding quality class. Their proposed method relies on four key water parameters: temperature, pH, turbidity, and coliforms. The application of multiple regression algorithms has proven to be crucial and effective in predicting the water quality index. Additionally, the adoption of artificial neural networks represents the most efficient means for classifying water quality.

This paper presents the construction of a dataset of water surface macrophyte biological signals from the canals, rivers, and streams of An Giang province, as well as a machine learning model to predict the environmental pollution level based on the distribution and frequency of aquatic plants. Several machine learning models have developed, compared, and evaluated the predictive model against various machine learning methods, including Artificial Neural Networks (ANN) [14], k-nearest neighbors algorithm (kNN) [15], Support vector machines (SVMs) [16], Stochastic Gradient Descent [17], Random Forest [18], Linear regression, Gradient Boosting [19]. Experimental results show that the Gradient Boosting model successfully simulated the surface water pollution data, achieving the highest predictive accuracy with a coefficient of determination (R2) of 91.7% for the dependent variable MIR compared to the independent variables.

2. MATERIALS AND METHODS

2.1. Data collection

Surface water samples were collected at 18 sampling locations affected by aquaculture areas. The sample collection schedule was in March and June 2021. Fig. 1 shows the location map of surface water sampling. Table 1 shows the code, coordinates, and features of sampling locations.

https://doi.org/10.31436/iiumej.v25i2.3129

Table 1. Location of surface water sampling.

No.	Code	Coordinate	Location	Feature
1	AQ1	559.880, 1.188.560	Long Hoa commune, Phu Tan district	The end of wastewater source of the Long Hoa raft floating fish farming
2	AQ2	570.753 1.168.426	Kien An commune, Cho Moi district	The end of wastewater source of pen culture and pond fish farming area in Kien An
3	AQ3	540.104 1.184.937	Chau Phu A ward, Chau Doc city	The end of wastewater source of the Vinh Nguon raft floating fish farming
4	AQ4	540.708 1.184.988	Da Phuoc commune, An Phu district	The end of wastewater source of the Da Phuoc raft floating fish farming
5	AQ5	548.722 1.172.710	Vinh Thanh Trung commune, Chau Phu district	Impact of wastewater from fish farming ponds on Xang Vinh Tre canal
6	AQ6	567.496 1.156.042	Binh Thanh commune, Chau Thanh district	The end of wastewater source of pond fish farming area in Binh Thanh
7	AQ8	576.289 1.151.944	My Hoa Hung commune, Long Xuyen city	The end of wastewater source of pen culture and pond fish farming area in My Hoa Hung
8	AQ11	574.103 1.139.929	My Thoi ward, Long Xuyen city	Impact of wastewater from fish farming ponds
9	AQ13	552.267 1.153.651	Vinh Thanh commune, Chau Thanh district	Impact of wastewater from rice-shrimp farming area
10	AQ14	572.673 1.139.250	Phu Thuan commnue, Thoai Son district	The beginning of the Don Dong canal, adjacent to the Moi canal
11	AQ15	565.734 1.136.318	Vinh Khanh commune, Thoai Son district	The beginning of the Don Dong channel, adjacent to the Ong Co canal
12	AQ19	584.521 1.157.065	My An commune, Cho Moi district	Impact of wastewater from raft floating fish farming
13	AQ20	548.134 1.201.285	Vinh Hoa commune, Tan Chau town	Impact of wastewater from pond fish farming area
14	AQ21	546.162 1.169.627	Phu My commune, Chau Phu district	Impact of wastewater from Loc Kim Chi fish farming area, the confluence between Hao De Lon canal and Xang Vinh Tre canal
15	AQ22	548.706 1.158.636	Binh Phu commune, Chau Phu district	Impact from wastewater from Nam Viet Binh Phu fish farming area, the confluence between 13 canal and Xang Cay Duong canal
16	AQ23	552.578 1.171.504	Phu Binh commune, Phu Tan district	Impact of wastewater from Pangasius farming area in Phu Binh commune
17	AQ24	561.948 1.145.178	Vinh Trach commune, Thoai Son district	Impacts from aquatic discharge source
18	AQ25	550.885 1.177.362	Hoa Lac commune, Phu Tan district	Impact of wastewater from pond fish farming area of Hoa Lac

https://doi.org/10.31436/iiumej.v25i2.3129

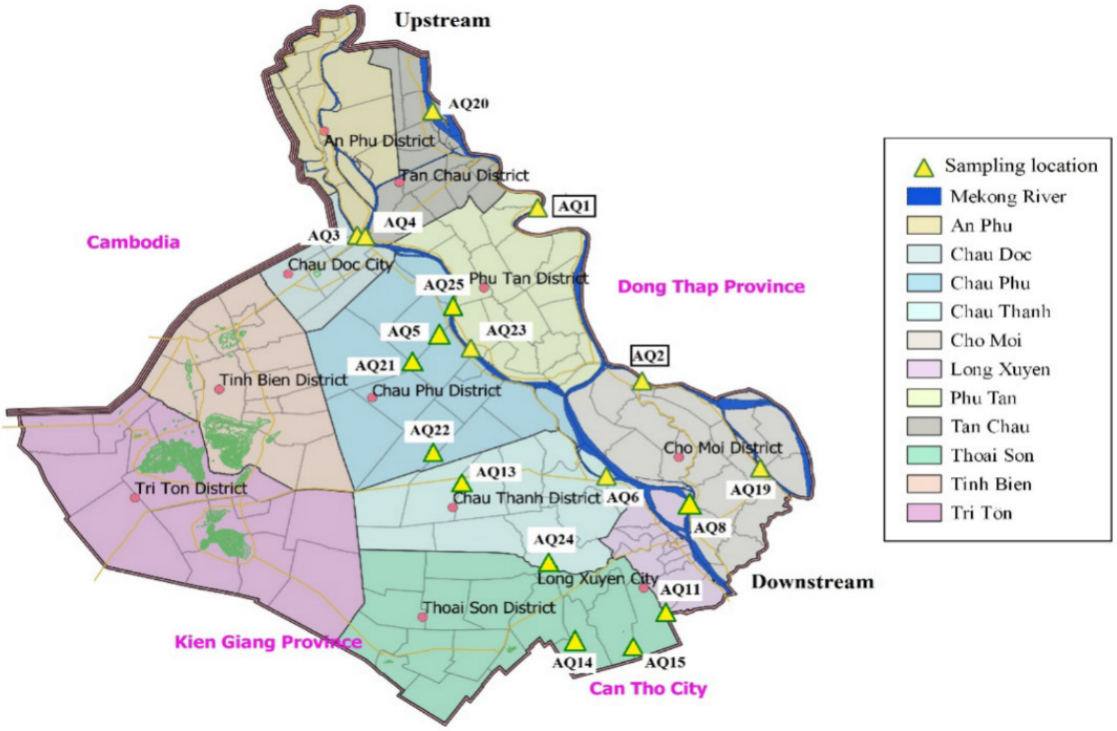


Figure 1. The location map of surface water sampling.

2.2. Methods of Sampling, Preserving, and Analyzing Water Samples

The assessment of 18 water samples involved the evaluation of various parameters, including Temperature, pH, Dissolved Oxygen (DO), Chemical oxygen demand (COD), Biological oxygen demand (BOD5), Total Suspended Solids (TSS), Ammonium (N-NH4+), Nitrate (N-NO3-), Phosphate (P-PO43-), Total Nitrogen (TN), Total Phosphorus (TP), and Coliform. The procedures for environmental monitoring followed the guidelines set forth in Circular 24/2017/TT-BTNMT, which outlines technical regulations for environmental monitoring and adhered to Vietnamese standards and regulations established by the Vietnam Ministry of Natural Resources and Environment in 2017.

To collect and preserve surface water and wastewater samples, the methods specified in TCVN 6663-1:2011 and TCVN 6663:3:2016 were employed, as outlined by the Vietnam Ministry of Science and Technology in 2011 and 2016, respectively. Sample analysis was conducted in accordance with the procedures described in Standard Methods [20] and the Vietnam National Standard on Water Quality, as specified by the Vietnam Ministry of Science and Technology in 1995, 1996, 2000, 2011, and 2016 (Table 2 and Table 3). For the measurement of temperature, DO, and pH, a PCD 650/pH 600 Eutech measuring probe was utilized.

The study assessed the plant growth performance within the wastewater treatment tank by tallying the branches of each plant during various experimental stages and gauging the length of the plant's stem both before and after a 28-day experiment. This research was conducted at two chosen agricultural households in Chau Phu district, An Giang province. The experiment was organized in the following manner:

Bamboo poles enclose the wastewater tank and feature a waterproof rubber lining, measuring 1.5 meters by 1.5 meters by 1 meter, resulting in a total water storage capacity of 1.8 cubic meters. Aquatic plants floating on the water surface were gathered from the

wastewater discharge area in accordance with the density observed in our survey. The preparation proceeded as follows:

Table 2. Method of analyzing/measuring water samples

No.	Parameter	Method of analyzing/measuring
1	Chemical oxygen demand (COD)	SMEWW 5220C:2017
2	Biological oxygen demand (BOD_5)	SMEWW 5210B:2017
3	Total Suspended Solids (TSS)	SMEWW 2540D:2017
4	Ammonium ($N\text{-}NH_4^+$)	TCVN 5988:1995
5	Total Nitrogen (TN)	TCVN 6638:2000
6	Total Phosphorus (TP)	SMEWW 4500.P.B&E:2017
7	Nitrate ($N\text{-}NO_3^-$)	SMEWW 4500-NO_3^--E:2017
8	Phosphate ($P\text{-}PO_4^{3-}$),	SMEWW 4500-P.E:2017
9	Coliform	TCVN 6187-2:1996
10	pH	TCVN 6492:2011
11	Temperature	SMEWW 2550B:2012
12	DO	TCVN 7325:2016

Water hyacinth: Select plants with a length of 20 cm and 4-5 leaves. Remove the roots, damaged stems, and leaves from the water hyacinth and place them in clean water for a period of 7 days prior to the experiment [21]. Water spinach and Climbing dayflower: Select young, healthy plants that are uniform in color and height, measuring 20 cm in height. Prior to the experiment, these plants should be washed with distilled water, as outlined in the study by [22].

The framework for the aquatic floating raft was constructed using D90mm PVC pipes, with the raft's dimensions measuring 1 meter by 1 meter. The mesh supporting the groups of aquatic plants is made of polyethylene with 2 cm-sized openings. As a result, the total area of the raft structure is 1 square meter. At each stage, wastewater was gathered from the discharge pipes of two fish ponds within the chosen farming households.

Aquatic plants were positioned on floating rafts, with a spacing of 30 cm between individual plants and 30 cm between rows, spanning across the wastewater tank. Within each floating raft, a total of 3 plant types were incorporated, resulting in 9 plant groups, each comprising 2-3 plant types. The plant density within the rafts was arranged based on the observed densities within the surveyed area. Subsequently, the rafts were secured within the wastewater tank using ropes. Fig. 2 illustrates the wastewater tank and the floating raft, while Fig. 3 provides an overview of the experimental design. Table 3 displays the water quality rating based on Water Quality Index ranges and recommendations for surface water usage.

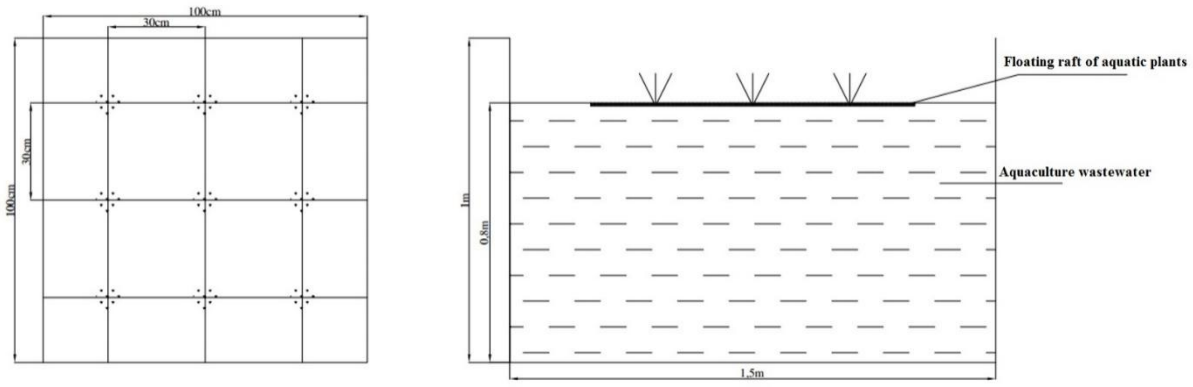


Figure 2. The wastewater tank and the floating raft of aquatic plants

Figure 3. Design of Experiment.

Table 3. Water Quality Rating and recommendation of usage

Water Quality Index Range	Water Quality Rating	Color	Intended use
91-100	Excellent		Good for water supply
76-90	Good		For water supply but requires appropriate treatment measures
51-75	Medium		For irrigation and other similar purposes
26-50	Poor		For water transport and other similar purposes
10-25	Polluted		Water is heavily polluted, requires future treatment
<10	Seriously polluted		Water is poisoned, requires treatment

Note. From "Decision 1460/QD-TCMT on promulgating technical guidelines for calculation and Vietnam water quality index (VN_WQI)" by Vietnam Environment Administration, 2019.

2.3. Macrophyte Index for Rivers

Macrophytes are widely used to assess the environmental conditions of various river ecosystems [23]. To obtain a more objective assessment of river water quality in addition to using hydrochemical indicators, the Macrophyte Index for Rivers (MIR) has been implemented [24]. The research method utilizes floating plants as a biological indicator to predict the surface water quality in An Giang province. Based on the collected macrophyte data, the Macrophyte Index for Rivers (MIR) was adapted by WFD to suit the water environment of the Mekong Delta region. In this method, the value of the MIR index depends on the presence of identified species in the river and their indicator values. According to the study conducted by the authors [7], the MIR (Macrophyte Index for Rivers) is calculated from field records using Eq. (1).

$$IR = \frac{\sum_{i=1}^{N} l_i.w_i.p_i}{\sum_{i=1}^{N} w_i.p_i} \times 10 \quad (1)$$

where MIR is the Macrophyte Index for the river at the sampling point, N is the number of species at the sampling point, l_i is the indicator value for classification unit i, w_i is the weighting factor of classification unit i, and p_i is the coverage density of classification unit i on a 9-point scale. Specifically: 1 for 0.1%, 2 for 0.1–1%, 3 for 1–2.5%, 4 for 2.5–5%, 5 for 5–10%, 6 for 10–25%, 7 for 25–50%, 8 for 50–75%, and 9 for 75%. The classification of water quality in the research areas is based on the MIR index, as studied in the paper by [24], and is presented in Table 4.

https://doi.org/10.31436/iiumej.v25i2.3129

Table 4. Water quality classification based on the MIR index.

No.	Ecological state	MIR	Water quality
1	Very good	>=44.7	I
2	Good	(44.7 – 36.5]	II
3	Medium	(36.5 – 28.2]	III
4	Bad	(28.2 – 20.0]	IV
5	Very bad	<20.0	V

2.4. Creating GIS map and database

The establishment of the current status map of aquatic plants using traditional methods has been in existence for a long time but comes with numerous limitations in implementation, requiring significant investments of time and effort in data collection and consolidation at various local levels. The lengthy process of aggregating and constructing the current status map of aquatic plants for a region leads to outdated and inaccurate information on the map. Therefore, there is a need for an alternative method to overcome the shortcomings of the traditional approach in investigating and researching the current status of aquatic plants, analyzing geographically relevant information, and meeting the urgent demands of practical production and scientific research. The choice of how to represent data on a map is a crucial step in geographic analysis. In our study, we utilized Geographic Information Systems (GIS) to create a digital map illustrating the aquatic plants found in An Giang province. GIS is a computerized database designed for managing and utilizing spatial data effectively. We carefully determined the structure of the main folders and subfolders within the GIS database. The plant data collected during field surveys was gathered using GPS and subsequently integrated into the GIS system. A thorough validation process was conducted using field forms to ensure data accuracy. We employed a georelational data model within the GIS to store attribute data alongside spatial data. Additionally, we adopted an object-based data model that consolidates spatial data with other attributes within a single system. This approach simplifies the management process by eliminating the complexity of coordinating and synchronizing two separate data sets, as is often required in a split system. The initial plant layer in our current database lacked sufficient information. To address this issue, we added columns to the attribute table of the existing plant layers to input relevant information, as described in studies by [25] and [8].

In the context of the Mekong Delta region, several studies have been conducted to survey the biodiversity and map the distribution of various biological groups, predominantly focusing on aquatic plant maps [26], [27]. GIS technology has been applied to create GIS-based maps illustrating the current status of aquatic plant coverage in areas under investigation. Within these maps, foundational information layers in the GIS database, along with detailed attribute layers containing species composition, dominant species, and area coverage of aquatic plants, serve as the basis for managing, utilizing, conserving, and restoring aquatic plant communities within the study region [10]. This paper presents the process of using remote sensing analysis to establish the current status map of aquatic plants, specifically in An Giang province. The current status map of aquatic plants in An Giang province has been developed with the following content:

- Determining the presence of floating aquatic plant families along river and stream segments flowing through An Giang province at 18 observation points:
- Designing the standard Quadrat frame and conducting practical surveys at 18 observation points.

- Determine the number of species and calculate the species density.
- Research on using floating plants as a biological indicator to predict surface water quality in An Giang province.
- Building an online GIS database and map of aquatic plant species data. Building an online aquatic GIS map system using open-source WebGIS technology. The general model of the system is presented in Fig. 5. The components of the model include:
- MySQL is a place to store spatial and non-spatial data. This data is stored in the form of tables. Therefore, this is the place to build a database of aquatic information at all survey points of the project in An Giang province.
- Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps, used to overlay maps, calculate spatial information, and receive and display information from WMS services and WebService, combining with HTML and JavaScript in designing interfaces and displaying spatial databases on the Web.
- QGIS is used to edit, edit, map, and serve information management of land plots in raw material production areas.

2.5. Method for building an online GIS database and map of aquatic plant species data

Surveying is the professional process of collecting information and data on aquatic life and water environment monitoring by research staff and specialized units at the Department of Natural Resources and Environment of An Giang Province. Research the structure of relevant forms and reports specified in documents on the water environment and aquatic life management. Object-Oriented Analysis and Design (OOAD) system analysis, from focusing on user needs expressed through building use case diagrams to database design data and processing, creating software, using, testing, and maintaining software.

2.5.1. Predicting Model Based on Machine Learning Methods

Building a predictive model for environmental pollution indicators based on the distribution and occurrence frequency of invasive alien species in An Giang. The predictive model is constructed using state-of-the-art advanced machine learning techniques such as the Adaptive Neuro-Fuzzy Inference System (ANFIS), Radial Basis Function Neural Networks (RBF-ANN), Multi-Layer Perceptron Neural Networks (MLP-ANN), and deep learning techniques.

2.5.2. Data preprocessing, removing noisy data during the data collection process

Constructing a predictive model for environmental pollution indicators is based on the best techniques and methods in the fields of machine learning and artificial intelligence, such as neural networks and deep learning. The predictive model is researched and built and can be described in a general way, as shown in Fig. 4. Evaluating the predictive model, comparing it with other traditional methods, and validating the predictive model by consulting with experts' opinions.

The predictive model for the MIR biological index has been researched and constructed and can be described in a general manner, as shown in Fig. 4. In this model, the input values consist of data on the aquatic plant density of four species: Eichornia crassipes, Polygonum hydropiper, Ipomoea aquatica, and Centrostachys aquatica, with a total of 68 samples as presented in Table 11. We have divided this dataset into two parts: one for training, comprising 80% of the samples, and the remaining 20% for testing and assessing the accuracy of the machine learning models.

https://doi.org/10.31436/iiumej.v25i2.3129

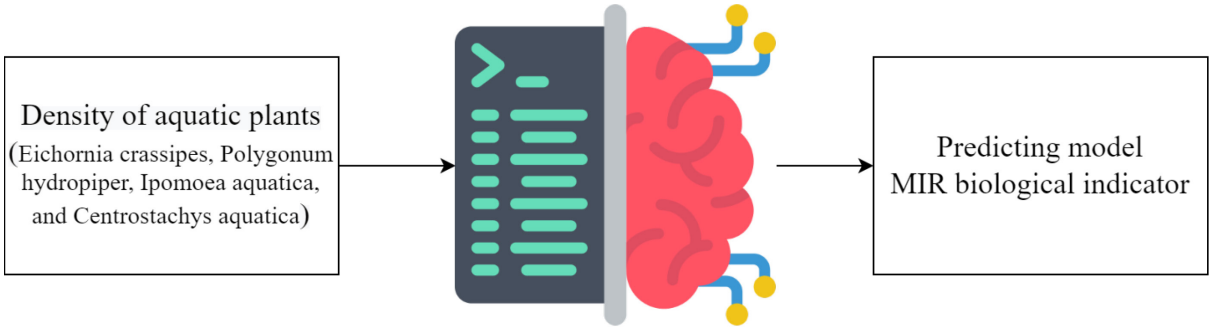
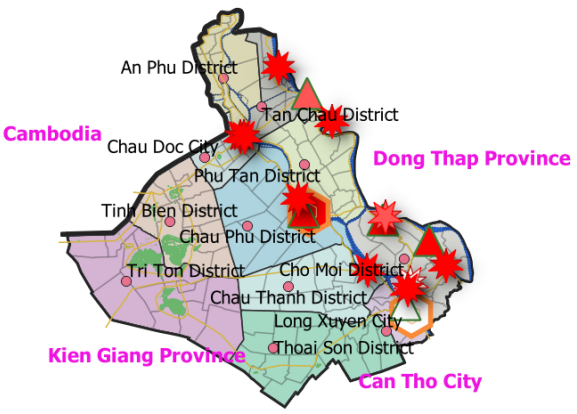


Figure 4. Machine learning model predicts MIR index

3. RESULTS AND DISCUSSION

3.1. MIR GIS Map in An Giang Province

Fig. 5 describes a GIS map constructed with information layers, including the locations of observation points, MIR index collection, and aquatic plant species. QGIS software was used to edit the map and display the layers, while VN2000 coordinates were used for map positioning. The content of both general maps and aquatic plant distribution maps consists of two fundamental parts closely related to each other: geographical foundation elements and specialized content elements. The geographical foundation elements are considered the background of the thematic map and are seen as the framework for determining the thematic elements. In the case of aquatic plant distribution maps, the geographical foundation elements serve as the basis for orientation and clarifying the characteristics and patterns of distribution of the thematic map's content. Based on available 1:50,000 topographic base map data and thematic maps at the same scale, as well as acquired remote sensing imagery, the current state map of vegetation in An Giang Province was established at a scale of 1:50,000 using the UTM WGS 84 coordinate system. The thematic content elements represent the spatial distribution of corresponding types of vegetation cover for each plot. These plots accurately depict the location, area, and shape of each vegetation cover type in proportion to the map's scale.



Feature	Value
Fishsery	
Title	
(Derived)	
(Actions)	
Code	TS1 (TĐ)-PT
x	10,747287999999999
y	105,299304000000006
Address	Long Hoa Commune, Phu Tan District
Feature	At the end of the waste stream of Long Hoa raf...
Intended_use	Monitoring the impact of aquaculture activities ...
Map_code	FS1
Eichhornia_crassipes	433
Polygonum_tomentosum	0
Ipomoea_aquatica	54
Centrostachys_aquatica	0
MIR	20

Figure 5. The map is constructed with information layers pertaining to the Fishery aquaculture region.

3.2. Survey of the Frequency of Occurrence of Aquatic Plant Species and MIR Index in An Giang

Field surveys at 18 locations along rivers and waterways within the jurisdiction of An Giang Province revealed that aquatic plant species such as water hyacinth, water spinach, morning

https://doi.org/10.31436/iiumej.v25i2.3129

glory, and water mimosa were found. However, there was no presence of lotus, red water lily, or water lily at any of the 18 survey locations. These aquatic plant species are all listed in the book by [28], and some of them were also studied by [29] in Can Tho City. It can be observed that, although there is not a great variety of species at the surveyed locations, these species are valuable for water pollution treatment, as demonstrated in numerous studies. Among them, water hyacinth is the most predominant and is present throughout the year at all survey locations. Actual photos of these aquatic plant species in the surveyed area are presented in Fig. 6.

Figure 6. Photos of aquatic plant species in the surveyed area

The area affected by urbanization: The results of the survey on the frequency of occurrence and density of aquatic plant species in the urban-affected area in 2021 are presented in Table 5. Water hyacinth was present in all seasons throughout the year. Water hyacinth predominated over other species with an occurrence frequency of over 85%. It had a 100% occurrence rate in all four seasons, with a density ranging from 10.25 to 19.3 plants/m^2 in the urban-affected areas. Water spinach was relatively abundant, with an occurrence frequency ranging from 20% to 80%. It appeared in most locations during all four seasons; some locations had it in only three or one season, with a density ranging from 0.55 to 5.45 plants/m^2. On the other hand, water mimosa were relatively scarce. Water mimosa mainly appeared in March, June, and September but was not observed in November, with an occurrence frequency ranging from 15% to 95% and a density of 0.4 to 6.8 plants/m^2. Some water mimosa appeared in only one to two seasons, mainly in September and November, with an occurrence frequency ranging from 15% to 90% and a density ranging from 0.4 to 6.1 plants/m^2.

https://doi.org/10.31436/iiumej.v25i2.3129

Table 5. The frequency of occurrence and species density in the urban-affected area in 2021.

No	Location code	Total	Frequency of appearance (%)							Density (plant/m^2)						
			Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbo naceae	Eichhornia crassipes	Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbon aceae	Eichhornia crassipes
1	MT1 (TĐ-ĐT)-TC	T3	100	25	35	-	-	-	-	14,8	0,6	3,3	-	-	-	-
		T6	100	-	60	-	-	-	-	18,45	-	5,45	-	-	-	-
		T9	100	15	80	25	-	-	-	15,4	0,4	4,0	0,6	-	-	-
		T11	100	-	80	25	-	-	-	14,2	-	4,4	0,7	-	-	-
2	MT2 (TĐ-ĐT)-CM	T3	100	-	20	-	-	-	-	17,2	-	0,65	-	-	-	-
		T6	90	-	20	-	-	-	-	14,35	-	1,3	-	-	-	-
		T9	100	-	60	45	-	-	-	12,95	-	1,15	4,75	-	-	-
		T11	100	-	20	90	-	-	-	13,8	-	0,55	6,1	-	-	-
3	MT3 (TĐ-ĐT)-CM	T3	100	-	35	-	-	-	-	19,3	-	2,9	-	-	-	-
		T6	100	-	45	-	-	-	-	17,75	-	3,55	-	-	-	-
		T9	100	40	40	-	-	-	-	14,3	1,4	0,6	-	-	-	-
		T11	90	-	-	55	-	-	-	13,4	-	-	2,35	-	-	-
4	MH3 (TĐ-ĐT)-LX	T3	100	-	55	-	-	-	-	10,25	-	2,85	-	-	-	-
		T6	90	-	30	-	-	-	-	12,45	-	1,8	-	-	-	-
		T9	100	-	35	15	-	-	-	14,6	-	1,3	0,4	-	-	-
		T11	100	-	55	-	-	-	-	15,6	-	4,05	-	-	-	-
5	MH2 (TĐ-ĐT)-CP	T3	100	75	-	-	-	-	-	10,85	6,8	-	-	-	-	-
		T6	85	95	-	-	-	-	-	15,05	6,5	-	-	-	-	-
		T9	100	75	-	-	-	-	-	14,55	4,8	-	-	-	-	-
		T11	100	-	30	80	-	-	-	14,65	-	0,8	2,55	-	-	-

The area affected by industrial zones/clusters: Table 6 presents the frequency of occurrence and species density in the area affected by industrial zones in 2021. In this table, both Water Hyacinth and Water Spinach were present in all seasons. Water Hyacinth predominated over other species, with an occurrence frequency of 80-100% in the locations, with densities ranging from 6.2 to 20.15 plants/m^2. Water Spinach was present with an occurrence frequency ranging from 15% to 100% and densities ranging from 0.5 to 8.45 plants/m^2. Water Mimosa was present at 2 locations during 3 seasons of the year, appearing at both locations in September and November, with occurrence frequencies ranging from 20% to 60% and densities of 0.7 to 4.05 plants/m^2. Nymphoides were present at one location in March and June, with an occurrence frequency ranging from 60% to 65% and densities of 3.8 to 4.05 plants/m^2, while it was not observed at the remaining locations.

Table 6. The frequency of occurrence and species density in the area affected by industrial zones in 2021.

No	Location code	Total	Frequency of appearance (%)							Density (plant/m^2)						
			Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbo naceae	Eichhornia crassipes	Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbon aceae	Eichhornia crassipes
6	MH1 (TĐ-CN)-CP	T3	100	-	40	20	-	-	-	20,15	-	2,4	0,95	-	-	-
		T6	100	-	15	-	-	-	-	24,1	-	0,5	-	-	-	-
		T9	100	-	70	30	-	-	-	15,6	-	6,05	0,7	-	-	-
		T11	100	-	80	45	-	-	-	16,95	-	5,7	3,75	-	-	-
7	MH2 (TĐ-CN)-LX	T3	100	65	55	-	-	-	-	17,8	3,8	4,25	-	-	-	-
		T6	100	60	100	60	-	-	-	6,2	4,05	8,45	4,05	-	-	-
		T9	80	-	90	50	-	-	-	8,85	-	6,6	3,8	-	-	-
		T11	90	-	80	35	-	-	-	11,85	-	7,3	2,8	-	-	-

The area affected by aquaculture: Table 7 presents the frequency of occurrence and species density in the area affected by aquaculture in 2021. Only water hyacinth was present

https://doi.org/10.31436/iiumej.v25i2.3129

in all seasons, with occurrence frequencies ranging from 35% to 100%. It had a 100% occurrence rate in all four seasons, with densities ranging from 5.7 to 28.05 plants/m^2. Water spinach was present with occurrence frequencies ranging from 10% to 100% and densities ranging from 0.4 to 9.95 plants/m^2 at locations where it was present, but it had a lower occurrence compared to water hyacinth. Only two locations had water spinach in all four seasons, and one location showed no presence. Meanwhile, Water Mimosa were present with relatively low occurrence frequencies at a few locations, with some locations showing no presence. Water mimosa appeared at only one location in three seasons, while the remaining locations mostly appeared in one to two seasons, with occurrence frequencies ranging from 10% to 90% and densities from 0.15 to 4.45 plants/m^2. Additionally, Water Mimosa was present at only 2 locations and appeared in only one season throughout the year.

Table 7. The frequency of occurrence and species density in the area affected by aquaculture in 2021

No	Location code	Total	Frequency of appearance (%)							Density (plant/m^2)						
			Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbonaceae	Eichhornia crassipes	Eichornia crassipes	Polygonum hydropiper	Ipomoea aquatica	Centrostachys aquatica	Nulumbo nucifera	Nelumbonaceae	Eichhornia crassipes
1	TS1(TĐ)-PT	T3	80	-	60	-	-	-	-	21,65	-	2,7	-	-	-	-
		T6	95	-	80	10	-	-	-	14,15	-	4,1	0,15	-	-	-
		T9	100	-	40	-	-	-	-	20,4	-	1,65	-	-	-	-
		T11	100	-	-	-	-	-	-	23,45	-	-	-	-	-	-
2	TS2(TĐ)-CM	T3	100	-	-	45	-	-	-	8,4	-	-	0,15	-	-	-
		T6	100	-	-	40	-	-	-	9,1	-	-	0,75	-	-	-
		T9	100	-	-	30	-	-	-	8,85	-	-	0,6	-	-	-
		T11	100	-	-	-	-	-	-	8,45	-	-	-	-	-	-
3	TS3(TĐ)-CĐ	T3	90	-	35	-	-	-	-	11,65	-	1,35	-	-	-	-
		T6	100	-	-	-	-	-	-	25,35	-	-	-	-	-	-
		T9	100	-	-	-	-	-	-	28,05	-	-	-	-	-	-
		T11	100	-	-	-	-	-	-	24,9	-	-	-	-	-	-
4	TS4(TĐ)-AP	T3	100	-	70	-	-	-	-	16,85	-	5,1	-	-	-	-
		T6	85	-	65	-	-	-	-	14,35	-	4,0	-	-	-	-
		T9	90	-	55	10	-	-	-	15,8	-	3,45	0,35	-	-	-
		T11	35	-	65	70	-	-	-	2,95	-	4,45	6,95	-	-	-
5	TS6(TĐ)-CT	T3	100	-	40	-	-	-	-	21,05	-	1,05	-	-	-	-
		T6	100	-	30	-	-	-	-	15,2	-	1,0	-	-	-	-
		T9	100	-	-	-	-	-	-	22,95	-	-	-	-	-	-
		T11	80	-	-	-	-	-	-	12,6	-	-	-	-	-	-
6	TS7(TĐ)-LX	T3	100	25		35	-	-	-	23,4	1,55	-	2,1	-	-	-
		T6	100	-	65	-	-	-	-	12,35	-	2,1	-	-	-	-
		T9	100	-	55	-	-	-	-	15,9	-	1,3	-	-	-	-
		T11	70	-	45	75	-	-	-	9,35	-	2,95	7,95	-	-	-
7	TS8(TĐ)-LX	T3	100	-	40	-	-	-	-	22,95	-	3,1	-	-	-	-
		T6	100	-	25	-	-	-	-	29,3	-	0,4	-	-	-	-
		T9	100	-	20	-	-	-	-	28,5	-	0,9	-	-	-	-
		T11	45	70	10	90	-	-	-	5,7	6,8	0,75	8,95	-	-	-
8	TS19(TĐ)-CM	T3	100	-	-	-	-	-	-	10,25	-	-	-	-	-	-
		T6	100	-	30	-	-	-	-	12,25	-	0,5	-	-	-	-
		T9	100	-	25	-	-	-	-	13,05	-	0,4	-	-	-	-
		T11	100	-	-	-	-	-	-	12,3	-	-	-	-	-	-
9	TS20(TĐ)-TC	T3	100	-	25	-	-	-	-	18,45	-	0,6	-	-	-	-
		T6	100	-	-	-	-	-	-	19,75	-	-	-	-	-	-
		T9	100	-	-	-	-	-	-	20,05	-	-	-	-	-	-
		T11	100	-	-	-	-	-	-	18,8	-	-	-	-	-	-
10	TS23(TĐ)-PT	T3	100	-	70	-	-	-	-	17,15	-	4,9	-	-	-	-
		T6	90	-	85	-	-	-	-	12,95	-	9,95	-	-	-	-
		T9	100	-	-	-	-	-	-	24,0	-	-	-	-	-	-
		T11	100	-	35	85	-	-	-	14,9	-	1,0	6,25	-	-	-

https://doi.org/10.31436/iiumej.v25i2.3129

The MIR index values for the points affected by urban areas, industrial zones/clusters, and aquaculture areas have been determined. The average MIR values for each season and year indicate that the water quality at the observation points is very poor. This is specifically shown in Table 8, Fig. 7 and Fig. 8. The MIR index has a correlation with the TSS, PO43-, and NH4+-N parameters at significance levels of 1%, 5%, and 5%, respectively. This finding is consistent with studies by [30] and [31], where these variables can play an important role as determinants of the presence of macrophyte species in human-impacted water sources [32].

Table 8. The average MIR values for each season and year at the affected points.

Time \ Impact point	Urban	Industry	Fisher area
1 st	19,12	18,19	19,52
2 sd	19,39	18,30	19,75
3 rd	17,97	18,36	19,71
4 th	18,31	18,09	18,84
Average year	**18,7**	**18,24**	**19,46**

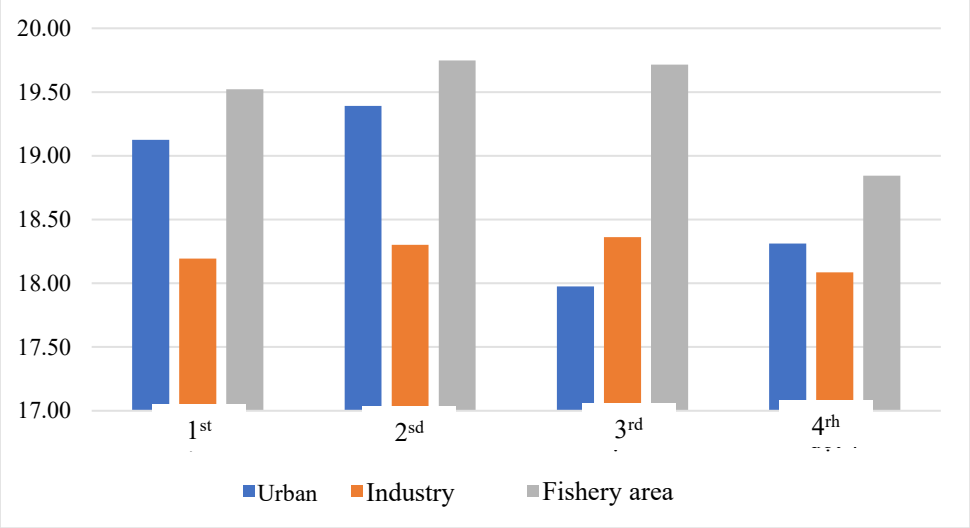


Figure 7. Average MIR value of impact points across each phase

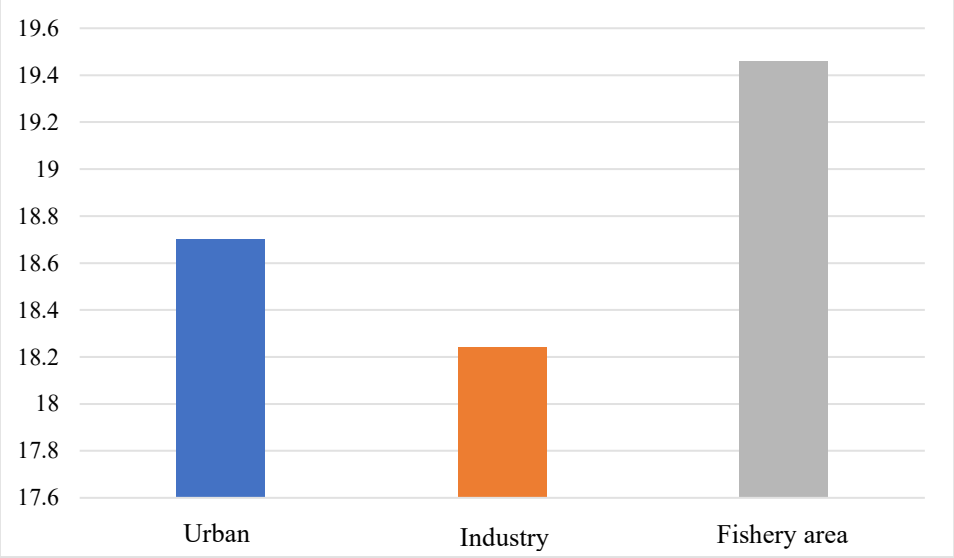


Figure 8. Average MIR value of impact points during the year

https://doi.org/10.31436/iiumej.v25i2.3129

3.3 Build a predicting model

In this study, the Orange Data Mining Toolbox [33] was used to build and evaluate predictive models for the MIR biological index, as illustrated in Fig. 9. This is an open-source platform with machine-learning algorithms and data visualization capabilities. This system can help create visual data analysis workflows with numerous tools and utilities for users and can handle large datasets.

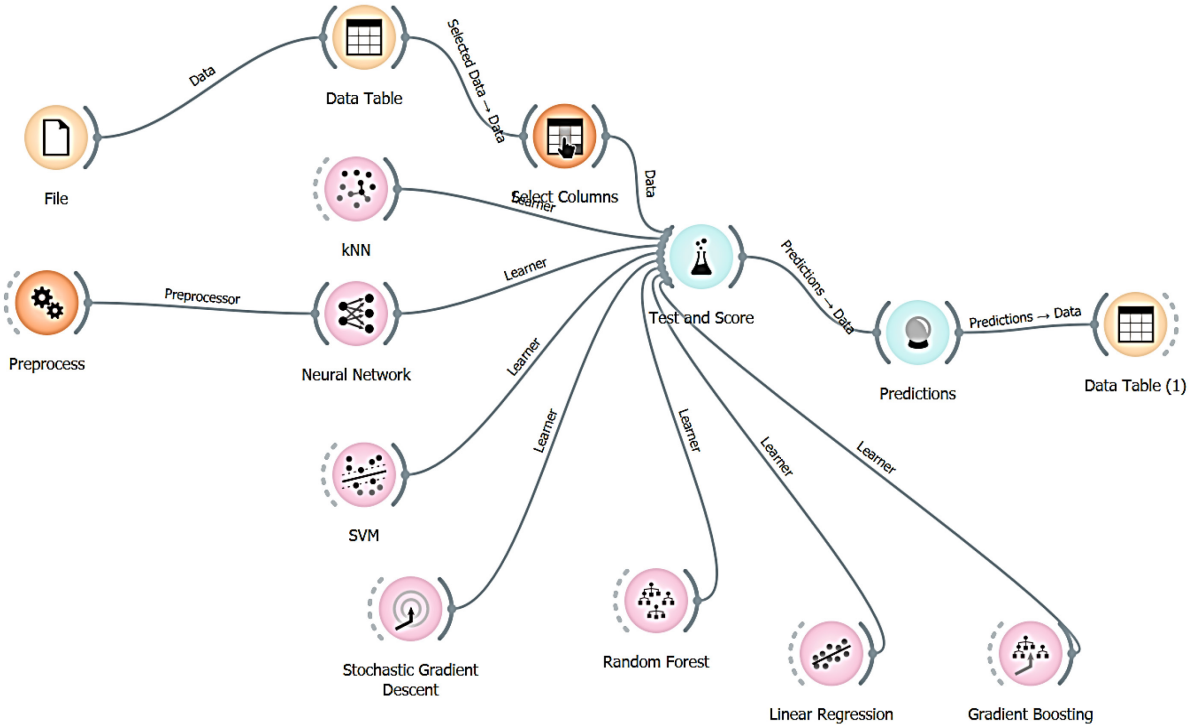


Figure 9. MIR biological indicator prediction model using Orange Data Mining Toolbox

The configuration used to train the machine learning models is as follows: 1) Sampling type: 3-fold cross-validation, 2) Data instances: 68, 3) Features: 4, 4) Meta attributes: 8, 5) Target: Numeric variable 'MIR'. The machine learning models used to predict the MIR index include kNN, Random Forest, Linear Regression, and Gradient Boosting, with the following specific configurations (Table 9):

Table 9. Specific configurations of machine learning models

Methods	Configuration
kNN	Number of neighbor = 5, Distance metric: Euclidean, Weight: Uniform
SVM	C=1, epsilon = 0.1, RBF Kernel, Numerical tolerance = 0.001, Iteration limit = 100
SGD	Classification: Hinge, Regression: Squared Loss, Regularization: Ridge (L2), Strength (a) = 0.00001, Learning rate: Constant, Initial learning rate = 0.01, Number of iterations = 1000, Tolerance = 0.001
Random Forest	Number of trees = 10, Limit depth of individual trees = 3, Do not split subsets smaller than 5
Neural Network	Neurons in hidden layers = 500, Activation = ReLu, Solver: Adam, Maximal number of iterations = 1000
Linear Regression	Fit intercept, No regularization, Alpha = 0.0001
Gradient Boosting	Number of trees = 100, Learning rate = 0.1, Limit depth of individual trees = 3, Do not split subsets smaller than 2, Fraction of training instances = 1

https://doi.org/10.31436/iiumej.v25i2.3129

To evaluate and compare the effectiveness of the machine learning models on this dataset, considering the number of species and the calculated MIR index, we used the following model evaluation metrics: MSE, RMS, MAE, and R2. MSE measures the average of the squared errors or deviations (the difference between the estimated value and the actual value). RMSE is the square root of the average of the squares of a set of numbers (a measure of the degree of imperfection of the fit of the estimator to the data). MAE is used to measure the level of closeness of a prediction to the actual result. R2 is the proportion of the variance in the dependent variable that can be predicted from the independent variable.

Table 10 shows the results of the MSE, RMSE, MAE, and R2 scores for the prediction model with Gradient Boosting, which is the best when compared to other machine learning models. This algorithm typically demands less training time and computational resources compared to deep learning algorithms, rendering them particularly suitable for smaller datasets or scenarios requiring swift model development. Additionally, gradient boosted tree models often exhibit superior accuracy compared to neural networks and greater interpretability than linear models [34]. The R2 ratio of variance in the dependent variable MIR compared to the independent variables is 91.7%. Therefore, the prediction results of this model are highly reliable and can be used to forecast the MIR biological index in practice in An Giang Province. The table provides detailed forecasting figures from the machine learning models using the independent variables: density of Eichornia crassipes, Polygonum hydropiper, Ipomoea aquatica, and Centrostachys Aquatica.

Table 10. Compare the prediction results of machine learning models on data sets at 18 monitoring points in An Giang

Model	MSE	RMSE	MAE	R2
kNN	1.105	1.051	0.752	0.210
SVM	0.572	0.756	0.515	0.590
SGD	0.505	0.711	0.571	0.638
Random Forest	0.129	0.359	0.174	0.9075
Neural Network	5.131	2.265	1.112	-2.667
Linear Regression	0.506	0.711	0.565	0.637
Gradient Boosting	0.115	0.340	0.146	0.917

Table 11 presents the predicted values from the machine learning models at the 18 observation points in An Giang Province. The results show that the machine learning models have successfully learned from the training data and can predict the MIR index. This demonstrates the effectiveness of the machine learning models when applied to real-world prediction tasks at the observation points.

The graphs of the machine learning models trained to predict the MIR biological index are presented in Fig. 10. The results show that the derived curves of the machine learning model closely match the MIR graph. This indicates that these models have effectively captured the data presented in the table with the number of species and the calculated MIR index.

https://doi.org/10.31436/iiumej.v25i2.3129

Table 11. Performance of machine learning models on monitoring dataset at 18 points in An Giang Province

No	MIR	Neural Network	SVM	SGD	kNN	Random Forest	Linear Regression	Gradient Boosting	Eichhornia crassipes	Polygonum tomentosum	Ipomoea aquatica	Centrostachys aquatica
1	18.6667	18.5922	19.4028	19.2884	18.9303	19.6408	19.2911	19.8184	296	12	66	0
2	20	20.1572	19.885	19.5363	19.7831	20	19.5525	19.9982	344	0	13	0
3	20	19.9862	19.7539	19.4665	19.4302	19.9226	19.4811	19.9997	287	0	26	0
4	17.5294	18.562	17.7207	17.411	17.8566	17.6805	17.3726	17.5879	276	0	11	122
5	20	19.7507	19.7959	19.5496	19.5758	20	19.558	19.9982	355	0	71	0
6	20	19.86	19.9293	19.6695	19.6774	20	19.6796	19.9997	453	0	57	0
7	16.9565	22.044	18.1176	17.7168	19.2414	17.5261	17.6618	17.1922	608	136	0	0
8	17.9545	15.7738	17.9873	18.6202	18.4583	17.819	18.6131	17.7938	293	0	16	51
9	16.6038	14.0625	18.2146	16.635	17.8691	17.3972	16.5506	17.2708	124	81	169	81
10	20	19.7415	19.9495	19.6451	19.6774	20	19.6556	19.9997	433	0	54	0
11	20	20.0947	19.9457	19.6892	19.2414	20	19.707	19.9988	469	0	0	0
12	20	19.7394	19.5298	19.4004	19.7831	19.9226	19.415	20	233	0	27	0
13	20	19.6406	19.4947	19.4664	18.8745	19.9226	19.4737	19.9997	287	0	80	0
14	18.9655	18.6997	19.4867	19.3847	19.1472	19.1122	19.3903	18.9256	316	0	69	7
15	20	20.0618	19.8061	19.4873	19.4302	20	19.5027	19.997	304	0	20	0
16	20	20.1023	19.9418	19.677	20	20	19.6948	19.9988	459	0	0	0
17	20	19.9572	19.5679	19.4237	19.6471	19.9226	19.4419	19.9978	252	0	0	0
18	20	19.8161	19.5493	19.4175	19.6403	19.9226	19.43	20	247	0	42	0
19	20	20.0827	19.91	19.8127	20	20	19.8279	19.9998	570	0	18	0
20	15.8	13.4451	18.2523	14.1162	17.7802	17.2127	13.9792	16.7172	114	136	15	179
21	20	19.9067	19.6363	19.4347	19.4302	19.9226	19.4518	20	261	0	8	0
22	20	20.1969	19.8939	19.5987	20	20	19.6166	19.9991	395	0	0	0
23	20	27.7457	18.4205	19.4319	18.9157	19.9226	19.4231	20.0072	259	0	199	0
24	18.5714	18.4532	18.9162	19.307	18.9593	18.3467	19.3224	17.9571	284	0	88	14
25	17.6404	15.8079	17.6843	18.4737	18.8799	17.8711	18.4274	17.5786	259	0	23	95
26	20	19.9687	19.8423	19.6184	19.6774	20	19.6462	20	421	0	58	0
27	17.5	17.6556	18.7199	19.1562	19.374	17.9602	19.1415	17.691	268	0	0	47
28	20	19.8443	19.7763	19.635	20	20	19.6546	19.9998	249	0	36	0
29	20	19.7411	19.3252	19.5146	18.9593	20	19.5401	19.9996	312	0	81	0
30	16.9565	27.6958	18.0199	15.723	18.6035	17.5188	15.6783	17.1392	301	213	0	0
31	18	13.6666	17.705	18.0742	19.1022	17.5188	18.079	17.7949	356	76	85	0
32	18.0851	17.4166	17.7899	18.3526	18.6951	17.8711	18.3248	17.838	177	0	132	76
33	18.0851	18.156	18.1337	18.5808	19.1602	17.8711	18.5701	17.9475	237	0	146	56
34	20	19.931	19.9713	19.6919	19.6774	20	19.7173	20	408	0	33	0
35	19	20.4839	19.8023	19.6842	20	19.4254	19.6959	19.286	168	0	0	3
36	18.1818	19.6098	19.6083	19.5363	20	18.3467	19.5402	17.957	182	0	0	15
37	18.1818	19.8335	19.664	19.5728	20	18.3467	19.5786	17.948	177	0	0	12
38	20	19.7877	19.1695	19.457	19.0961	20	19.4851	19.9996	337	0	102	0
39	20	19.9586	20.0059	19.7329	19.6774	20	19.7579	20	421	0	21	0
40	16.2069	12.5687	18.5693	18.6916	20	17.9249	18.6786	18.133	468	31	0	42
41	20	20.1003	19.8844	19.6866	19.5762	20	19.7081	20	318	0	26	0
42	20	20.4068	19.89	19.7333	20	20	19.7485	19.9998	205	0	0	0
43	20	20.1649	19.8888	19.7845	19.6774	20	19.8061	20	376	0	0	0
44	20	19.8155	19.2232	19.4712	19.0961	20	19.4992	19.9996	343	0	98	0
45	20	20.0201	19.9597	19.8156	20	20	19.8412	20	480	0	0	0
46	17.6404	17.0757	17.6139	18.1142	18.0142	17.6844	18.0486	17.7948	298	0	20	125
47	20	19.8254	19.6561	19.6676	19.7931	20	19.6839	20.0002	369	0	109	0
48	17.8788	17.9322	19.4881	19.3085	19.2407	18.2561	19.3161	18.3355	308	8	80	12
49	18.2353	15.7899	19.4891	19.2135	20	18.0966	19.2196	18.1419	286	28	12	0
50	18.9157	19.4316	19.7674	19.4632	19.7333	18.8093	19.4727	18.9704	292	0	26	8
51	17.2	11.5502	17.9209	18.2739	20	17.4975	18.2609	17.3525	291	96	0	0
52	18.3871	18.4138	19.4915	19.3141	20	18.2313	19.3146	18.1762	403	0	48	19
53	20	19.8625	19.938	19.6684	19.2414	20	19.675	20.0002	482	0	10	0
54	18.6364	18.6427	19.3101	19.4005	19.7143	18.2561	19.4114	18.5599	312	0	121	14
55	18.0851	15.7789	18.0273	18.3117	19.5074	17.9114	18.2857	17.9832	339	0	114	75
56	19.2857	19.6641	19.7722	19.5722	19.2407	19.2409	19.5882	19.077	283	0	82	3
57	20	19.6687	19.1712	19.5526	19.0727	20	19.5695	19.9221	169	0	0	0
58	20	20.1972	19.8288	19.6733	19.2414	20	19.6785	19.9993	507	0	0	0
59	20	20.9454	19.6528	19.6926	19.2414	20	19.6959	19.9993	561	0	0	0
60	20	20.1372	19.8491	19.6701	19.2414	20	19.6756	19.9993	498	0	0	0
61	17.4074	22.3486	17.9996	17.0522	17.5341	17.7819	16.9971	17.8931	59	0	89	139
62	17.5862	21.5141	17.9737	16.7268	17.7961	17.7819	16.6538	17.8908	187	0	59	159
63	20	19.9472	19.8507	19.6809	20	20	19.6914	20.0002	459	0	62	0
64	20	21.0837	19.5714	19.7048	19.2414	20	19.7076	20.0002	586	0	8	0
65	20	19.8574	19.7528	19.5837	20	20	19.5986	19.9998	245	0	10	0
66	20	19.869	19.6595	19.5801	20	20	19.5943	19.9989	246	0	0	0
67	20	20.0412	20.0446	19.6288	20	20	19.6395	20.0002	369	0	12	0
68	20	20.1205	19.9502	19.6355	20	20	19.6443	19.9993	401	0	0	0

https://doi.org/10.31436/iiumej.v25i2.3129

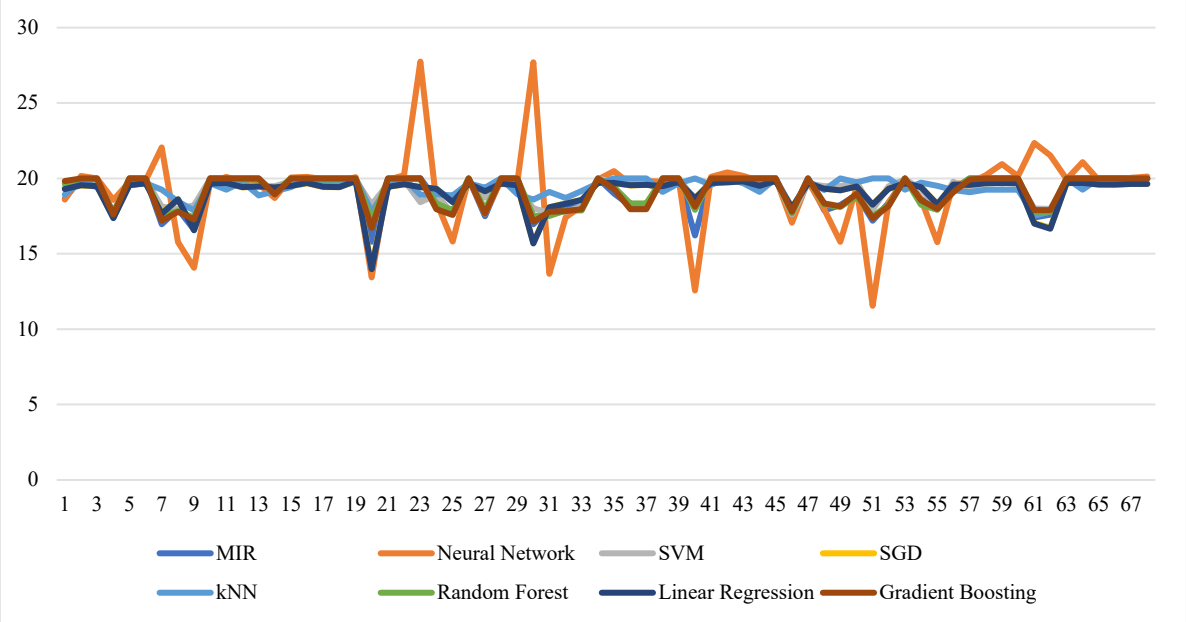


Fig. 10. Graph of MIR index prediction results of machine learning models

Fig. 11 presents a visual map of the predicted biological index at 18 observation points in An Giang Province. This helps users easily observe, compare, evaluate, and suggest appropriate solutions. The star symbols indicate the locations of the MIR pollution index, with darker red colors indicating higher levels of pollution, corresponding to lower water quality.

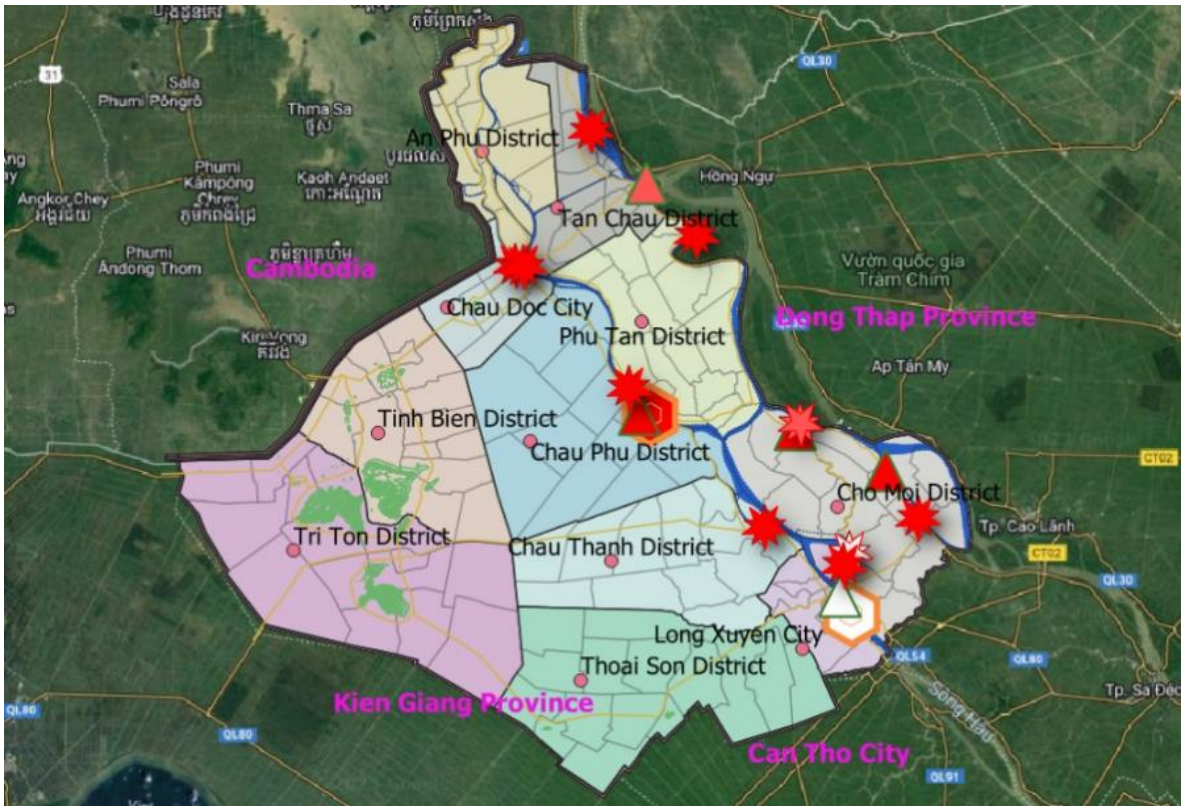


Figure 11. Map showing the results of MIR biological indicator forecasting in An Giang province.

4. CONCLUSION

This research collected and established a surface water quality monitoring database at 18 observation points. The results of surface water quality monitoring affected by urban areas, industrial clusters, and aquaculture areas in 2021 showed that 7 out of 14 parameters exceeded the allowable standards, including DO, TSS, COD, BOD5, PO43--P, NH4+-N, and Coliform; while the remaining parameters were within limits specified in the QCVN 08-MT:2015/BTNMT, column A1 - National technical standards for surface water quality. Four floating aquatic plant species were identified: Water Hyacinth, Water Mimosa, Water Spinach, and Water Mimosa. There was no presence of lotus, water lily, or water hyacinth. Among them, water hyacinth was the most dominant, present at all times of the year at all observation locations, while water spinach and water mimosa appeared relatively infrequently and mainly in scattered locations. The MIR index at points affected by urban areas, industrial clusters, and aquaculture areas was determined, and the average MIR values for each period and year indicated that water quality at the monitoring points was very poor. The results also showed that the calculated MIR index had a significant correlation with TSS, PO43-, and NH4+-N at significance levels of 1%, 5%, and 5%, respectively. As a result, this study constructed a GIS map of surface water monitoring, the number of plant species, the MIR index, and a prediction model for the MIR biological index with an accuracy of over 91.7%. This helps users easily observe, compare, evaluate, predict, and propose suitable solutions for managing the quality of surface water in An Giang Province.

ACKNOWLEDGEMENT

The authors would like to thank An Giang University and Vietnam National University in Ho Chi Minh City, Vietnam.

REFERENCES

[1] An Giang Center for Environmental Monitoring and Techniques Resources. (2019) Report on results of environmental monitoring in An Giang province November, 2019, People's Committee of An Giang province

[2] Nguyen, K.; Yoshiaki, K.; Xiao, L.; Endo, R.; Shibuya, T. (2015) Microalgae Culture with Digestate from Methane Fermentation, *Eco-Engineering*, Vol. 27, No. 1, 7–11. doi:10.11450/seitaikogaku.27.7

[3] People's Committee of An Giang province. (2019). An Giang Department of Natural Resources and Environment. (2020). Report on environmental status of An Giang province in the period 2016-2020.

[4] Nguyen, V. P.; Nguyen, K. T. T.; Ton, L. T.; Nguyen, D. T.; Nguyen, K. Q.; Vu, M. T.; Tran, H. N. (2020) Dual-Electronic Nanomaterial (Synthetic Clay) for Effective Removal of Toxic Cationic and Oxyanionic Metal Ions from Water, *Journal of Nanomaterials*, Vol. 2020, 1783749. doi:10.1155/2020/1783749

[5] An Giang province monitoring center. (2019) Report on environmental monitoring results of An Giang province

[6] Demars, B. O. L.; Edwards, A. C. (2009) Distribution of aquatic macrophytes in contrasting river systems: A critique of compositional-based assessment of water quality, *Science of The Total Environment*, Vol. 407, No. 2, 975–990. doi:https://doi.org/10.1016/j.scitotenv.2008.09.012

[7] Szoszkiewicz, K.; Jusik, S.; Pietruczuk, K.; Gebler, D. (2020) The macrophyte index for rivers (MIR) as an advantageous approach to running water assessment in local geographical conditions, *Water (Switzerland)*, Vol. 12, No. 1. doi:10.3390/w12010108

[8] Le Hoang Tu, Phan Thi Ha, Vo Ngoc Quynh Tram, Nguyen Ngoc Thuy, Dang Nguyen Dong Phuong, Tran Thong Nhat, N. K. L. (2023) GIS Application in Environmental Management: A Review, *VNU Journal of Science: Earth and Environmental Sciences,* Vol. 39, No. 2

[9] Tran Triet, N. P. B. (2002) *Correlation between hydrophytes and water environment at Tram Chim National Park, Dong Thap*

[10] Dũng, P.; Tin, H.; Pháp, T. (2013) Species composition and distribution of mangrove plants in lap an swamp, phu loc district, Thua Thien Hue province, *Hue University Journal of Science: Agriculture and Rural Development*, Vol. 75. doi:10.26459/jard.v75i6.3157

[11] Najah Ahmed, A.; Binti Othman, F.; Abdulmohsin Afan, H.; Khaleel Ibrahim, R.; Ming Fai, C.; Shabbir Hossain, M.; Ehteram, M.; Elshafie, A. (2019) Machine learning methods for better water quality prediction, *Journal of Hydrology*, Vol. 578, 124084. doi:https://doi.org/10.1016/j.jhydrol.2019.124084

[12] Shams, M. Y.; Elshewey, A. M.; El-kenawy, E.-S. M.; Ibrahim, A.; Talaat, F. M.; Tarek, Z. (2023). Water quality prediction using machine learning models based on grid search method, *Multimedia Tools and Applications*. doi:10.1007/s11042-023-16737-4

[13] Azrour, M.; Mabrouki, J.; Fattah, G.; Guezzaz, A.; Aziz, F. (2022) Machine learning algorithms for efficient water quality prediction, *Modeling Earth Systems and Environment*, Vol. 8, 1–9. doi:10.1007/s40808-021-01266-6

[14] McCulloch, W. S.; Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, 115–133. doi:https://doi.org/10.1007/BF02478259

[15] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression, *The American Statistician*, Vol. 46, No. 3, 175–185

[16] Cortes, C.; Vapnik, V. (1995) Support-vector networks, *Machine Learning*, Vol. 20, No. 3, 273–297

[17] Ruder, S. (2016) An overview of gradient descent optimization algorithms, *ArXiv Preprint ArXiv:1609.04747*

[18] Ho, T. K. (1995) Random decision forests, *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1), 278–282

[19] Friedman, J. H. (2001) Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 1189–1232

[20] Lipps, W. C., Baxter, T. E., & Braun-Howland, E. (n.d.). Standard methods for the examination of water and wastewater, *Standard Methods For the Examination of Water and Wastewater*. doi:10.2105/SMWW.2882.030

[21] Nguyen, L.; Linh, N.; Thịnh, Đ.; Hang, P.; Nguyen Vo Chau, N.; Thu, V. (2015) Evaluation of treatment efficiency of domestic wastewater by aquatic plants, *Can Tho University Journal of Science*

[22] Khôi, C. M.; Dũng, N. V. C.; Nhiên, C. T. (2012) Ability to treat dissolved nitrogen and organic phosphorus pollution in pangasius pond wastewater by Luc Binh (Eichhorinna crassipes) and Vertiver grass (Vertiver zizanioides), *Can Tho University Journal of Science*, 151–160

[23] Ceschin, S.; Ferrante, G.; Mariani, F.; Traversetti, L.; Ellwood, N. T. W. (2020) Habitat change and alteration of plant and invertebrate communities in waterbodies dominated by the invasive alien macrophyte Lemna minuta Kunth, *Biological Invasions*, Vol. 22, No. 4, 1325–1337. doi:10.1007/s10530-019-02185-5

[24] Nekos, Alla and BoiarynMariia and Lugowska, Maria and Tsos, Oksana and Netrobchuk, I. (2021) Assessment of the ecological condition of the Western Bug river basin according to the macrophyte index for rivers (MIR), *Visnyk of V.N. Karazin Kharkiv National University, Series Geology. Geography. Ecology*

[25] Ismail, M. H. (2010) Determining and Mapping of Vegetation using GIS and Phytosociological Approach in Mount Tahan, Malaysia, *Journal of Agricultural Science*, Vol. 2, No. 2. doi:10.5539/jas.v2n2p80

[26] Nguyen Quang Tuan; Trần Văn No; Do Thi Viet Huong. (2010) Applying GIS and remote sensing in creating a map of the current state of vegetation in 2008 at a scale of 1:50,000 in Ky Anh district, Ha Tinh province, *Hue University Journal of Science*

[27] Bareth, G.; Waldhoff, G. (2017) GIS for Mapping Vegetation, *Reference Module in Earth Systems and Environmental Sciences*. doi:10.1016/B978-0-12-409548-9.09636-6

[28] Đỗ Huy Bích; Đặng Quang Chung; Bùi Xuân Chương; Nguyễn Thượng Dong; Đỗ Trung Đàm; Phạm Ngọc Hiển; Vũ Ngọc Lộ; Phạm Duy Mai; Phạm Kim Mãn; Đoàn Thị Nhu; Nguyễn Tập; Trần Toàn. (2006) *Medicinal Plants and Medicinal Animals in Vietnam (Volume 1)*, Nhà xuất bản Khoa học và Kỹ thuật, Hà Nội

[29] Đan, T. H.; Duy, N. P.; Thọ, B. T. (2012) The distribution of aquatic plants in the organic polluted canals in Can Tho City in rainy season, *Tạp Chí Khoa Học Đại Học Cần Thơ*, No. 23a, 283–293

[30] Bytyçi, P.; Shala-Abazi, A.; Zhushi-Etemi, F.; Bonifazi, G.; Hyseni-Spahiu, M.; Fetoshi, O.; Çadraku, H.; Feka, F.; Millaku, F. (2022) The Macrophyte Indices for Rivers to Assess the Ecological Conditions in the Klina River in the Republic of Kosovo, *Plants*. doi:10.3390/plants11111469

[31] Tarkowska-Kukuryk, M.; Grzywna, A. (2022) Macrophyte communities as indicators of the ecological status of drainage canals and regulated rivers (Eastern Poland), *Environmental Monitoring and Assessment*, Vol. 194, No. 3, 210. doi:10.1007/s10661-022-09777-0

[32] Bytyçi, P.; Shala-Abazi, A.; Hyseni Spahiu, M.; Feka, F.; Fetoshi, O.; Bonifazi, G.; Millaku, F. (2021) *The Macrophyte Index for Rivers (MIR) to Running Water Assessment in River Klina to the Republic of Kosovo*

[33] Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; Štajdohar, M.; Umek, L.; Žagar, L.; Žbontar, J.; Žitnik, M.; Zupan, B. (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research*, Vol. 14, 2349–2353

[34] Lundberg, S.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. (2020) From Local Explanations to Global Understanding with Explainable AI for Trees, *Nature Machine Intelligence*, Vol. 2. doi:10.1038/s42256-019-0138-9

https://doi.org/10.31436/iiumej. v25i2.3279

AREA-BASED RAINFALL RATE MODEL FOR SPECIFIC ATTENUATION IN THE EQUATORIAL REGION

YASSER ASRUL AHMAD[1*], AHMAD FADZIL ISMAIL[1], MUHAMMAD NASRIN AQIL ABDUL HAMID[1], MOHD FAIZAL JAMLOS[2,3]

[1]Dept. of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University of Malaysia, Kuala Lumpur, Malaysia

[2] Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia

[3] Centre for Automotive Engineering, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia

**Corresponding author: yasser@iium.edu.my*

(Received: 30 April 2024; Accepted: 20 May 2024; Published online: 15 July 2024)

ABSTRACT: The advent of new telecommunication systems with large bandwidths like 5G/6G and satellites operating at higher frequency bands such as Ku, Ka-band, and even Q/V bands, has brought to the forefront the issue of rain attenuation, particularly in equatorial and tropical regions. Many existing models developed to address these inaccuracies in the rainfall rate prediction rely on a single rain gauge measurement, which still can lead to inaccuracy when the model is generalized for a larger area. This research, therefore, is of utmost importance as it aims to develop an area-based rainfall rate model using multiple rain gauges spread across many locations in a specific area. The area of focus for this research is the Klang Valley area, a crucial economic territory that includes the capital city, Kuala Lumpur, in Malaysia. Five rain gauges distributed in Klang Valley were chosen to measure the rainfall rate. The rainfall rate model was then developed based on the data from these five rain gauges. The results indicate that each rain gauge's rainfall rate at $R_{0.01\%}$ exceedance level varied greatly from 102 mm/hr to 138 mm/hr and exceeded Malaysia's recommended ITU-R at 100 mm/hr. The new model, presented herein, accounts for the variation of rainfall rate across a larger area, which can provide accurate modeling of specific attenuation and rain attenuation in the equatorial regions, thereby enhancing the reliability of communication systems.

ABSTRAK: Kemunculan sistem telekomunikasi jalur lebar baru seperti 5G/6G dan satelit yang beroperasi pada jalur frekuensi tinggi seperti Ku, Ka-band, dan jalur Q/V, menyebabkan penurunan ketara amplitud gelombang akibat hujan, terutamanya di kawasan khatulistiwa dan tropika. Kebanyakan model sedia ada yang dibangunkan bagi menangani ketidak tepatan ramalan kadar hujan hanya berdasarkan pengukuran dari satu tolok hujan, ini menyebabkan ketidaktepatan model itu apabila digeneralisasikan bagi kawasan lebih luas. Oleh itu, kajian ini adalah amat penting kerana ia bertujuan bagi membangunkan model kadar hujan berasaskan kawasan menggunakan beberapa tolok hujan yang tersebar di pelbagai lokasi pada satu kawasan tertentu. Kawasan tumpuan kajian ini adalah kawasan Lembah Klang, sebuah wilayah ekonomi penting yang melingkupi ibu negara, Kuala Lumpur, Malaysia. Lima tolok hujan yang tersebar di Lembah Klang dipilih bagi mengukur kadar hujan. Model kadar hujan kemudiannya dibangunkan berdasarkan data dari kelima-lima tolok hujan ini. Dapatan menunjukkan kadar hujan $R_{0.01\%}$ bagi lima tolok hujan adalah antara 102 mm/jam hingga 138 mm/jam dan melebihi kadar cadangan ITU-R Malaysia iaitu pada 100 mm/jam. Model baru yang mengambil kira variasi kadar hujan pada suatu kawasan lebih luas menggunakan banyak tolok hujan, di mana dapat memberikan pemodelan amplitud gelombang yang lebih tepat bagi

hujan di kawasan khatulistiwa, dengan itu meningkatkan kebolehpercayaan sistem komunikasi.

KEYWORDS: *Rain gauge network, Areal rainfall analysis, Equatorial climate dynamics, Rainfall rate modeling, and Satellite communication attenuation.*

1. INTRODUCTION

New telecommunication systems such as 5G, 6G, and high throughput satellites (HTS) employ a higher frequency band such as Ku, Ka, and above. These microwave and millimeter frequency bands are well known to be highly susceptible to rain attenuation, which is a greater issue in the rainy tropical and equatorial regions. Not much research distinguishes the tropical region from the equatorial region, which can significantly affect the rain attenuation model. In a tropical region, the climate typically exhibits two distinct seasons: a dry season and a wet season. During the dry season, the average rainfall is approximately 50 mm/hr, whereas the wet season experiences significantly higher rainfall rates [1,2]. The equatorial region has a hot and wet season throughout the year with an average rainfall value of at least 60 mm/hr, with heavy rainfall usually occurring in the afternoon throughout the year [1,2]. Thus, the equatorial region also has heavy rainfall but with a well-distributed rainfall rate throughout the year. The rain formation and raindrop sizes exhibit similar features for the equatorial and tropical regions. As a comparison, temperate regions have a smaller raindrop size and a lower rainfall rate of around 25 mm/hr for an annual average with an interval not exceeding one hour [3].

The rainfall rate in equatorial regions greatly varies according to different geographical areas [4]. Therefore, understanding the characteristics of an area-based rainfall rate is imperative and should be based on local areas and climatic conditions. Malaysia is unique as it experiences an equatorial climate with tropical monsoons [5].

Malaysia has a very high rainfall rate that can exceed 200 mm/hr with an equatorial climate mixed with tropical monsoon [6]. The seasons in Malaysia are generally divided into two main seasons: heavy rainfall rate is typically distributed throughout the twelve months of the year, and heavy rainfall rate occurs in the afternoon [7]. Malaysia faces two monsoon seasons, namely, the Southwest Monsoon, which occurs from late May to September, and the Northeast Monsoon, occurring from November to March. During the monsoon regime in the East peninsular, heavy showers are expected to pour for an hour or so. Rain patterns are usually consistent and can be predicted for about a week before changing. In the Western peninsular, nonstop rain could occur for weeks at a time [8].

From the point of view of electromagnetic wave propagation, precipitation can be divided into two types, specifically stratiform rain and convective rain [9]. Stratiform rain is a product of stratus clouds. The stratiform rain features light rainfall intensity and limited vertical extent of rain height, spread over a large area. Whereas the cumulus clouds produce convective rain with high rainfall rates and extensive rain height that covers a smaller area [9]. Convective rain is the major cause of communication link outages in the equatorial region and Southeast Asia region [4].

The stratiform and convective rain can be discriminated or differentiated based on the rainfall rate. In the Equatorial region, stratiform rain has a rainfall rate of less than 10 mm/hr while convective rain can be considered above 10 mm/hr [10]. Rainfall rate can be obtained from the rainfall rate distribution model or direct measurement using an instrument such as a rain gauge. One of the major issues in predicting the rainfall rate in the tropical region is that most of the rain models are developed in the temperate climate [10].

https://doi.org/10.31436/iiumej. v25i2.3279

The rainfall rate model is the major parameter in determining specific attenuation based on the power law relationship [11]. The ITU-R P.837-7 [12] provides a global rainfall rate model for exhibiting the propagation of electromagnetic waves under the condition of precipitation. However, the ITU-R rainfall rate model developed in temperate regions tends to underestimate the rainfall rate in other areas with heavier rainfall [13]. The ITU-R P.837-7 and the Singh & Acharya model focus on predicting the rainfall rate at $R_{0.01\%}$ to be suitable with the ITU-R 618 rainfall rate prediction model. However, rainfall rate distribution <1% is more critical due to the variation in rainfall rate in the equatorial climate [14]. Intriguingly, there is a developed model based on the local area, the Moupfouma model for Southern India [15].

The most common and direct way to measure rainfall rate is to use a rain gauge. Rain gauge measures the point rainfall rate [16]. Classic rain gauges may have an issue in terms of accuracy, but over the years, rain gauge design has seen improvement in terms of accuracy [17]. A single rain gauge may not give accuracy over a large area. Thus, hydrologists have resorted to a network of rain gauges to improve area average rainfall estimates [18].

In analyzing communication links, the rainfall rate is analyzed to produce a cumulative distribution based on 1-minute rainfall rate in accordance with the standard set by ITU-R [12]. Recent measurement [7] shows a lower rainfall rate than the older measurement [6, 19, 20]. Thus, this research aims to achieve the following objectives: to revise, measure, and analyze the area-based rainfall rate exceedance in the Klang Valley, Malaysia, and to develop and propose a more accurate area-based rainfall rate model based on multiple rain gauges for the equatorial region.

2. METHODOLOGY

2.1 Rain Gauge Measurement Set Up

In Malaysia, the Department of Irrigation and Drainage (DID) owns and operates a meteorological standard hydrological station network all over the country, which accurately provides tropical rainfall rate information. Thus, rain data could be obtained from the DID rain gauge. Fig. 1 shows a DID rain gauge.

Figure 1. DID Rain Gauge

The DID rain gauge is a remote-sensing tipping twin bucket type, which is able to measure the volume of rainfall continuously, accurately, and reliably. It is a meteorological standard rain gauge with a portable, rugged, lightweight unit that is able to continuously measure the volume of rainfall using a twin bucket, hence providing highly accurate and reliable

measurement. Each tip of the bucket can measure 0.5 mm of rainfall. The accuracy of the rainfall intensity is +/- 1% at 250 mm/hr. The diameter of the funnel is 200 mm with a sharply beveled edge in conformance with best meteorological practice. The funnel has a 0.09 m height to reduce sampling loss due to splashing. The data is recorded on an electromagnetic strip chart recorder in less than 120 msec. The data is recorded in each one-minute interval and sent via modem to the internet. Then, it is made available at the DID portal.

The specification of the tipping bucket rain gauge used by DID is listed below[21].

- The rain gauge shall be a remote-sensing tipping bucket type. It shall be a rugged, lightweight and portable unit, which can be used to measure continuously volume of rainfall with high accuracy and reliability.
- The `unit' of measurement shall be in the metric system, with an accuracy of ± 1% over a rainfall intensity of 250mm/hr and an accuracy of ± 3% to 300mm/hr. The rain gauge shall be calibrated against a rainfall intensity of 250mm/hr to 300mm/hr.
- The rain gauge collecting funnel shall have a sharply beveled edge that conforms with the best meteorological practice. The diameter of the funnel shall be 200mm ± 0.2mm.
- The funnel wall shall be a minimum of 0.09m high to reduce the sampling loss due to splashing.
- The funnel shall have a large round sieve to prevent falling leaves, debris, etc., from entering the measuring element of the rain gauge.
- The funnel shall have a siphon mechanism for the rainwater to constantly flow into the measuring element.
- The secondary funnel shall be provided to feed the rainwater to the measuring element. Its function is to stabilize the feed water and to allow dust to settle.
- The measuring element of the rain gauge shall be the twin bucket type. It shall be pivoted in the centre and so balanced that one of the buckets shall always be under the funnel, ready to be filled with water. When a given amount of water is collected the bucket shall tip and bring the second bucket into position under the funnel.
- The bucket and movable part shall be made of a metal base material (plastic base material shall not be accepted).
- Each tip of the bucket shall represent a measure of 0.5mm of rainfall, and every tip of the bucket shall also cause a make-and-break closure of two reeds or switches at the same time with a two-output terminal.
- Enough closure time of the switches shall be allowed for data to be recorded on an electromagnetic strip chart recorder, which should not be longer than 120 ms.
- The measuring element shall have two drainage openings, one for each bucket. The openings shall be covered with screens to prevent insects from entering the measuring element. The two openings shall have a facility to connect 4 to 10-mm diameter rubber tubing.
- All components of the rain gauge shall be made from corrosion-resistant materials.

2.2. Area-based Rainfall Rate

The rainfall rate measured by a single rain gauge will provide a point rainfall rate associated with the rain attenuation measurement. To obtain an area-based rainfall rate model in the surrounding area, multiple rainfall rate measurements at different places should be made. The surrounding area of this experiment is the Klang Valley in Malaysia. Inside the Klang Valley is Kuala Lumpur, the main capital of Malaysia, with surrounding urban areas from the state of Selangor adjoining Kuala Lumpur. As the most urban and important economic area in

https://doi.org/10.31436/iiumej. v25i2.3279

Malaysia, a reliable communication link is critical in the Klang Valley. The areal rainfall rate information is critical to satisfy the specific attenuation accurately.

Fig. 2 shows a typical DID hydrological station. The DID rain gauge is used to obtain the rainfall data. The DID rain gauge is part of the DID Hydrological station network. DID has a network of Hydrological stations across Malaysia. DID Selangor's Hydrological Network includes Rain gauges, Water Level Stations, Evaporation stations, Suspended Sediment Stations, Stream of Flow Stations, and Water Quality Stations. The specification of the rain gauge is the same as the rain gauge described in the previous section. The hydrological station's database can be found at HydroNET [22]. However, not all DID rain gauge stations provide measurements at 1-minute intervals. Thus, only 1-minute interval stations are selected from the HydroNET database. Table 1 lists the selected rainfall stations from HydroNET for the Klang Valley area. The selected stations record rain data for every 1-minute interval. All the rain data will be used to obtain the rainfall rate. 1-minute interval rainfall data from all stations from Jan 2015 to Dec 2016 will be used based on the existing database. The average rainfall rate data from the five stations will be obtained, and the rainfall rate will be modeled from the average data to give the rainfall rate for the Klang Valley area.

Figure 2. DID Hydrological Station

Table 1. Rainfall Stations in Klang Valley Area

Station No	Station Name	Latitude [N]	Longitude [E]	Area
2816139	Payah Indah	02 52 44	101 37 09	Cyberjaya
2916001	Puncak Niaga	02 54 40	101 41 56	Putrajaya
2917002	Sungai Merab	02 56 43	101 44 60	Bangi
3211001	Taman Templer	03 17 49	101 37 58	Gombak
3214057	Bdr. Tasik Puteri	03 17 59	101 29 18	Rawang

https://doi.org/10.31436/iiumej. v25i2.3279

The rainfall data collected by the rain gauge is measured in mm thus it has to be converted to mm/hr for an actual rainfall rate. [23] suggested that the rainfall rate can be determined using Eq. **Error! Reference source not found.**1), where L is the maximum rainfall (mm) and T is the time interval (min).

$$R(mm/h) = L * \frac{60}{T} \quad (1)$$

Table 2 shows a sample of the rainfall rate converted from the "rain rainfall" to "rainfall rate." It shows that the increment of 0.1 mm is equivalent to an increment of 6 mm/hr. The effect of the conversion will cause the cumulative distribution function graph to be staggered.

3. RESULT

3.1. Rainfall Measurement and Analysis

The rainfall measurement presented in this section is based on the Paya Indah rain gauge data. The Paya Indah station is used as a benchmark because it is near the Earth stations located in Cyberjaya Malaysia. Similar analyses were done for the other four rain gauge stations presented in Table 1, which will be plotted together with the developed model in Fig. 9. The rain gauge measures rainfall in mm for every minute. This section presents the results and findings for the monthly rainfall volume (mm), monthly rainfall rate (mm/hr) and annual rainfall rate (mm/hr) for the Paya Indah Station. Data availability for rainfall during the two years is 100%.

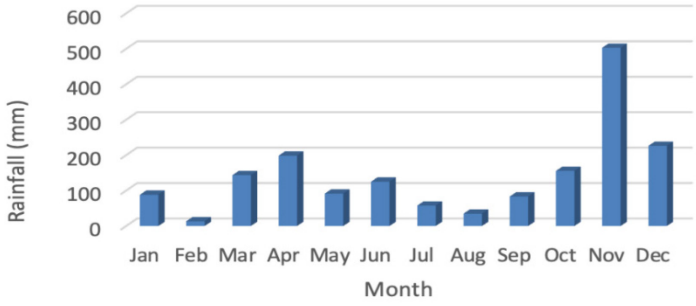


Figure 3. Monthly rainfall volume in 2015 for Paya Indah rain gauge

Table 2: Rainfall (mm) converted to rainfall rate (mm/hr)

Date	Time	Rainfall rate [mm/hr]	Rainfall [mm]
13/12/15	13:30:00	6	0.1
13/12/15	13:31:00	6	0.1
13/12/15	13:32:00	6	0.1
13/12/15	13:33:00	6	0.1
13/12/15	13:34:00	6	0.1
13/12/15	13:35:00	6	0.1
13/12/15	13:36:00	6	0.1
13/12/15	13:37:00	6	0.1
13/12/15	13:38:00	6	0.1
13/12/15	13:39:00	12	0.2
13/12/15	13:40:00	12	0.2
13/12/15	13:41:00	12	0.2
13/12/15	13:42:00	24	0.4
13/12/15	13:43:00	48	0.8
13/12/15	13:44:00	54	0.9
13/12/15	13:45:00	48	0.8

https://doi.org/10.31436/iiumej. v25i2.3279

3.2 Equatorial Rainfall Volume

The measured rainfall was taken from the DID rain gauge in Paya Indah for two years, from 2015 to 2016. Fig 3 and Fig 4 present the rainfall recorded by the rain gauge. The rain volume or intensity collected annually for one year resembles the characteristic of an equatorial climate rain. It is observed that rain volume, on average, is higher throughout the year of 2016. In 2015, February and August presented lower rainfall values than March, April, May, October, November, December, and January. Rain volume provides a good idea of which month has higher rainfall. The rainfall volume in 2015 was much lower than in 2016 because of Malaysia's long dry season due to the El Nino phenomenon. The year 2016 shows typical equatorial rainfall for Malaysia, and the rainfall is well distributed throughout the year, which corresponds to the discussion based on the literature in the introduction section. The rainfall rate is required to provide valuable information associated with signal attenuation to analyze the satellite communication link. Thus, rainfall volume (mm) must be converted to rainfall rate (mm/hr).

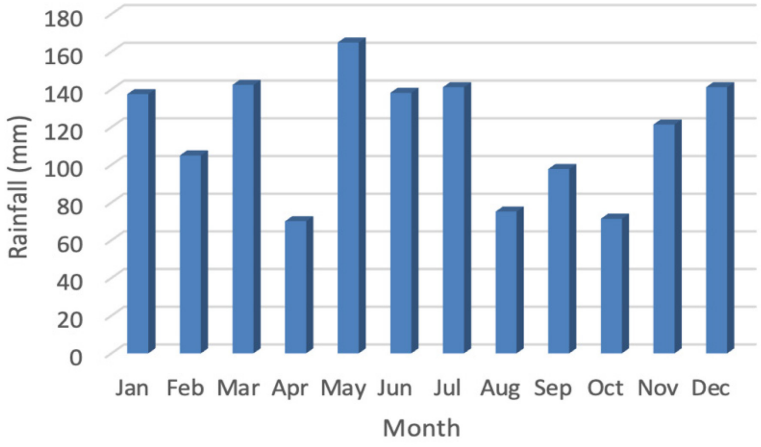


Figure 4. Monthly rainfall volume in 2016 for Paya Indah rain gauge

3.3. Rainfall Rate

Before analysis, the rainfall volume data were converted into rainfall rate. The rainfall rate, Probability Density Function (PDF), and Cumulative Density Function (CDF) were analyzed to produce the time exceedance on a monthly basis for the years 2015 and 2016. The monthly CDF for 2016 and 2015 are presented in Fig. 5 and Fig. 6, respectively. Fig. 7 presents the annual cumulative distribution of the average rainfall rate for 2015 and 2016. There is a small difference between the measurement of the average cumulative distribution for 2015 and 2016. The annual average rainfall rate measured at 0.01% of time exceedance for years 2015 and 2016 are 120 mm/hr and 96 mm/hr, respectively. Despite the dry season with lower rainfall volume, the average annual rainfall rate for 2015 was higher than in 2016. The average point rainfall rate measured at 0.01% of time exceedance was 108 mm/hr. The $R_{0.1}$ for 2015, 2016, and the two-year average were 48 mm/hr, 42 mm/hr, and 42 mm/hr, respectively. The rainfall rate was very high, reaching 222 mm/hr in 2015 and 174 mm/hr in 2016. Compared to the ITU-R P.837-7 [12] value, the measured local value of $R_{0.01}$ at 108 mm/hr is higher than the ITU-R $R_{0.01}$ which was 100 mm/hr for Peninsular Malaysia. The value of the typical time exceedance for the rainfall rate is presented in the next section.

https://doi.org/10.31436/iiumej. v25i2.3279

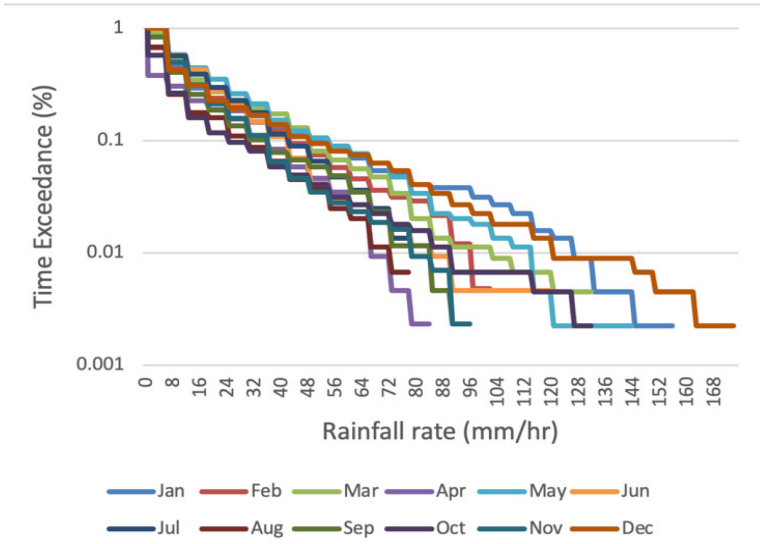


Figure 5. Monthly cumulative distribution of rainfall rate for year 2016

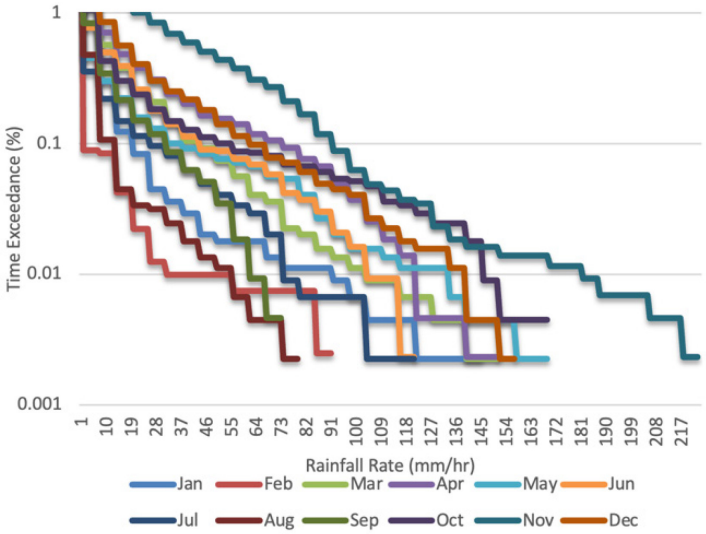


Figure 6. Monthly cumulative distribution of rainfall rate for year 2015.

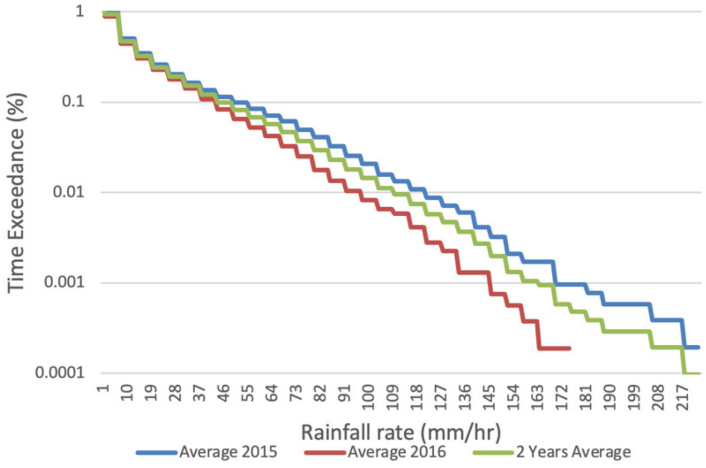


Figure 7. Annual cumulative distributions of rainfall rate for Paya Indah rain gauge

https://doi.org/10.31436/iiumej. v25i2.3279

3.4. Rainfall Rate Model

This section presents the results to achieve the rainfall rate model, which will be the critical input to the specific attenuation and rain attenuation models. The specific attenuation is a function of rainfall rate and k and α coefficients. The rainfall rate measured in time exceedance for Paya Indah will be useful for finding the values of specific attenuation for Cyberjaya only. The rainfall rate model is derived based on steps similar to those of the Paya Indah Rain Gauge. In this section, another rainfall rate analysis was made from four different locations in the Klang Valley area. The additional rain gauges were located in Putrajaya, Sg Merab, Rawang, and Gombak. Together with Paya Indah, five measurements of rainfall rate areas are available. An area-based rainfall rate model based on the average rainfall rate could be derived for the two-year period from the five locations of rainfall rate. The average could be projected to provide a model for the rainfall rate in Klang Valley. The ITU-R model P.837-7 generalized the rainfall rate for the whole of Malaysia territory to be 100 mm/hr at 0.01% exceedance. The specific measurements made for the rainfall rate in five areas could potentially be referred to as the rainfall rate of Malaysia, replacing the ITU-R P.837-7 for a more accurate representation.

The CDF of the monthly rainfall rate for each of the five areas was established. Next, the annual averages were obtained. 10 measurements of rainfall rate annual averages were acquired for the two-year period. The average of these 10 measurements was then calculated in Microsoft Excel. The annual average was subsequently exported to Matlab Curve Fit Tool to obtain a suitable model of the Klang Valley's rainfall rate, which could be generalized as Malaysia's rainfall rate. Fig. 8 presents the curve fit for the average rainfall rate based on 2 years of rainfall rate from five combined areas. In the Y-axis, R_P represents the rainfall rate in mm/hr while p represents the probability or the time exceedance in %. The R^2 of the curve is 0.9832. Subsequently, Eq. (2) represents the derived rainfall rate model.

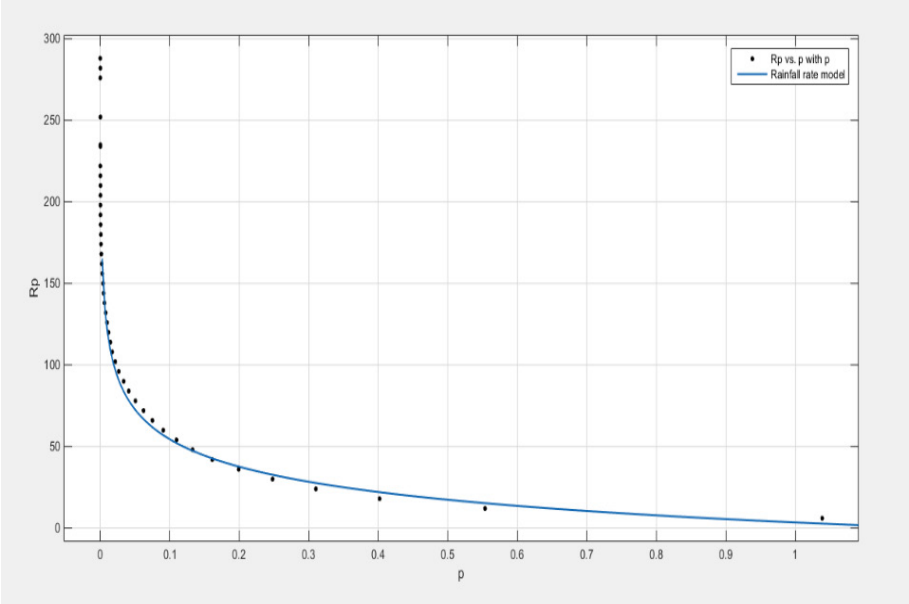


Figure 8. Rainfall rate curve fit for Klang Valley

$$R_{p\%} = 162p\%^{-0.12} - 158.6. \quad (2)$$

Eq. (2) is plotted together with the annual averages of the five rain gauges in the Klang Valley area for the years 2015 and 2016 in Fig. 9. The dots in the figure are the annual average rainfall rates for five different areas. The names of the areas are according to the legend and label in Fig. 9. The red line represents the Rp% model. The orange curve is the average of the

dataset. Based on Fig. 9, the annual average rainfall rate varies from a minimum value of 96 mm/hr to a maximum value of 132 mm/hr. The measurement for 96 mm/hr at 0.01% time exceedance is from Paya Indah rain gauge data in 2016, while the Sg Merab rain gauges measured the maximum at 132 mm/hr in 2016. The average rainfall rate is 126 mm/hr, while the derived model produced 123 mm/hr for the annual rainfall rate average.

As a benchmark, other rainfall rate measurement studies in Malaysia also demonstrated that the annual average $R_{0.01}$ is typically 120 mm/hr with a monthly variation of 58 mm/hr to 136 mm/hr [7]. Thus, this model can be used to accurately obtain the time exceedance for Malaysian territory and equatorial regions. Considering the rainfall rate benchmark of 100 mm/hr at 0.01% by ITU-R, the rainfall rate model derived in this study could be considered a better approximation to the prediction of rainfall rate in Malaysia. The obtained $R_{0.01}$ is crucial for determining the specific attenuation and rain attenuation. This finding is valuable for defining a more accurate model of the rainfall rate in Malaysia, particularly for communication links.

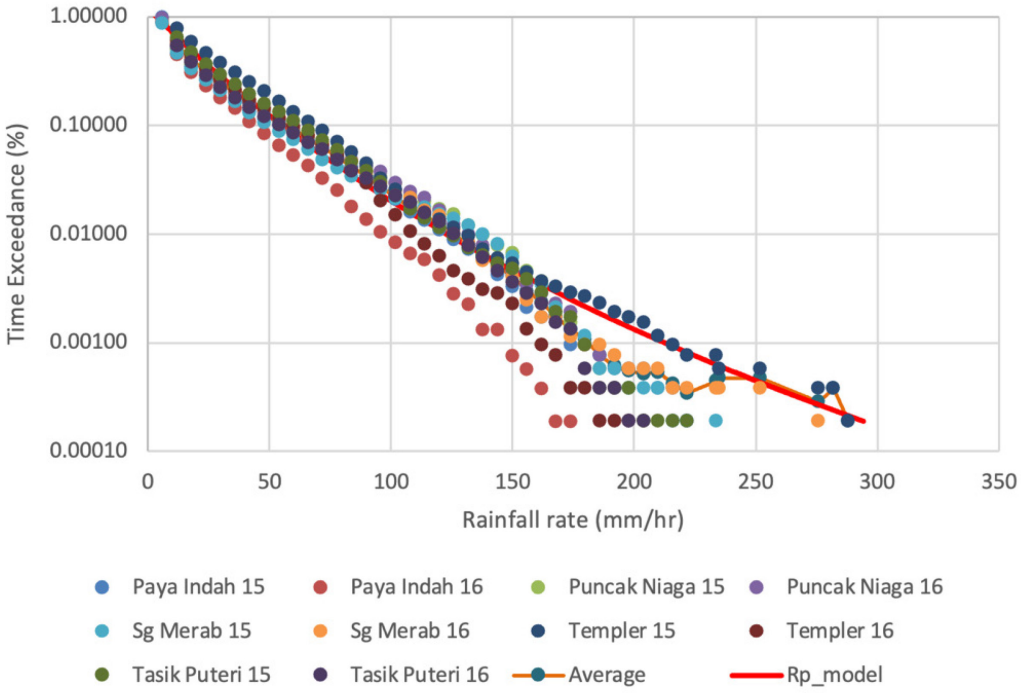


Figure 9. $R_{p\%}$ model plotted for 2 years average of rainfall rate from 5 areas.

4. CONCLUSION

The ITU-R stated that the rainfall rate for the region of Malaysia is 100 mm/hr [12]. Most measurements have shown that the ITU-R underestimated the rainfall rate distribution in equatorial regions, particularly in Malaysia [6], [19], [20], and newer measurements show that the measured rainfall rate is lower than the older one but still higher than the ITU-R model [7]. Most measurements, however, employed a single rain gauge to conclude that the ITU-R underestimated the predictions of rainfall rate. Multiple rain gauges should be used to obtain an accurate rainfall rate over a large area [18]. The methods of acquiring area-based rainfall rates using multiple rain gauges were discussed in the methodology section. As listed in Table 1, five rainfall stations provided rainfall data for the Klang Valley area. The results of the area-based rainfall rate measurements were discussed in the result section. The outcome from five rainfall stations indicates that the annual rainfall rate distribution for $R_{0.01}$ is from 96 mm/hr to

132 mm/hr, and the annual average is 126 mm/hr. From these findings, certain areas have a lower rainfall rate than 100 mm/hr, as recommended by ITU-R. However, most locations have shown results of higher rainfall rates than the ITU-R recommendation. Thus, the annual average is higher than the ITU-R recommendation. Based on the rainfall rate distribution of <1%, a rainfall rate model was developed and shown in Equation (2). The model produces a prediction for rainfall rate distribution exceedance of <1% for the Klang Valley in particular and Malaysia and equatorial regions in general. Thus, this model can contribute to a more accurate specific attenuation and rain attenuation model used in communication link modeling and analysis for the equatorial region.

ACKNOWLEDGEMENT

This project's funding comes from the University Malaysia Pahang (UMP) and International Islamic University Malaysia (IIUM) sustainable research collaboration 2022 grant scheme no. IUMP-SRCG22-001-0001. The authors would also like to acknowledge the Department of Irrigation and Drainage Malaysia, which provided invaluable rainfall data for the academic institution free of charge.

REFERENCES

[1] B. Wang, LinHo, B. Wang, and LinHo, "Rainy Season of the Asian–Pacific Summer Monsoon*," *J. Clim.*, vol. 15, no. 4, pp. 386–398, Feb. 2002, doi: 10.1175/1520-0442(2002)015<0386:RSOTAP>2.0.CO;2.

[2] C.-P. Chang *et al.*, "Annual Cycle of Southeast Asia—Maritime Continent Rainfall and the Asymmetric Monsoon Transition," *J. Clim.*, vol. 18, no. 2, pp. 287–301, Jan. 2005, doi: 10.1175/JCLI-3257.1.

[3] S. A. Zabidi, I. M. Rafiqul, and A. K. Wajdi, "Rain attenuation prediction of optical wireless system in tropical region," *2013 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2013*, no. November, pp. 26–27, 2013, doi: 10.1109/ICSIMA.2013.6717965.

[4] M. Alhilali, H. Y. Lam, and J. Din, "Comparison of Raindrop Size Distribution Characteristics across the Southeast Asia Region," *Telkomnika*, vol. 16, no. 6, pp. 2522–2527, 2018, doi: 10.12928/TELKOMNIKA.v16i6.10091.

[5] K. H. D. Tang, "Climate change in Malaysia: Trends, contributors, impacts, mitigation and adaptations," *Sci. Total Environ.*, vol. 650, no. October, pp. 1858–1871, 2019, doi: 10.1016/j.scitotenv.2018.09.316.

[6] K. N. H. and J. Din, "Rainfall Rate from Meteorological Radar Data for Microwave Applications in Malaysia," in *2005 13th IEEE International Conference on Networks Jointly held with the 2005 IEEE 7th Malaysia International Conf on Communic*, 2005, vol. 2, pp. 1008–1010, doi: 10.1109/ICON.2005.1635660.

[7] I. Shayea, T. A. Rahman, M. H. Azmi, and M. R. Islam, "Real Measurement Study for Rain Rate and Rain Attenuation Conducted Over 26 GHz Microwave 5G Link System in Malaysia," *IEEE Access*, vol. 3536, no. c, pp. 1–1, 2018, doi: 10.1109/ACCESS.2018.2810855.

[8] W. S. Hee, H. S. Lim, M. Z. M. Jafri, S. Lolli, and K. W. Ying, "Vertical profiling of aerosol types observed across monsoon seasons with a Raman Lidar in Penang Island, Malaysia," *Aerosol Air Qual. Res.*, vol. 16, no. 11, pp. 2843–2854, 2016, doi: 10.4209/aaqr.2015.07.0450.

[9] C. Capsoni, L. Luini, A. Paraboni, C. Riva, and A. Martellucci, "A New Prediction Model of Rain Attenuation That Separately Accounts for Stratiform and Convective Rain," *IEEE Trans. Antennas Propag.*, vol. 57, no. 1, pp. 196–204, Jan. 2009, doi: 10.1109/TAP.2008.2009698.

[10] H. Y. Lam, L. Luini, J. Din., C. Capsoni, and A. D. Panagopoulos, "Stratiform and convective rain discrimination for equatorial region," in *2010 IEEE Student Conference on Research and Development (SCOReD)*, Dec. 2010, pp. 112–116, doi: 10.1109/SCORED.2010.5703983.

[11] ITU-R P.838-3, "Specific attenuation model for rain for use in prediction methods," *Recomm. ITU-R P.838-3*, pp. 1–5, 2005.

[12] ITU-R P.837-7, "Characteristics of precipitation for propagation modelling P Series Radiowave

propagation," *Radiowave Propag.*, vol. 6, 2017.
[13] R. Singh and R. Acharya, "Development of a New Global Model for Estimating One-Minute Rainfall Rate," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6462–6468, Nov. 2018, doi: 10.1109/TGRS.2018.2839024.
[14] H. Y. Lam, L. Luini, J. Din, C. Capsoni, and A. D. Panagopoulos, "Assessment of seasonal Asia monsoon rain impact on the Earth-space propagation in equatorial Kuala Lumpur," *IEEE Antennas Propag. Soc. AP-S Int. Symp.*, pp. 1461–1464, 2012.
[15] P. Chandrika, S. Vijaya Bhaskara Rao, N. V. P. Kirankumar, and T. Narayana Rao, "Review and testing analysis of Moupfouma rain rate model for Southern India," *J. Atmos. Solar-Terrestrial Phys.*, vol. 132, pp. 33–36, Sep. 2015, doi: 10.1016/J.JASTP.2015.06.010.
[16] L. Csurgai-Horvath and J. Bito, "Rain Intensity Estimation Using Satellite Beacon Signal Measurements A Dual Frequency Study," in *2018 International Symposium on Networks, Computers and Communications (ISNCC)*, Jun. 2018, pp. 1–5, doi: 10.1109/ISNCC.2018.8530986.
[17] R. K. Das and N. R. Prakash, "Design of an improvised tipping bucket rain gauge for measurement of rain and snow precipitation," *Int. J. Instrum. Technol.*, vol. 1, no. 1, p. 44, 2011, doi: 10.1504/IJIT.2011.043597.
[18] M. L. Tan and H. Santo, "Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia," *Atmos. Res.*, vol. 202, pp. 63–76, Apr. 2018, doi: 10.1016/j.atmosres.2017.11.006.
[19] R. Islam, T. A. Rahman, and Y. Karfaa, "Worst-month rain attenuation statistics for radio wave propagation study in Malaysia," in *9th Asia-Pacific Conference on Communications (IEEE Cat. No.03EX732)*, 2003, pp. 1066–1069, doi: 10.1109/APCC.2003.1274262.
[20] J. S. Mustapha, A; Nazri, F. A; Mandeep, "Kajian Pelemahan Hujan pada Jalur K u dan Jalur K a di Malaysia," vol. 26, no. 2014, pp. 45–54, 2015.
[21] H. M. Ibrahim, "General Specification for Tipping Bucket used by DID," 2017.
[22] HydroNET, "HydroNet. - Water Resources Management and Hydrology Division," 2018. http://h2o.water.gov.my/v2/ (accessed Apr. 10, 2019).
[23] M. C. Kestwal, S. Joshi, and L. S. Garia, "Prediction of rain attenuation and impact of rain in wave propagation at microwave frequency for tropical region (Uttarakhand, India)," *Int. J. Microw. Sci. Technol.*, vol. 2014, 2014, doi: 10.1155/2014/958498.

https://doi.org/10.31436/iiumej.v25i2.3058

PARTIAL REPLACEMENT OF SILICA NANOPARTICLES IN CEMENT PASTE FOR CO_2 REGENERATION CAPTURE APPLICATION

MUDRIKAH SOFIA MAHMUD[1], FARAH DIANA MOHD DAUD[1*], NORSHAHIDA SARIFUDDIN[1], HAFIZAH HANIM MOHD ZAKI[1], NORHUDA HIDAYAH NORDIN[1], NUR FARAHIYAH MOHAMMAD[2]

[1]Department of Manufacturing and Materials Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100, Kuala Lumpur, Malaysia
[2] Faculty of Electronics Engineering Technology, Universiti Malaysia Perlis, Pauh Putra, 02600 Arau, Perlis, Malaysia

**Corresponding author: farah_diana@iium.edu.my*

(Received: 3 October 2023; Accepted: 13 May 2024; Published online: 15 July 2024)

ABSTRACT: A promising, supplementary material, highly reactive silica nanoparticles tend to react with calcium species (Ca^{2+}) in cement, which produces more hydrated phases. The high amount of calcium species and the potential of capturing carbon dioxide (CO_2) through carbonation implies the suitability of cement paste as a CO_2 sorbent material. However, limited studies could be found highlighting the effect of silica nanoparticle inclusion on the CO_2 capture property of hardened cement paste. Thus, this study investigates the effect of silica nanoparticle partial replacement on the CO_2 capture capacity and regeneration performance of hardened cement paste. XRD and SEM analysis proved the formation of more hydrated phases upon partially replacing silica nanoparticles in cement. Partial replacement of 3.00 % with silica nanoparticles significantly improved CO_2 regeneration capture at room temperature among the other cement paste samples. This study found that the presence of silica nanoparticles in cement paste triggered the formation of more hydrated phases, which served as active sites for CO_2 capture to occur and, thus, improved the CO_2 capture capacity and regeneration performance.

ABSTRAK: Sebagai salah satu bahan tambahan yang berpotensi, nanozarah silika sangat reaktif dan cenderung bertindak balas dengan spesies kalsium (Ca^{2+}) dalam simen di mana terhasilnya lebih banyak fasa terhidrat. Jumlah spesies kalsium yang tinggi dan berpotensi menangkap karbon dioksida (CO_2) melalui pengkarbonan menunjukkan kesesuaian pes simen digunakan sebagai bahan penyerap CO_2. Walau bagaimanapun, kajian adalah terhad terhadap kesan kemasukan nanozarah silika pada sifat tangkapan pes simen CO_2 yang dikeraskan. Oleh itu, kajian ini adalah berkenaan kesan penggantian separa nanopartikel silika ke atas kapasiti tangkapan CO_2 dan prestasi penjanaan semula pes simen yang dikeraskan. Analisis XRD dan SEM membuktikan pembentukan lebih banyak fasa terhidrat apabila penggantian separa nanopartikel silika pada simen. Penggantian separa nanopartikel silika 3% menunjukkan peningkatan ketara tangkapan penjanaan semula CO_2 pada suhu bilik di antara sampel pes simen lain. Kajian ini mendapati bahawa kehadiran nanozarah silika dalam pes simen mencetuskan pembentukan lebih banyak fasa terhidrat yang berfungsi sebagai tapak aktif bagi penangkapan CO_2 berlaku dan dengan ini, membawa kepada peningkatan kapasiti penangkapan CO_2 dan prestasi penjanaan semula.

KEYWORDS: *Silica nanoparticles, cement paste, hydrated phases and CO_2 regeneration capture.*

https://doi.org/10.31436/iiumej.v25i2.3058

1. INTRODUCTION

Cement-based material could be one of the potential calcium-based materials for the CO_2 capture application since it contains a high amount of calcium species and possesses an ability to react with CO_2. A similar argument was made by Chai et al. [1]. This ability to react with carbon dioxide is known as carbonation [2-4]. CO_2 tends to react with the main hydration compounds in the cement-based materials matrix, calcium hydroxide ($Ca(OH)_2$) and calcium silicate hydrate (C-S-H), which leads to the formation of calcium carbonate ($CaCO_3$). Thus, carbonation proved that CO_2 capture was possible for cement-based materials. Since the CO_2 capture of cement-based materials would likely proceed even after demolition [1,3], the potential of cement-based materials to capture CO_2 could be an excellent advancement in CO_2 capture applications.

Production of more of these hydration products would likely improve the properties of cement-based materials, including the CO_2 capture property. Previously, it had been reported that the inclusion of silica nanoparticles had significantly improved the properties of the hardened cement-based materials due to the occurrence of an aggressive hydration reaction that led to the formation of more hydrated phases [5-11]. With the presence of water, the nano-sized silica tends to react with calcium species (Ca^{2+}) in cement and lead to the formation of C-S-H and secondary C-S-H. Thus, improvement in properties could be expected. However, limited studies could be found that highlight the effect of silica nanoparticle inclusion on the CO_2 capture property of the hardened cement paste. Thus, this study investigates the effect of silica nanoparticle partial replacement on the CO_2 capture capacity and regeneration performance of the hardened cement paste.

2. EXPERIMENTAL

2.1. Materials

Silica nanoparticle was produced from rice husk ash (RHA) using the precipitation method, according to a previous study [12]. The powder obtained contains high-purity (98.91%) amorphous silica nanoparticles, proven by XRF analysis. Fig. 1 shows the micrograph of silica particles in the synthesized powder that have a particle diameter of less than 50 nm. Meanwhile, the Ordinary Portland cement (OPC) used was brand YTL Cement and obtained from a local supplier.

2.2. Preparation of Cement Paste Samples

The percentage of silica nanoparticles was limited to a maximum of 3.00 % weight since the largest increment in compressive strength was achieved with partial replacement of 3.00 % silica nanoparticles, as stated in an earlier study by [13]. The water-to-cement (w/c) ratio was maintained at 0.4 to highlight the effect of partially replacing silica nanoparticles in cement paste. The samples were prepared based on Table 1. Initially, silica nanoparticles and OPC powders were mixed. Then, an appropriate amount of water was added to the mixture. Meanwhile, the procedure was to mix the OPC cement powder and water to prepare the reference sample. All the mixing procedures were conducted using a mortar mixer. Next, the paste was poured into the 50 mm x 50 mm x 50 mm cube-shaped mold. This steel mold was brushed with grease to ensure easy removal of cement cubes after they were hardened. After 24 hours, all the hardened cement/silica nanoparticle paste samples were unmoulded and proceeded to the curing process by submerging them entirely in water. At a standard curing

https://doi.org/10.31436/iiumej.v25i2.3058

duration of 28 days [14], each composition's hardened samples were taken out from the water, and the specified characterization proceeded.

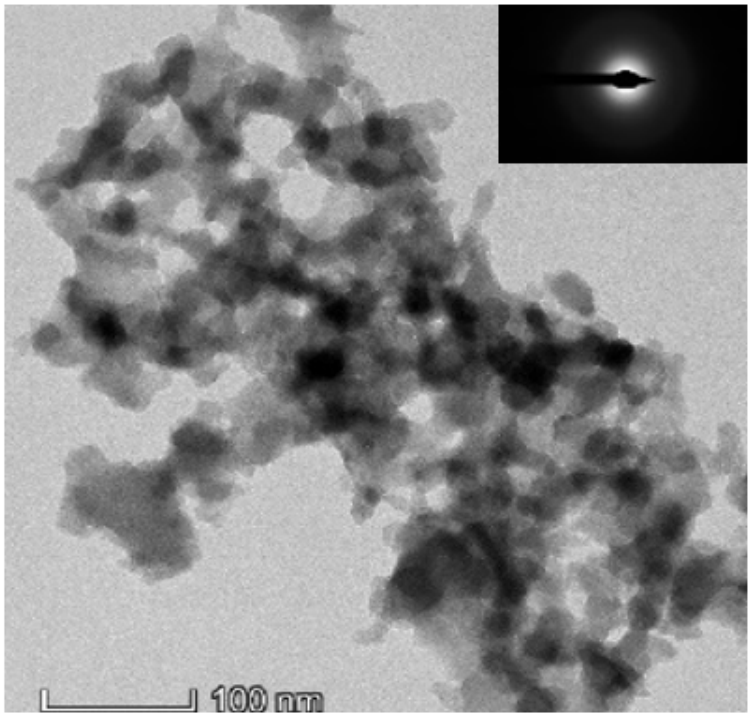


Figure 1. TEM micrograph of the synthesized silica nanoparticle powder [12].

Table 1. Composition of cement paste samples.

Sample	Cement (%)	Silica nanoparticle (%)
REF	100.00	0.00
NS0	99.25	0.75
NS1	98.50	1.50
NS2	97.75	2.25
NS3	97.00	3.00

2.2. Characterization

2.1.1. X-Ray Diffraction (XRD) Analysis

Diffraction (XRD) for phase structure analysis. This analysis uses the D2 Phaser machine by Bruker. The pattern was collected at 2θ range from 5° to 70°.

2.1.2. Scanning Electron Microscopy – Energy Dispersive X-Ray Spectroscopy (SEM-EDX) Analysis

The morphological characteristics observation and elemental mapping analysis were conducted through Scanning Electron Microscopy – Energy Dispersive X-Ray Spectroscopy (SEM-EDX) analysis using a JSM-IT100 machine by JEOL. The analysis was done by comparing the micrographs obtained with the ones from previous research [15,16]. The micrographs of the reference and cement paste samples with 3.00 % silica nanoparticle partial replacement were compared to highlight the hydration phase development. Before the analysis, the samples were oven-dried at 100°C overnight and coated with palladium using a mini sputter coater SC7620.

https://doi.org/10.31436/iiumej.v25i2.3058

2.1.3. CO_2 Regeneration Capture Performance

This CO_2 regeneration capture experiment utilized the prepared cubic cement samples, a container with gas ports, and a pure CO_2 gas supply alongside a timer and weighing scale. Each cubic sample's initial weight was measured before being sealed in a container where CO_2 gas was introduced for 30 minutes. Afterward, the sample was removed, and its weight was measured again. This process was repeated for a total of 150 minutes (5 cycles), which was broken into 30 minutes per cycle with a 10-minute gap between cycles. The CO_2 capture capacity for each cycle was calculated as follows:

$$CO_2\ capture\ capacity_x = \frac{W_x(g) - W_i(g)}{W_i(g)} \times 1000 \quad (1)$$

According to Eq. (1), the CO_2 capture capacity was determined by subtracting the initial weight of the sample (W_i) from its weight after 30 minutes of exposure (W_x), with x representing the cycle number. The CO_2 capture capacity was then graphed against the percentage of nano-silica partial replacement to observe the material's performance over 150 minutes of CO_2 exposure at 28 days of curing, providing insights into the prepared sample's performance in capturing CO_2 at room temperature. The experimental setup is shown in Fig. 2.

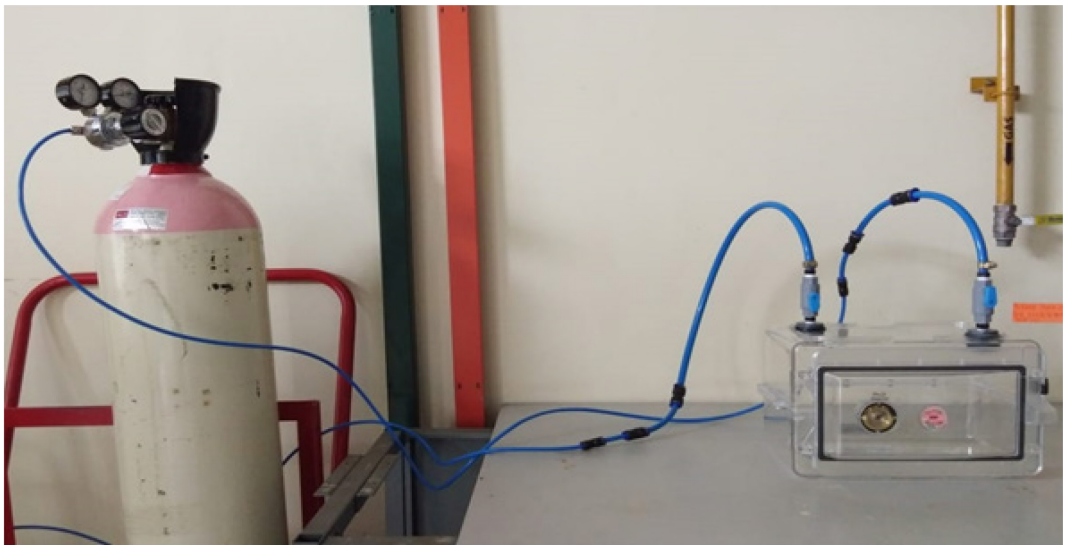


Figure 1. Customized experimental setup for CO_2 regeneration capture performance test.

3. RESULTS AND DISCUSSION

3.1. XRD Analysis: Development of Hydration Phases Upon Silica Nanoparticles Addition

Fig. 3 displays the XRD pattern of the samples containing different silica nanoparticle partial replacement percentages after 28 days of curing. For this analysis, the characteristic peaks of interest were $Ca(OH)_2$ and C-S-H phases. As can be observed, all samples exhibited a similar pattern, which consisted of similar characteristic peaks at 2θ similar locations, however, with different intensities. A similar trend was observed in previous studies [17-19], in which the peaks located at about 30.00° belonged to the C-S-H phase, while peaks at about 18.00° and 34.00° belonged to the $Ca(OH)_2$ phase. Previous research claimed that the peak intensity could reflect the content of the identified phase [19]. Thus, this finding could indicate the presence of similar phases in all samples, but in different proportions.

https://doi.org/10.31436/iiumej.v25i2.3058

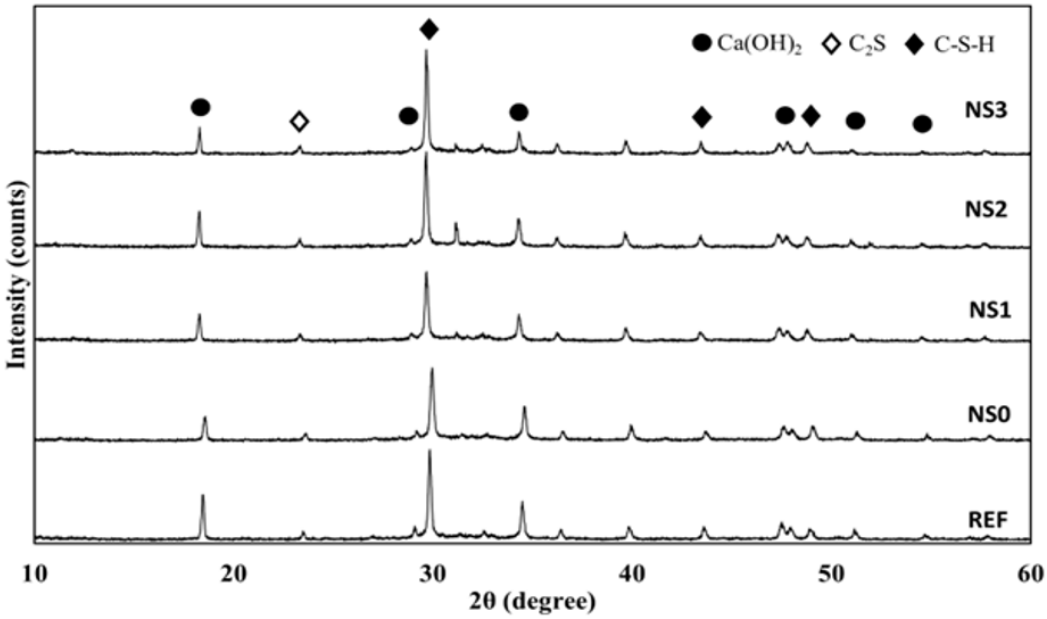


Figure 3. XRD patterns of the hardened cement paste samples at 28 days of curing.

Table 2 summarizes the peak intensity of the $Ca(OH)_2$ and C-S-H phases in samples at 28 days of curing upon different percentages of silica nanoparticle partial replacement. The result indicates that as the percentage of silica nanoparticles increases, there is a clear pattern of increasing and decreasing among these characteristic peaks. As can be observed, partial replacement of 3.00 % silica nanoparticle exhibited the highest increment in the intensity of the C-S-H peak, which was about 11%, and the highest decrement in intensity for both $Ca(OH)_2$ peaks as compared to the sample REF and other percentages of silica nanoparticle. According to this analysis, the presence of silica nanoparticles in cement paste may cause the $Ca(OH)_2$ phase to be consumed and an additional C-S-H phase to be produced.

Table 2. Intensities of the $Ca(OH)_2$ and C-S-H peaks in the XRD pattern.

Sample	Silica nanoparticle (%)	Peak intensity (cps)		
		~18°	~29°	~34°
REF	0.00	37.03	72.21	30.01
NS0	0.75	19.47	58.64	27.85
NS1	1.50	21.71	54.94	20.34
NS2	2.25	29.10	75.69	22.62
NS3	3.00	17.95	79.84	19.63

It has been reported that silica nanoparticles could promote hydration reactions in hardened cement paste materials [6,11,18,20,21]. This is a result of silica nanoparticles' tendency to react with calcium species (Ca^{2+}) in cement, which increases the production of an additional C-S-H phase, known as the secondary C-S-H phase. Hence, the availability of more C-S-H phases in the hardened cement paste structure could significantly improve the properties of the hardened cement paste samples.

3.2. Morphological Observation and Elemental Mapping Analysis

The development of these hydrated phases in the hardened cement paste samples was further investigated by SEM-EDX analysis. Fig. 4 shows the SEM micrographs of all samples at 28 days of curing at 1500X magnification. The micrographs displayed the formation of more fibrous and flocculated structure throughout the observed morphology upon increasing

https://doi.org/10.31436/iiumej.v25i2.3058

percentage of silica nanoparticles partial replacement. As compared with a previous study by [15], the micrographs of all samples displayed the common morphological traits of the hydration phases that were present in cement-based materials.

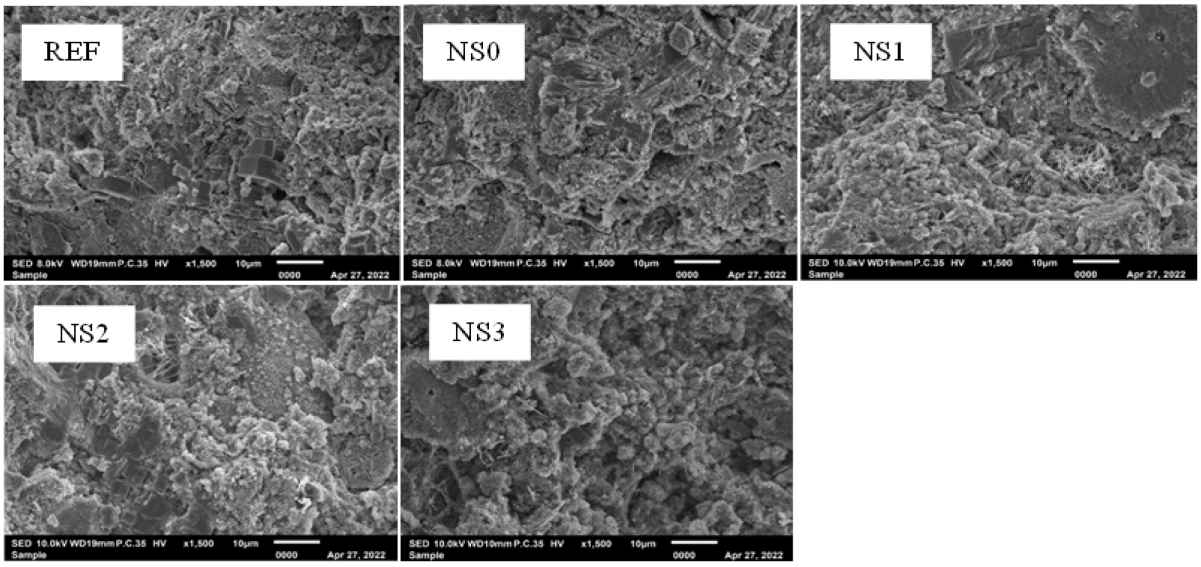


Figure 4. SEM micrographs of cement paste samples upon different percentages of silica nanoparticle partial replacement at 28 days of curing.

To better understand these morphological characteristics, the micrographs of the reference sample and cement paste sample with partial replacement of 3.00 % silica nanoparticle (sample NS3) were compared in Fig. 4 at 2500X magnification.

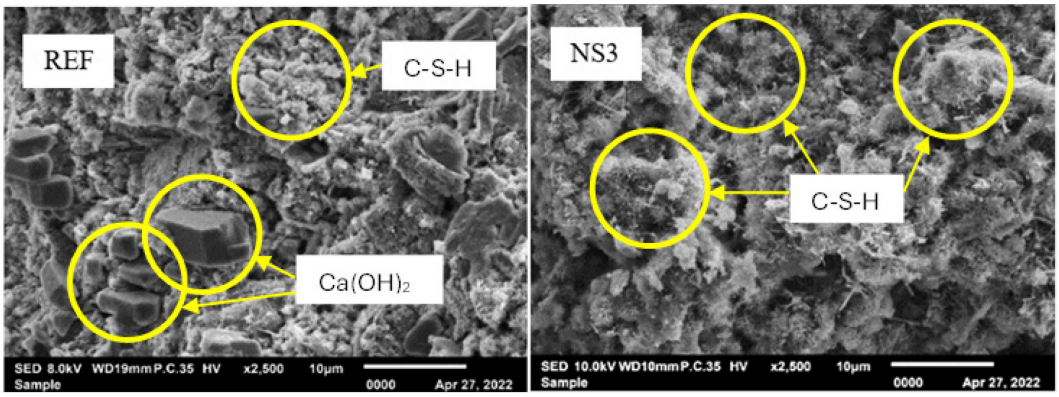


Figure 5. SEM micrographs of the sample REF and NS3 at 28 days of curing.

Previously, researchers have reported that the sheet-like structure indicates the presence of the $Ca(OH)_2$ phase; meanwhile, the C-S-H phase could be represented in different types of structures, such as fibrous and flock-like structures [15,16]. According to the micrographs obtained, sample NS3 has a morphology that is mostly covered with the C-S-H phase compared to sample REF. This observation suggests that more formation of the C-S-H phase could be expected upon increasing the percentage of silica nanoparticle partial replacement in cement.

Moreover, this finding is supported by Table 3 which summarizes the Ca/Si atomic ratio of the samples at 28 days of curing which was obtained from EDX analysis. Table 3 shows that the values of the Ca/Si ratio of the cement paste samples decreased upon the increasing percentage of silica nanoparticle partial replacement, in which the highest percentage of decrement of about 40 % was observed in sample NS3 compared to the sample REF. Previous

researchers claimed that the decrement in Ca/Si ratio value indicates the presence of additional C-S-H phases in cement that led to the development of a densified and compact microstructure [10,17]. Thus, it could be concluded that more C-S-H phase was formed when a 3.00 % silica nanoparticle was partially replaced compared to the other samples, which agreed with the findings in XRD and SEM analyses.

Table 3. Ca/Si ratio of each cement paste sample.

Sample	Silica nanoparticle (%)	Ca/Si
REF	0.00	5.08
NS0	0.75	4.56
NS1	1.50	4.16
NS2	2.25	4.20
NS3	3.00	3.05

These phase analyses suggest that silica nanoparticles could modify the morphology of the cured cement paste by promoting a more intense hydration reaction. The properties could significantly improve as the additional hydration products formed in the hardened cement paste structure. Hence, better CO_2 regeneration capture of the hardened cement paste sample could be expected upon the partial replacement of silica nanoparticles.

3.3. CO_2 Regeneration Capture Performance

Fig. 6 shows the samples' CO_2 capture capacity and regeneration performance based on the percentage of silica nanoparticle partial replacement. Based on the result, the CO_2 capture capacity and the regeneration performance of the samples increased as the percentage of silica nanoparticle partial replacement increased throughout the 150 minutes of CO_2 exposure.

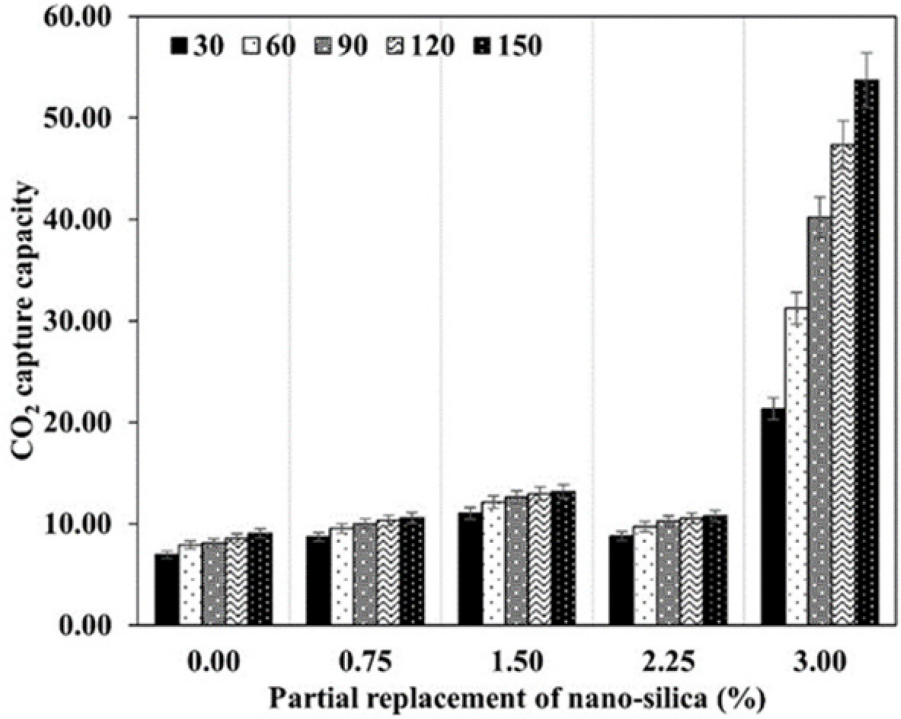


Figure 6. CO_2 regeneration captures the performance of all cement paste samples at 28 days of curing.

The partial replacement of 3.00 % silica nanoparticles yields a remarkable performance. For the initial CO_2 capture capacity (the first 30 minutes of CO_2 exposure), up to 208.25 % increment in CO_2 capture capacity was achieved upon the 3.00 % silica nanoparticle partial replacement (21.44 mg CO_2 captured/mg sample) compared to the reference sample (6.95 mg CO_2 captured/mg sample)

Based on the result, the CO_2 capture capacity and the regeneration performance of the samples increased as the percentage of silica nanoparticle partial replacement increased throughout the 150 minutes of CO_2 exposure. The partial replacement of 3.00 % silica nanoparticles yielded a remarkable performance. For the initial CO_2 capture capacity (the first 30 minutes of CO_2 exposure), up to 208.25 % increment in CO_2 capture capacity was achieved upon the 3.00 % silica nanoparticle partial replacement (21.44 mg CO_2 captured/mg sample) compared to the reference sample (6.95 mg CO_2 captured/mg sample).

During the 150 minutes of CO_2 exposure, cement samples with silica nanoparticles exhibited better CO_2 regeneration capture performance than the reference sample. The sample with 3.00 % silica nanoparticle partial replacement exhibited a total of 53.88 mg CO_2 captured/mg sample capture capacity, which indicates about a 493.76 % increment in CO_2 regeneration capture performance throughout the 5 cycles of CO_2 exposure compared to the sample REF, which only exhibited a total of 9.07 mg CO_2 captured/mg sample capture capacity. According to the result obtained, the CO_2 capture capacity could experience further increment if the experimental duration was prolonged, as the result still indicates an increasing pattern in the CO_2 capture capacity.

More C-S-H phases have formed upon the partial replacement of silica nanoparticles in cement. As explained in the carbonation reaction, these hydration products are responsible for the CO_2 capture property. So, the presence of more C-S-H phases allows more CO_2 to be captured. This led to a significant increment in the CO_2 capture capacity, and the development of cement paste with better regeneration capture performance could be expected. Similar findings were obtained from previous studies [2,22].

These findings indicate that the presence of silica nanoparticles in cement affects the CO_2 regeneration capture performance of the hardened cement paste samples. A significant improvement was achieved with a partial replacement of 3.00 % silica nanoparticles.

4. CONCLUSION

This research's findings indicate that silica nanoparticles altered the morphology of the hardened cement paste samples and affected its CO_2 capture property and regeneration performance. A significant improvement was exhibited by the hardened cement paste sample with partial replacement of 3.00 % silica nanoparticle, compared to the other samples. The inclusion of silica nanoparticles boosts cement hydration and increases C-S-H phase production. Production of more C-S-H phases had remarkably influenced the CO_2 regeneration capture ability of the hardened cement paste sample. Upon the presence of silica nanoparticles in cement, a higher capture capacity at each cycle of CO_2 exposure could be expected, which implies better regeneration performance.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2019/TK05/UIAM/03/1).

REFERENCES

[1] Cai J, Wang S, Xiao Z (2018). A study on the CO_2 capture and attrition performance of construction and demolition waste. Fuel, 222:232-242.

[2] Abid K, Gholami R, Elochukwu H, Mostofi M, Chua HB, Muktadir G (2018). A methodology to improve nanosilica based cements used in CO_2 sequestration sites. Petroleum, 4:198-208.

[3] Possan E, Thomaz WA, Aleandri GA, Felix EF, Santos A (2017). CO_2 uptake potential due to concrete carbonation: a case study. Case Studies in Construction Materials, 6:147–161.

[4] Hirata M, Jimbo I (2016). Utilization of concrete waste to capture CO_2 with zeolite. Proceedings of the School of Engineering, Tokai University, Series E, 41:9-13.

[5] Karakouzian M, Farhangi V, Farani MR, Joshaghani A, Zadehmohamad M, Ahmadzadeh M (2021). Mechanical characteristics of cement paste in the presence of carbon nanotubes and silica oxide nanoparticles: an experimental study. Materials, 14(1347): 1-14.

[6] Meng T, Ying K, Yang X, Hong Y (2021). Comparative study on mechanisms for improving mechanical properties and microstructure of cement paste modified by different types of nanomaterials. Nanotechnology Reviews, 10 (1):370-384.

[7] Wang X, Gong C, Lei J, Dai J, Lu L, Cheng X (2021). Effect of silica fume and nano-silica on hydration behavior and mechanism of high sulfate resistance portland cement. Construction and Building Materials, 279:122481.

[8] Raza A, Bhandari M, Kim H-K, Son H-M, Huang B, Nam I-W (2021). A study on mechanical characteristics of cement composites fabricated with nano-silica and carbon nanotube. Applied Science, 11:152, 1-16.

[9] Mendes T, Repette W (2021). Nano-silica added to Portland cement. Acta Scientiarum Technology, 43(e51699):1-10.

[10] Kim T, Hong S, Seo K-Y, Kang C (2019). Characteristics of ordinary Portland cement using the new colloidal nano-silica mixing method. Applied Science, 9(4358):1-20.

[11] Xiao H, Zhang F, Liu R, Zhang R, Liu Z, Liu H (2019). Effects of pozzolanic and nonpozzolanic nanomaterials on cement-based materials. Construction and Building Materials, 213:1-9.

[12] Mahmud MS, Mohd Daud FD, Sariffudin N, Mohd Zaki HH, Nordin NH, Mohammad NF (2022). high purity nano-silica from rice husk ash (rha) via chemical method as additive/stabilizing agent for CO_2 capture application. In Key Engineering Materials, 908:73–378. Trans Tech Publications, Ltd.

[13] Mahmud MS, Aadnan AF, Mohd Daud FD, Sariffudin N, Mohd Zaki HH, Nordin NH, Mohammad NF (2023). Cement-based with partial replacement of nano-silica for improvement in compressive strength. In: Maleque MA, Ahmad Azhar AZ, Sarifuddin N, Syed Shaharuddin SI, Mohd Ali A, Abdul Halim NFH, (eds) Proceeding of 5th International Conference on Advances in Manufacturing and Materials Engineering. Lecture Notes in Mechanical Engineering, 483–489. Springer, Singapore.

[14] Wu Z, Khayat KH, Shi C (2019). Changes in rheology and mechanical properties of ultra-high performance concrete with silica fume content. Cement and Concrete Research, 123: 105786.

[15] Franus W, Panek R, Wdowin M (2015). SEM investigation of microstructures in hydration products of Portland cement. In: 2nd International Multidisciplinary Microscopy and Microanalysis Congress, Springer Proceedings in Physics, 164:105-112.

[16] Bu Y, Hou X, Wang C, Du J (2018). Effect of colloidal nanosilica on early-age compressive strength of oil well cement stone at low temperature. Construction and Building Materials, 171:690-696.

[17] Snehal K, Das BB, Akanksha M (2018). Early age, hydration, mechanical and microstructure properties of nano-silica blended cementitious composites. Construction and Building Materials 233(117212):1-16.

[18] Sharkawi AM, Abd-Elaty MA, Khalifa OH (2018). Synergistic influence of micro-nano silica mixture on durability performance of cementitious materials. Construction and Building Materials, 164:579-588.

[19] Wang L, Zheng D, Zhang S, Cui H, Li D (2016). Effect of nano-SiO_2 on the hydration and microstructure of Portland cement. Nanomaterials, 6(12):1-15.

[20] Li G, Liu Q, Niu M, Cao L, Nan B, Shi C (2020). Characteristic of silica nanoparticles on mechanical performance and microstructure of sulphoaluminate cement/ordinary Portland cement binary blends. Construction and Buildings Materials, 242:118158.

[21] Rai S, Tiwari S (2018). Nano silica in cement hydration. Materials Today: Proceedings, 5:9196-9202.

[22] Pacheco-Torgal, F., Shasavandi, A. and Jalal, S. (2013). Eco-efficient concrete using industrial wastes: a review. Materials Science Forum, 730-732:581-586.

https://doi.org/10.31436/iiumej.v25i2.2964

A STUDY ON THE TERRESTRIAL MOBILITY OF A SPHERICAL AMPHIBIAN ROBOT

MOHD BAZLI BAHAR[1,2,3]*, SHAHRUM SHAH ABDULLAH[2], MOHD SHAHRIEEL MOHD ARAS[1,3], MOHAMAD HANIFF HARUN [1,3], MUHAMAD KHAIRI ARIFIN[1], FARIZ ALI@IBRAHIM[1]

[1]Fakulti Teknologi dan Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[2]Department of Electronic Systems Engineering, Malaysia-Japan International Institute of Technology, UTM KL Campus, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[3]Underwater Technology Research Group (UTeRG), Center for Robotics and Industrial Automation (CERIA), Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.

**Corresponding author: mohdbazli@utem.edu.my*

(Received: 17 August 2023; Accepted: 9 January 2024; Published online: 15 July 2024)

ABSTRACT: Amphibian spherical robots are an appealing and practical alternative that can move around on different surfaces and function in aquatic environments. Spherical robots boast remarkable mobility and robustness, enabling them to navigate and perform exploration and reconnaissance tasks even in challenging or harsh environments. This paper explores assessing the terrestrial travel capabilities of a proposed amphibian spherical robot. A rapid Prototyping machine (RPM) was used to print the prototype's main shell, yoke, and circuit holders. One main circuit was built on the yoke, while the other was positioned in the bottom shell. The driving principle used the barycentre offset notion, in which a pendulum mass is used to vary the location of the mass to generate a motion. Additional mass is added to the pendulum to determine the robot's performance when mass is altered. The results reveal that the robot can travel on land with a maximum velocity of 40.75 degrees per second while carrying 600 grams of weight and a turning angle of 22.8 degrees. The robot can only move when the additional mass exceeds 400 grams.

ABSTRAK: Robot sfera amfibia ialah satu alternatif menarik dan praktikal yang dapat bergerak di atas permukaan berbeza dan berfungsi dalam persekitaran akuatik. Robot sfera mempunyai mobiliti yang luar biasa dan tahan lasak. Ciri ini membolehkannya bergerak dan menjalankan tugas penerokaan dan peninjauan dalam persekitaran merbahaya atau buruk. Kajian ini adalah bagi menganalisis prestasi robot sfera amfibia ketika bergerak di darat. Mesin Pemprototaip Pantas (RPM) digunakan bagi mencetak badan utama prototaip, yok, dan pemegang litar. Satu litar utama dibina pada yok, manakala satu lagi diletakkan di bahagian bawah. Prinsip pemanduan adalah dengan menggunakan konsep penentuan kedudukan pusat jisim, di mana jisim pendulum digunakan bagi menentukan lokasi jisim sambil menghasilkan gerakan. Berat tambahan pada jisim pendulum ditambah bagi mendapatkan prestasi robot apabila jisim diubah. Dapatan kajian menunjukkan robot dapat bergerak atas darat dengan halaju maksimum 40.75 darjah sesaat sambil membawa 600 gram berat dengan sudut pusingan sebanyak 22.8 darjah. Robot hanya boleh bergerak apabila tambahan jisim melebihi 400 gram.

KEYWORDS: *spherical robot, terrestrial motion, pendulum*

https://doi.org/10.31436/iiumej.v25i2.2964

1. INTRODUCTION

The earth's surface has various structures like hills, cliffs, lakes, and rivers. The sea is said to cover 71% of the earth's surface [1]. In finding or collecting earth resources, workers face various problems, primarily related to safety [2-3]. Besides, the earth is often plagued by natural disasters [4-5]. Search and rescue tasks should be done as soon as possible and carried out safely for the rescuer and the victim. Therefore, a robotic system capable of carrying out activities to locate and collect resources and be used as a first responder when disasters occur is crucial [6]. These robots need to be equipped with the ability to move around on various surfaces and operate in watery areas. Besides, the robot's durability must also be emphasized to ensure it can successfully perform its task.

One attractive and practical solution to these issues is using a spherical robot. A spherical robot motion is based on the rolling concept inspired by the pangolins [7]. This locomotive mode is faster and safer as its spherical body becomes a protective shield. A spherical robot has high mobility and is robust against its surrounding environments [8], which allows it to perform exploration and reconnaissance tasks in unfriendly or harsh environments [9]. Furthermore, spherical robots are naturally stable and can recover from collisions [7]. They also can conceal and protect all the essential parts inside the sphere against environmental states such as moisture, radiation, dust, hazardous material, or water pressure [10]. The steering hydrodynamic force can be reduced as the robot is spherical, increasing movement flexibility [3]. Combining different driving mechanisms makes the robot amphibious, meaning it has terrestrial and aquatic abilities. The major challenge is mass distribution as it has a round shape nature [11]. Furthermore, limited physical size and compact structure require an optimal actuator for both conditions [12]. Therefore, a suitable driving principle is crucially needed to cater to both the design constraints and maneuverability.

An amphibian spherical robot commonly combines terrestrial and underwater driving principles. To optimize the actuation, both driving principles are set to work together, especially when operating underwater [3,13-14]. The terrestrial drive units control thrust direction to improve the system's maneuverability. Applying a legged actuator to move on land and underwater is another solution to cater to the maneuverability of an amphibian spherical robot [15-17]. Still, the design has more actuators, and complex motion planning is needed.

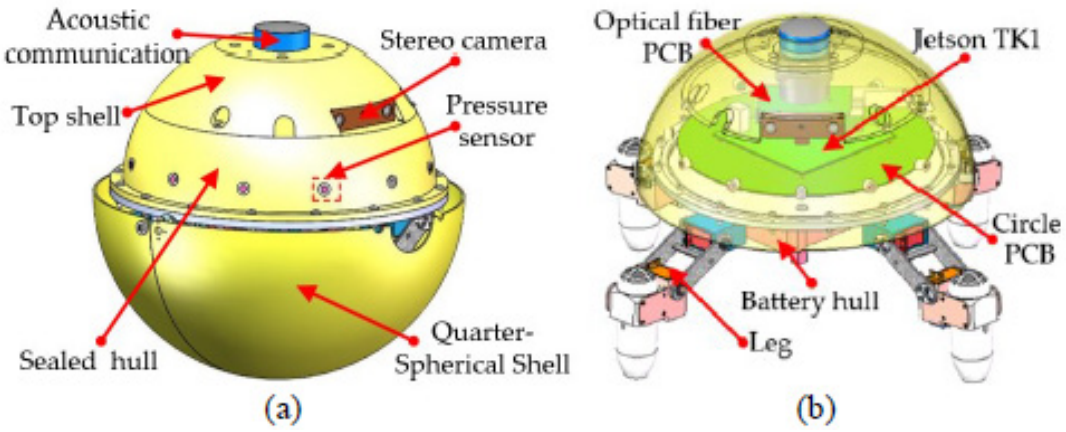


Fig. 1. The legged design proposed in [15] where 4 legs were used as actuators to move on terrestrial and a water jet was installed at the tip of the legs to provide thrust when traveling underwater.

Therefore, a novel amphibian robot was developed to reduce complexity while taking design restrictions and maneuverability into account. The research investigates the assessment of a proposed amphibian spherical robot's terrestrial travel capabilities in terms of motion pattern, turning angle, and motion velocity.

2. THEORETICAL BACKGROUND

Several driving principles have been found, generally divided into three (3) different types. These include methods that move the robot's center of mass (barycenter offset), methods that rely on changing the spherical shell's shape, and methods that generate variable gyrostatic momentum [18]. From this three-basis method, a combination of any one of these was also proposed [19]. The barycenter offset involves shifting the mass location within the sphere from one point to another. This necessitates integrating moving components within the sphere, which can take the form of a cart, pendulum, or slider [8]. Among these options, the rotating pendulum method is a prevalent choice in prior research endeavors. The design has at least two drive motors; one is used to roll the yoke, while another is used to control the pendulum angle [11,20-26].

However, Spherical robots driven by a pendulum method produce less force for the sphere to move. The problem can be reduced by adding more pendulums inside it, but the space is limited. A wheel fixed to the spherical shell was proposed to minimize this issue. The wheel is driven by a motor and placed on a yoke at the sphere's center [23].

Masses can be used as propulsion by swinging the pendulum or moving the masses straight from the sphere's center to its surface. Slippage does not occur for this method because no mass is in contact with the spherical surface. The amount of energy produced depends on the weight of the mass used. The mobility of this method can be improved by adding mass to the robot. Up to 4 pendulum units are used to move in all axes, increasing the robot's motion torque. However, the increasing number of pendulums will also increase the robot's weight and control method complexity so that it can perform a stable motion.

2.1. Barycenter offset

The barycenter offset driving principle implemented in the spherical robot applied several methods. A pendulum is one of the common ways to change the mass location inside the sphere. K. Landa and A.K. Pilat describe the dynamic of such a system using the Lagrange Eq. (1), where U_1 is the potential energy of the sphere with respect to its centroid, U_2, the potential energy of the pendulum with respect to the sphere's centroid, K_1, the kinetic energy of the sphere, K_2, the kinetic energy of the pendulum, T_1, the rotational energy of the sphere, and T_2 represents the rotational energy of the pendulum [27]. The model assumes that the system is a two rigid-body system with a single degree of freedom between them.

$$L = K_1 + K_2 + T_1 + T_2 - U_1 - U_2 \tag{1}$$

Based on Eq. (1), K. Landa and A.K. Pilat state in [28] that the transfer function that shows the relationship between the sphere inclination angle and pendulum angle as Eq. (2). $L_1(s)$ is the Laplace transform of the sphere inclination angle, $L_2(s)$ is the Laplace transform of the pendulum angle, and T_v is the coefficient of viscous friction.

$$\frac{L_1(s)}{L_2(s)} = \frac{(A_2 - 2A_3)^2 - 2A_4}{(A_1 - 3A_2 + A_3)s^2 - T_v s + 2A_4} \tag{2}$$

https://doi.org/10.31436/iiumej.v25i2.2964

Where

$A_1 = J_1 + J_2 + M_1R^2 + M_2R^2 + M_2l^2$

$A_2 = M_2Rl$

$A_3 = J_2 + M_2l^2$

$A_4 = M_2gl$

$J1$ – moment of inertia of the sphere around its centroid,
$J2$ – moment of inertia of the pendulum around the sphere's centroid,
M_1 – mass of the sphere,
M_2 – mass of the pendulum,
g – acceleration of gravity
R – sphere radius
l – pendulum length

From the transfer function, the position or sphere rotation depends on the moment of inertia of the pendulum. This is valid when the sphere's initial center of mass is aligned with the pendulum's origin. Therefore, the mass distribution inside the sphere must be equal on each side to ensure the pendulum motion can change the sphere orientation.

2.2. Static Equilibrium

An object with mass has inertia that is determined by the mass energy exerted on it [29]. This inertia will keep the object at its original state until external forces are given. The body is in static equilibrium if every material point in a body has the same constant velocity relative to an inertial system [30]Stefan Lindstrom states that a body's static equilibrium occurs when the force-couple system acting on it is a zero system. This means that the force and moment sum of the force-couple systems equals zero, as in Eq. (3).

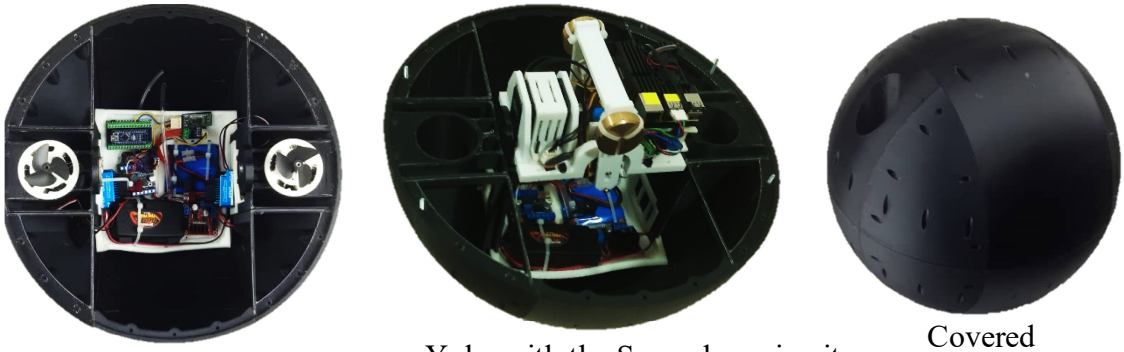
$$\begin{aligned} \sum F_x = 0,\ \sum M_{Ax} = 0, \\ \sum F_y = 0,\ \sum M_{Ay} = 0, \\ \sum F_z = 0,\ \sum M_{Az} = 0, \end{aligned} \tag{3}$$

3. METHODOLOGY

As indicated in Fig. 2A spherical robot was developed to test the proposed driving principle. A complete overview of the mechanical design was discussed in [31]. The proposed spherical robot electronic circuit was located on the yoke to control the terrestrial actuator and main sphere sensory unit. A secondary electronic circuit was used to control the underwater actuator and sensory situated at the bottom of the lower casing.

Two units of inertial measurement unit (IMU) were used to detect the yoke and the sphere motion. All the data was gathered by a raspberry located in the yoke. Additionally, the yoke controls all the actuators located on the yoke, while the secondary circuit controls the actuator for the underwater motion.

https://doi.org/10.31436/iiumej.v25i2.2964



Secondary circuit assembly | Yoke with the Secondary circuit | Covered

Fig. 2. Proposed design (Left: the secondary circuit assembly with the lower cover. Center: Assembly of the secondary circuit with the Yoke. Right: Top cover was installed to the proposed design).

3.1. A proposed terrestrial driving principle

The proposed design rolled on a pitch by changing the yoke angle, while the sphere roll angle was manipulated by swinging the pendulum. The pendulum was placed on top of the yoke, as in Fig. 2 to increase the offset mass inside the sphere. Detailed discussion regarding the proposed driving principle is in [32].

The proposed movement must be tested for its effectiveness in terms of moving, turning, and determining motion velocity with various settings. This information is necessary to obtain the optimum calibration for the robot before developing a suitable control system. Furthermore, experiments must be performed to establish the proposed robot's limitations.

3.2. Experimental Procedures

The experimental procedures began with the initial setup, a step adopted across all experiments to ensure data consistency. Subsequently, a detailed description of the testing methodology is provided.

3.2.1 Initial Setup

The procedure begins with the initialization of the yoke and pendulum angle. To ensure that the robot center is perpendicular to the ground, spirit level was used to measure the position while the yoke angle, θ_Y and pendulum angle, θ_P is tuned to identify the initial θ_{Yi} and θ_{Pi}.

From the pendulum's initial angle θ_{Pi}, the maximum angle to the left and right was identified. To do so, six (6) pieces of weight coin were attached to the pendulum as depicted in Fig. 3. 600g is the maximum mass that can be added to the pendulum due to the limited space. The tuning was done in this setup to ensure a consistent angle can be used for different added mass. From the test, the pendulum can rotate 7^0 to the left and right. Therefore, for 500g to 600g added mass, three pendulum angles, θ_P is tested, which is -7^0 , 0^0, and 7^0.

https://doi.org/10.31436/iiumej.v25i2.2964

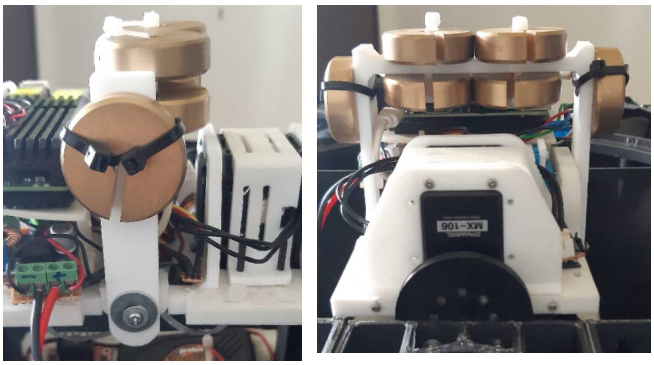


Fig. 3. A total of 600g (6 pieces) added mass attached to the pendulum.

Secondly, the environment was set as depicted in Fig. 4. The total test area is $3.25m^2$ and the same material was used for all the tests done. The start and center lines were marked on the test area to ensure the initial position was consistent.

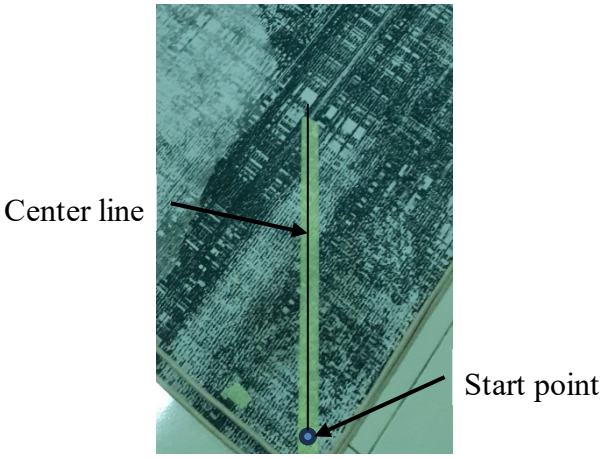


Fig. 4. Environment setup for the experiment.

3.2.2 Performing the Experiment

1. The proposed robot was placed on top of the Start point and aligned with the initial line by measuring the front and back of the robot using a spirit level, as in Fig. 5.

Fig. 5. Setting the initial position of the proposed robot.

2. The pendulum angle was set at 0 degrees with a 500g mass placed on it.

3. The motion was observed to identify the first and second cycles. At the end of each cycle, a mark was made on the floor to represent each completed cycle.

4. The angle of each cycle was measured based on the angle between the center line with a straight line to the cycle point.

5. Distance was measured from the start to the first cycle point, and from the first cycle point to the second. Fig. 6 shows the measured method.

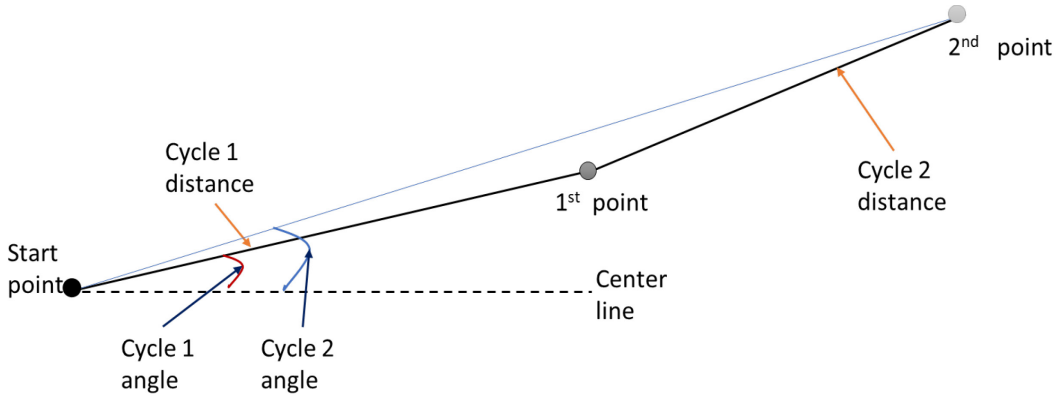


Fig. 6. Method to measure the data.

6. Step 1 until 5 was repeated 5 times.

7. Step 1 until 6 was repeated with pendulum angles of +7 and -7 degrees.

8. Step 1 until 7 was repeated with a 600g pendulum mass.

9. By installing the propeller to the proposed robot, steps 1 until 6 were repeated to observe the motion difference when the proposed robot mass increased.

The procedure was done to prove that the proposed driving principle can drive the proposed robot from one point to the other. Furthermore, the experiment was able to identify the maximum turning angle of the proposed robot. Fig. 7 represents the flowchart of the experiment procedure for a 500g mass located at the pendulum. The procedure was repeated with a 600g mass, followed by the installation of propellers with a 600g mass.

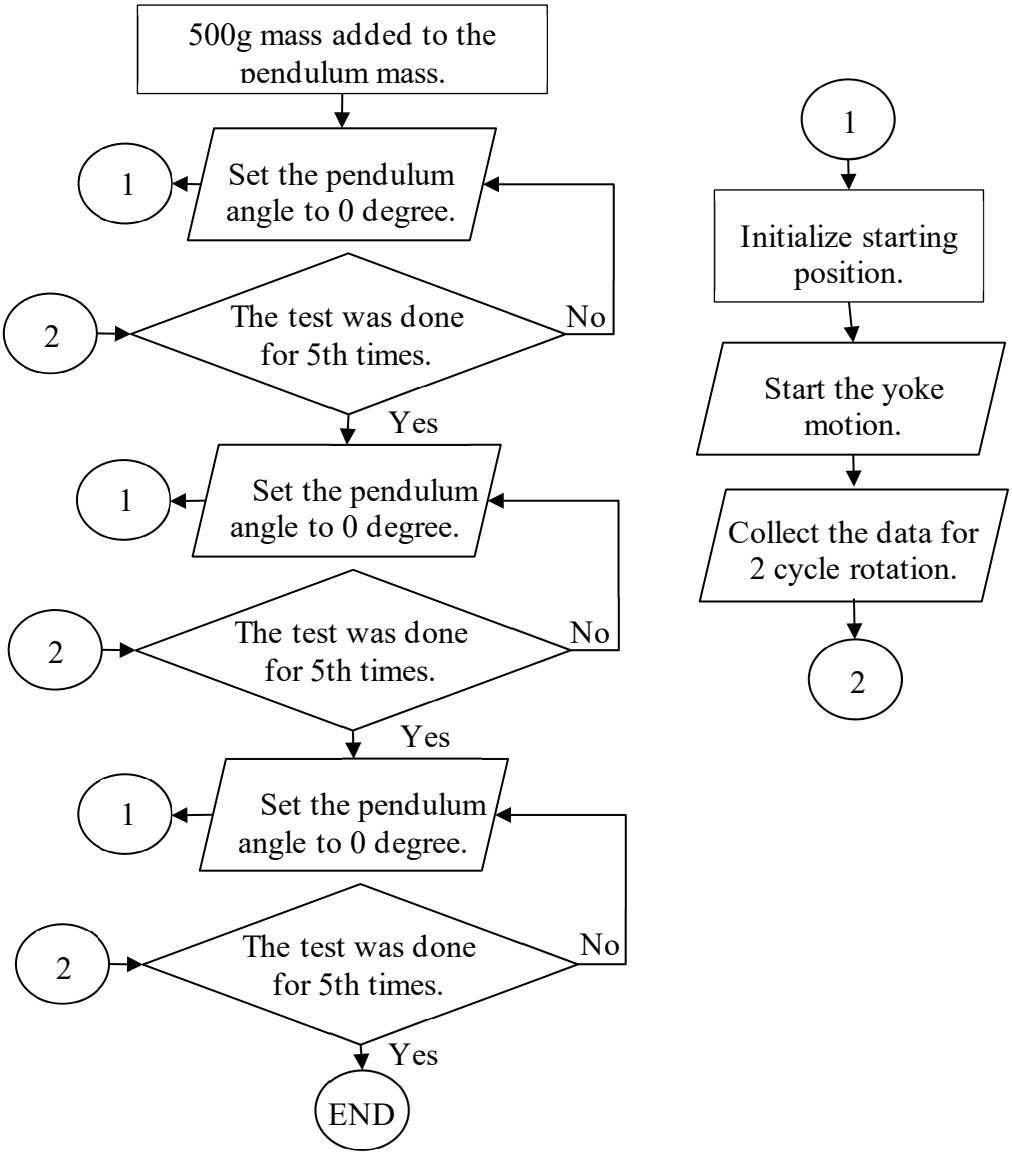


Fig. 7. Flowchart of the experiment procedure for 500g mass. The same step was used for 600g mass.

4. RESULTS AND DISCUSSIONS

The proposed spherical amphibian robot can travel on a dry surface with a minimum $30.84°/s$ and a maximum of $40.75°/s$. The average roll velocity for each test done is in Table 1.

Table 1: Average velocity of the proposed system when traveling on land.

Mass	Average Velocity (degree/second)	
	Cycle 1	Cycle 2
500g	32.97	39.21
600g	36.81	40.75
600g with propeller	30.84	32.79

Analyzing the velocities of the first and second cycles reveals that the system achieves its highest speed when bearing a weight of 600g. Notably, the second cycle consistently exhibits

https://doi.org/10.31436/iiumej.v25i2.2964

greater speed across all trials than the first. This phenomenon is attributed to the momentum acquired during the initial cycle being transferred to the subsequent one. In the initial cycle, the system contends with inertia, thereby impeding acceleration. Notably, the system experiences increased inertia when engaging the underwater circuit, resulting in reduced velocity.

Fig. 8 depicts the roll motion data for three distinct pendulum locations, with the top figure representing the roll angle of 7° to the left, the middle figure representing 0 degrees, and the bottom figure representing 7° to the right. The y-axis is the angle data provided by the IMU sensor while the step in the x-axis represents the time data was taken (loop time).

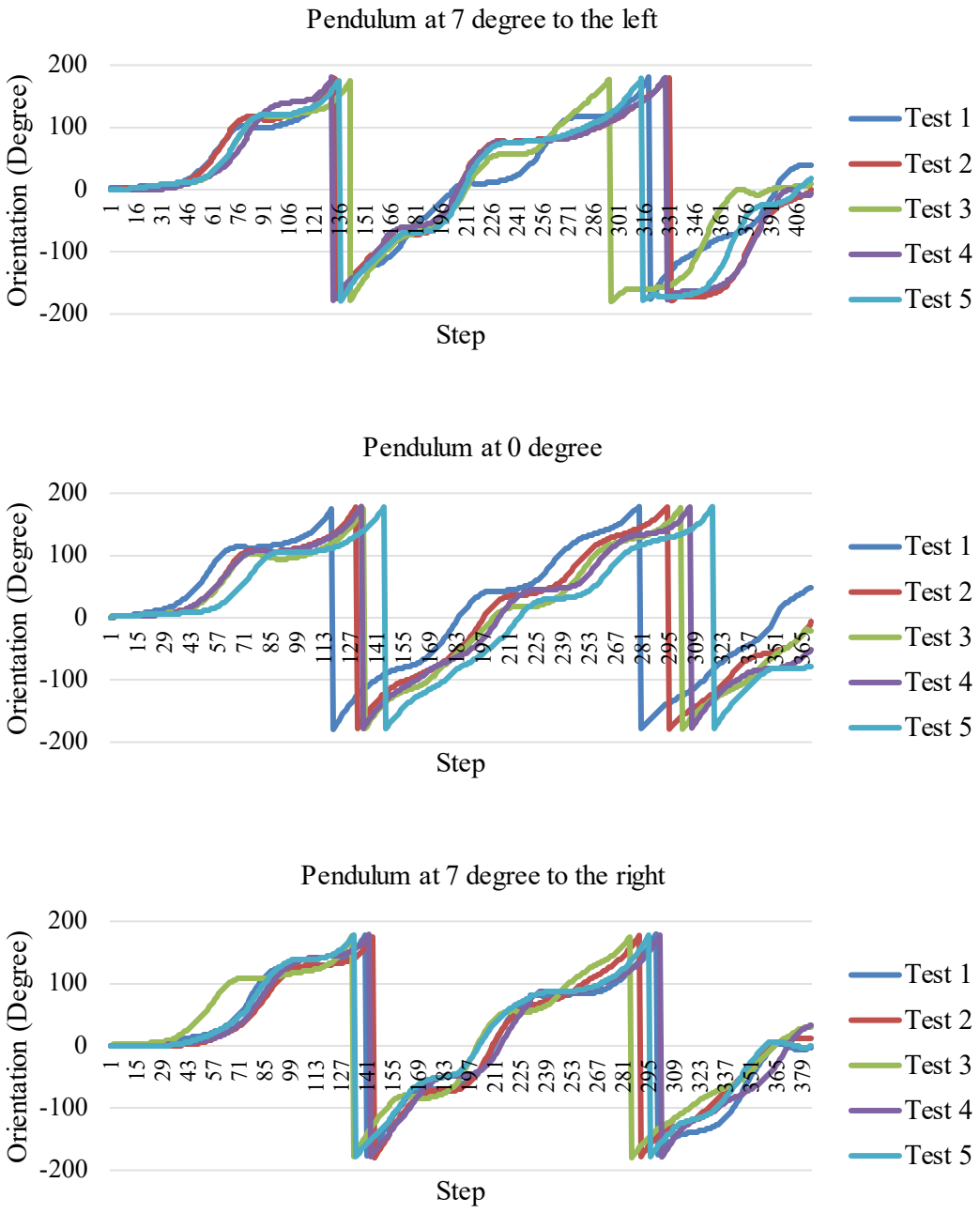


Fig. 8. Roll motion of the proposed system.

https://doi.org/10.31436/iiumej.v25i2.2964

Based on the roll angle obtained, the roll movement pattern was the same for all three pendulum locations. When the proposed system was in motion, the change in velocity was caused by the presence of a secondary circuit in the proposed system, causing the change in CoM to become uneven. This caused the system to decelerate temporarily until the yoke could overcome the secondary circuit mass. Fig. 9 depicts the relationship between angular velocity and the rolling motion of the proposed system, where the top graph shows the system's rolling together with the yoke motion (black line). In contrast, the graph at the bottom depicts the angular velocity profile.

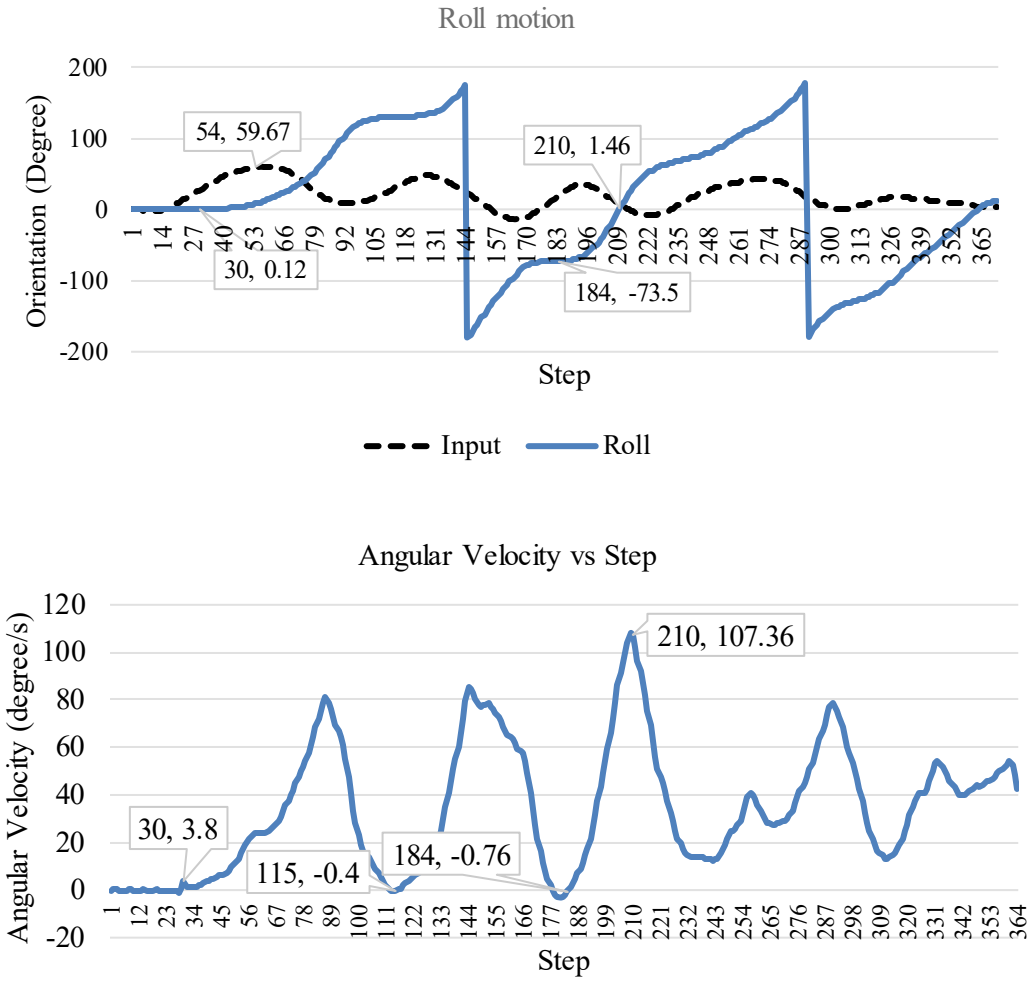


Fig. 9. Roll and angular velocity of the proposed system.

Fig. 10 depicts the pitch angle of the proposed system gathered from the secondary circuit's sensor readings. Each graph, like Fig. 8, illustrates a distinct position of the pendulum.

https://doi.org/10.31436/iiumej.v25i2.2964

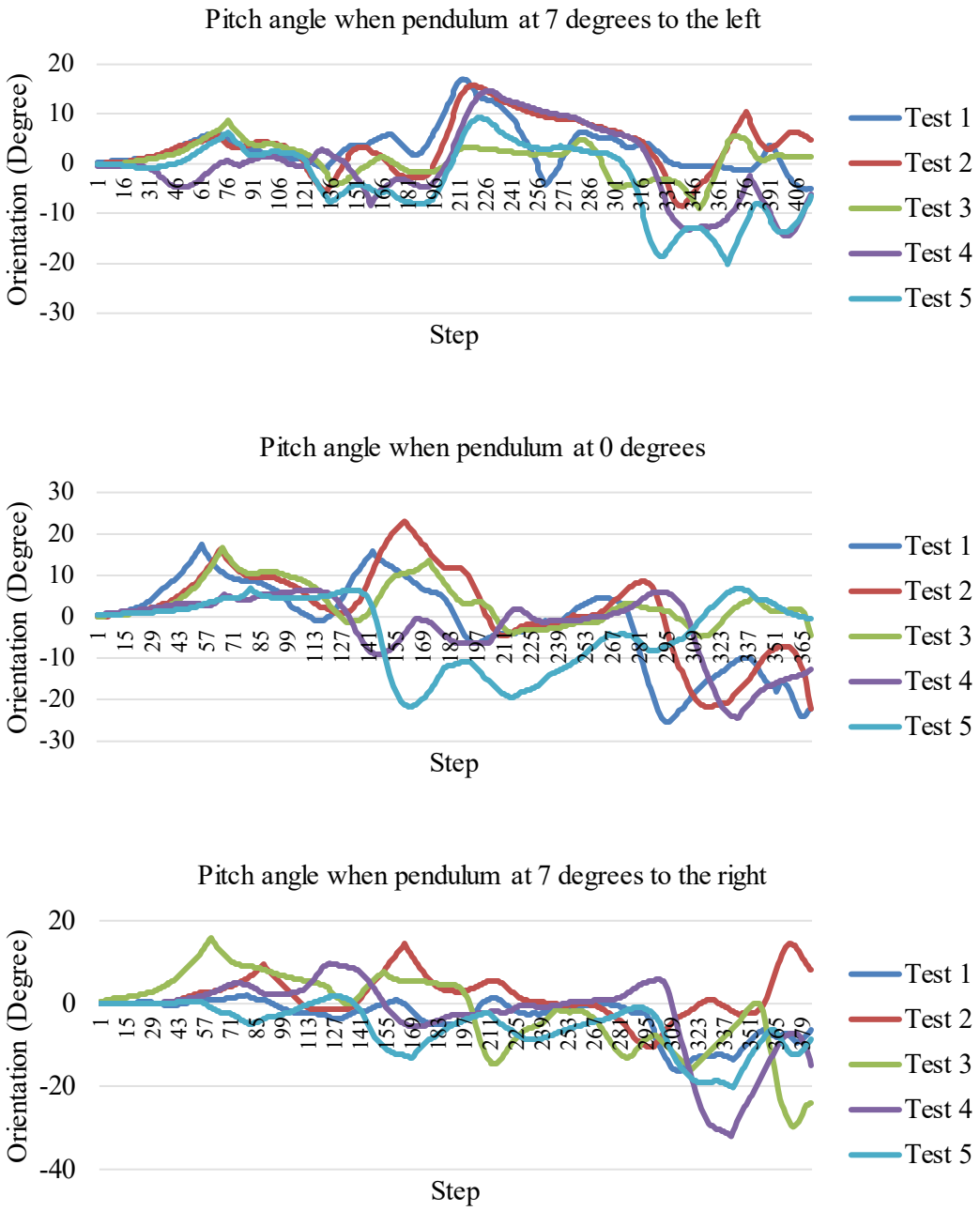


Fig. 10. Pitch motion of the proposed system.

Based on the pitch angle data, the suggested system consistently demonstrates lateral movement to the left and right. However, this behavior is not uniformly observed across all experiments for the three pendulum positions. The system initially exhibits a stable pattern that gradually becomes unstable over time. This transition can be attributed to the absence of a controlling mechanism throughout the testing phase, allowing the mounting momentum to impact the system's motion progressively. The system's average turning motion is depicted in Fig. 11.

https://doi.org/10.31436/iiumej.v25i2.2964

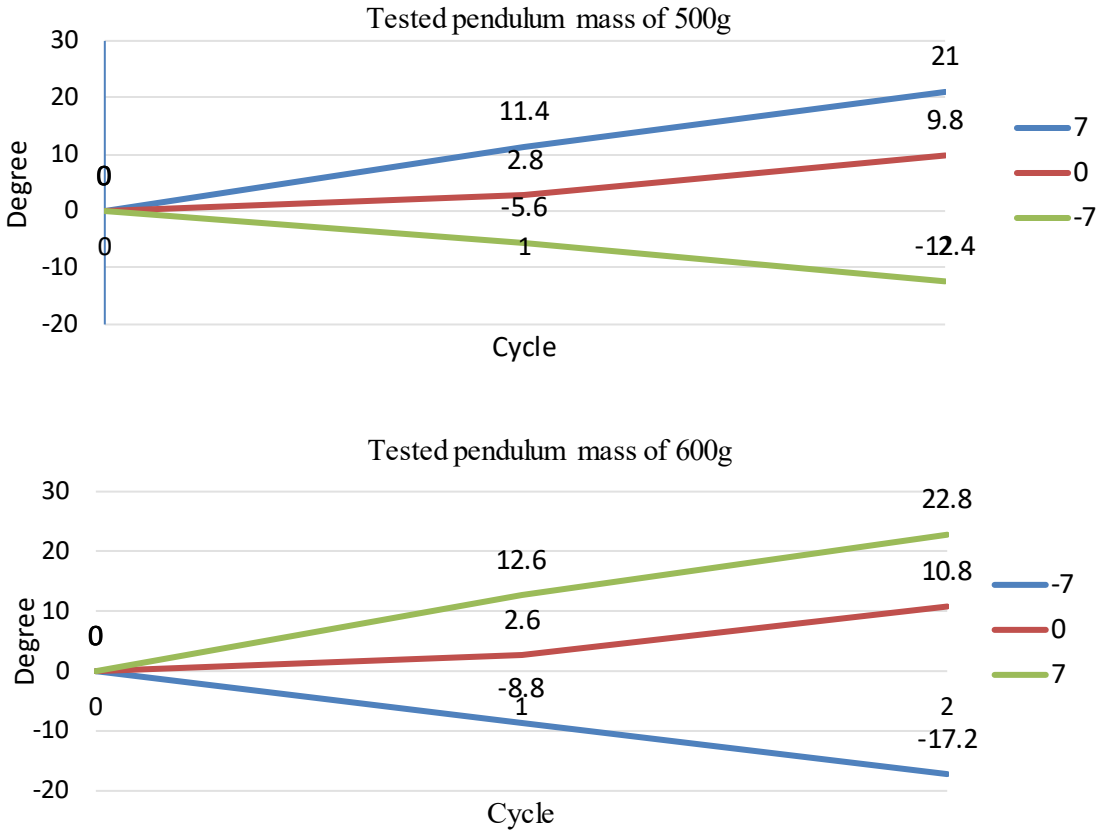


Fig. 11. Turning angle of the proposed system.

Due to the constraint of the pendulum movement, which can only travel 14 degrees, the proposed system requires a significant distance to turn. Based on the results, the maximum turning for the first cycle is 12.6 degrees, and the maximum turn for the second cycle is 22.8 degrees when 600g of weight is applied. On average, adding weights can increase the degree of rotation by 1.53 degrees in the first cycle and 2.53 degrees in the second cycle.

Fig. 12 depicts the distance traveled by the system during the first and second cycles with weights of 500g (top) and 600g (bottom). The longest distance is shown in the test that placed the pendulum in the middle (0 degrees), which is 212.64 cm, while changing the pendulum's angle minimizes the system's travel distance.

Fig. 13 represents the rolling motion of the proposed system (OUTPUT) when a 400g mass is added. While the yoke (INPUT) rolled to 90^0, the system is still in its position because the force created by the yoke motion is insufficient to encounter the inertial moment. Compared to the graph in Fig. 9 (top), the system started to roll before the yoke rolled to 90^0. Therefore, the mass needed to flee from the static equilibrium must be greater than 1113.8g, whereas 713.8g came from the yoke and pendulum, with an additional 400g added to the pendulum. The limitation of space inside the sphere limits the number of weight coins that can be placed at the pendulum. The increasing number of weight coins will decrease the pendulum rotation angle.

https://doi.org/10.31436/iiumej.v25i2.2964

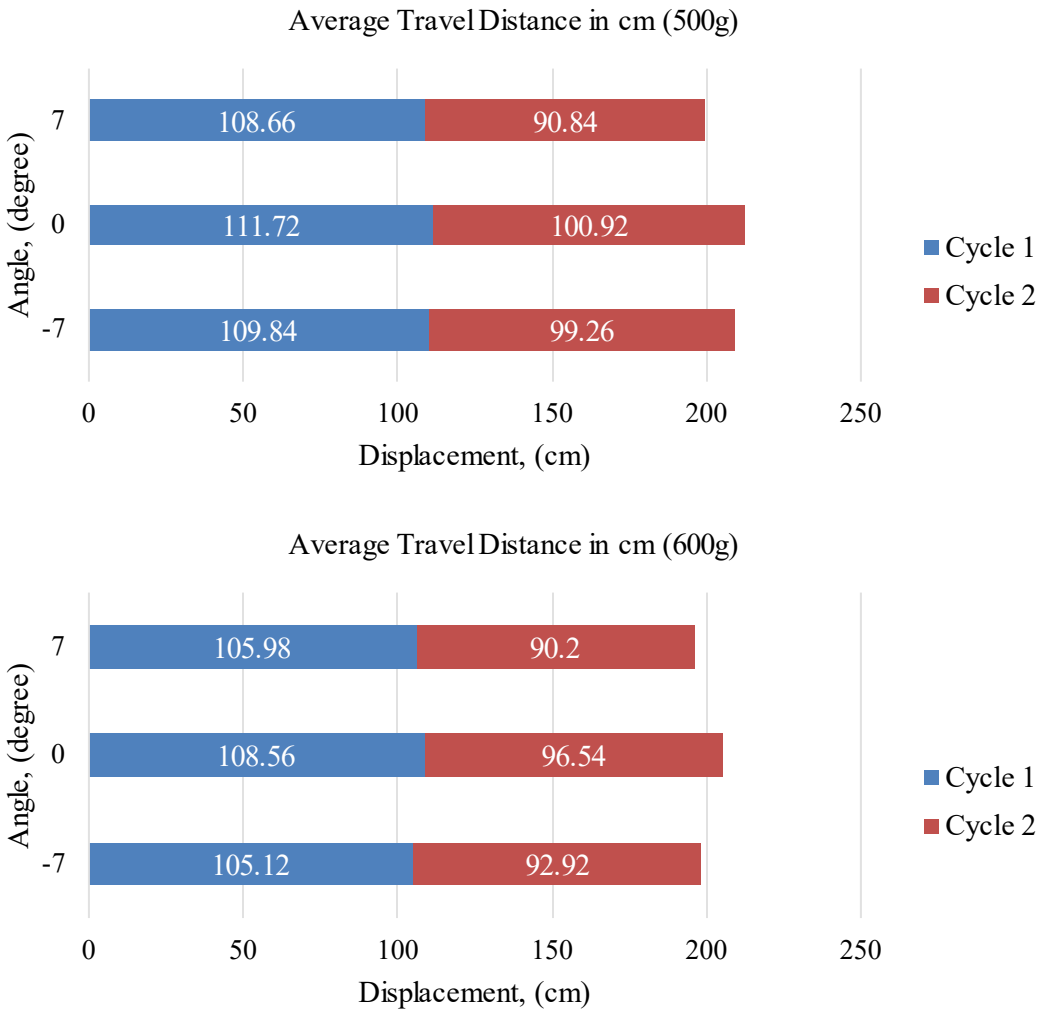


Fig. 12. Displacement of the proposed system.

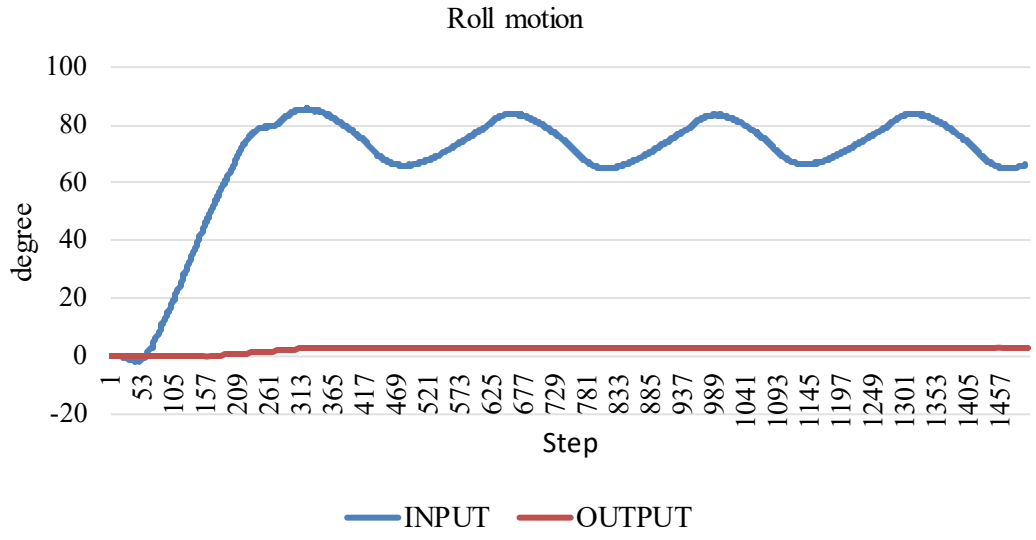


Fig. 13. Roll motion of yoke and the proposed system.

5. CONCLUSIONS

The proposed spherical amphibian robot can travel on a dry surfaces with a maximum velocity of 40.75 degree/s while carrying 600g of mass. It needs more than 400g of additional mass to encounter inertia to move. From the experiment, the increasing mass of the pendulum increases the velocity and turning angle of the proposed system. The propeller with its motor and drivers (549g) was excluded from the experiment so that the weight may be varied to two values, 500g, and 600g. The robot's roll motion remains consistent across all experiments, while its yaw motion varies. This is because the study focuses on overall system speed rather than specific control techniques for yaw. This paves the way for future control strategy decisions. Smaller actuators could provide space for weighted coin placement, and different materials may enhance the pendulum's weight.

ACKNOWLEDGEMENT

We wish to express our gratitude to the honorable Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education (MOHE) for sponsoring this work under grant no. PJP/2022/FKE/S01855. Special appreciation and gratitude especially for Universiti Teknologi Malaysia (UTM) KL Campus, Malaysia-Japan International Institute of Technology (MJIIT), Fakulti Teknologi dan Kejuruteraan Elektrik (FTKE) and Underwater Technology Research Group (UTeRG), Center for Robotics and Industrial Automation (CERIA) and Ministry of Higher Education (KPT) for supporting this research.

REFERENCES

[1] L. Shi, S. Su, S. Guo, K. Tang, S. Pan, Y. he, H. Xing, Z. Chen, and P. Guo. (2017) A fuzzy PID control method for the underwater spherical robot. 2017 IEEE International Conference on Mechatronics and Automation (ICMA)., 626–631.

[2] O. Lang (2010). The dangers of mining around the world. *BBC News*, Retrieved from https://www.bbc.com/news/world-latin-america-11533349

[3] Y. Li, M. Yang, H. Sun, Z. Liu, and Y. Zhang. (2018) A Novel Amphibious Spherical Robot Equipped with Flywheel, Pendulum, and Propeller. J. Intell. Robot. Syst., 89(3–4):485–501.

[4] S. Radu (2019). The Top 5 Deadliest Disasters in 2019. U.S.News, Retrieved from https://www.usnews.com/news/best-countries/slideshows/5-of-the-deadliest-natural-disasters-in-2019?slide=6

[5] L. DeMunno (2018). 9 natural disasters that took the lives of hundreds of thousands. Business Insider US, Retrieved from https://www.businessinsider.my/worst-natural-disasters-2018-7/?r=US&IR=T

[6] S. Guo, Y. Liu, L. Shi, P. Guo, H. Xing, X. Hou, Z. Chen, S. Su, and H. Liu. (2018) Binocular Camera-based a Docking System for an Amphibious Spherical Robot. 2018 IEEE International Conference on Mechatronics and Automation (ICMA)., 1621–1626.

[7] G. S. Soh, S. Foong, and K. Wood. (2017) De-coupled dynamics control of a spherical rolling robot for waypoint navigation. 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)., 562–567.

[8] A. Nakashima, S. Maruo, R. Nagai, and N. Sakamoto. (2018) 2-Dimensional Dynamical Modeling and Control of Spherical Robot Driven by Inner Car. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO).,1846–1851.

[9] J. Kim, T. Kim, and S. C. Yu. (2018) Conceptual Design of a Spherical Underwater Vehicle Equipped with Vertically Rotatable Thruster Units. 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV)., 1–4.

[10] M. T. Andani, S. Shahmiri, H. Pourgharibshahi, K. Yousefpour, and M. H. Imani. (2018) Fuzzy-Based Sliding Mode Control and Sliding Mode Control of a Spherical Robot. IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society., 2534–2539.

[11] J. V. Raj, K. Lashmi, and R. Shanmughasundaram. (2018) Design of a spherical robot with improved stability for planetary exploration. J. Adv. Res. Dyn. Control Syst., 10(3):1001–1005.

[12] S. Guo, Y. He, L. Shi, S. Pan, R. Xiao, K. Tang, and P. Guo. (2018) Modeling and experimental evaluation of an improved amphibious robot with compact structure. Robot. Comput. Integr. Manuf., 51:37–52.

[13] L. Geng, Z. Hu, Y. Lin, R. Yi, and C. Wang. (2015) A new concept spherical underwater robot with high mobility. 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)., 887–890.

[14] L. Jia, Z. Hu, L. Geng, Y. Yang, and C. Wang. (2016) The concept design of a mobile amphibious spherical robot for underwater operation. 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)., 411–415.

[15] H. Xing, S. Guo, L. Shi, X. Huo, S. Su, Z. Chen, Y. Liu, and H. Liu. (2018) Performance Evaluation of a Multi-Vectored Water-Jet Propellers Device for an Amphibious Spherical Robot. 2018 IEEE International Conference on Mechatronics and Automation (ICMA)., 1591–1596.

[16] X. Hou, S. Guo, L. Shi, H. Xing, Y. Liu, H. Liu, Y. Hu, D. Xia, and. Z. Li. (2019) Hydrodynamic analysis-based modeling and experimental verification of a new water-jet thruster for an amphibious spherical robot. Sensors, 19 (2): 259.

[17] H. Xing, S. Guo, L. Shi, X. Hou, Y. Liu, and H. Liu. (2019) Design, modeling and experimental evaluation of a legged, multi-vectored water-jet composite driving mechanism for an amphibious spherical robot. Microsyst. Technol., 1–13.

[18] B. Alexey, K. Alexander, K. Yury, and K. Anton. (2019) Stabilization of the motion of a spherical robot using feedbacks. Appl. Math. Model., 69: 583–592.

[19] B. B. Dey and M. Jenkin. (2018) Spherico: rapid prototyping a spherical robot. 2018 IEEE International Conference on Information and Automation (ICIA)., 957–962.

[20] Y. Ping, S. Hanxu, Q. Zhongjiang, and C. Jiazhen. (2016) Design and Motion Control of a Spherical robot with Stereovision. 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)., 1276–1282.

[21] M. Taheri Andani, Z. Ramezani, S. Moazami, J. Cao, M. M. Arefi, and H. Zargarzadeh. (2018) Observer-Based Sliding Mode Control for Path Tracking of a Spherical Robot. Complexity, 2018:1-15.

[22] Q. Zhan. (2015) Motion planning of a spherical mobile robot. Motion and Operation Planning of Robotic Systems. Springer., 361–381.

[23] S. Gajbhiye and R. N. Banavar. (2016) Geometric modeling and local controllability of a spherical mobile robot actuated by an internal pendulum. Int. J. Robust Nonlinear Control, 26(11):2436–2454.

[24] M. Roozegar, M. Ayati, and M. J. Mahjoob. (2017) Mathematical modelling and control of a nonholonomic spherical robot on a variable-slope inclined plane using terminal sliding mode control. Nonlinear Dyn., 90(2):971–981.

[25] N. Muraleedharan, D. S. Cohen, and D. R. Isenberg. (2016) Omnidirectional locomotion control of a pendulum driven spherical robot. SoutheastCon 2016.,1–6.

[26] S. Asiri, F. Khademianzadeh, A. Monadjemi, and P. Moallem. (2019) The Design and Development of a Dynamic Model of a Low-Power Consumption, Two-Pendulum Spherical Robot. IEEE/ASME Trans. Mechatronics., 24(5): 2406–2415.

[27] K. Landa and A. K. Pilat. (2015) Design and start-up of spherical robot with internal pendulum. 2015 10th International Workshop on Robot Motion and Control (RoMoCo)., 27–32.

[28] K. Landa and A. K. Pilat. (2016) Design of a controller for stabilization of spherical robot's sideway oscillations. 2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR).,484–489.

[29] I. Ciufolini and J. A. Wheeler. (1995) Gravitation and inertia, Princeton university press., 101.

[30] S. Lindström. (2019) Lectures on engineering mechanics: statics and dynamics.

[31] M. B. Bahar, S. S. Abdullah, M. S. M. Aras, N. F. Ibrahimjee, F. N. Zohedi, M. H. Harun, A. A. Yusof, and M. K. Aripin. (2022) A Spherical Amphibian Underwater Robot: Preliminary Mechanical Design. 2022 IEEE 9th International Conference on Underwater System Technology: Theory and Applications (USYS)., 1–6.

[32] M. B. Bahar, S. S. Abdullah, M. S. M. Aras, and F. N. Zohedi. (2022) Spherical Amphibian Robot Design with Novel Driving Principle. Proceedings of the 12th National Technical Seminar on Unmanned System Technology 2020., 69–77.

https://doi.org/10.31436/iiumej.v25i2.3017

EXPLORING THE EFFICACY OF PIEZOELECTRIC-BASED SENSORY SYSTEMS FOR HEART RATE MONITORING IN DIFFERENTIATING STRESS VS RELAX CONDITIONS

MUHAMMAD HAFIZ ZULHILMI ZULKIFLI [1], FARAHIYAH JASNI [1*], AIMI SHAZWANI GHAZALI [1], NOR HIDAYATI DIYANA NORDIN[1], SHAHRUL NA'IM SIDEK[1], SYAMIMI MARDIAH SHAHARUM[2], FADHLAN HAFIZHELMI KAMARU ZAMAN[3]

[1]Department of Mechatronics, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[2]Fakulti Teknologi Kejuruteraan Elektrik Dan Elektronik, Universiti Malaysia Pahang, Pahang, Malaysia
3School of Electrical Engineering, Universiti Teknologi MARA, Selangor, Malaysia.

**Corresponding author: farahiyahjasni@iium.edu.my*

(Received:14 September 2023, Accepted: 14 February 2024, Published online: 15 July 2024)

ABSTRACT: Stress has diverse effects on human physiological reactions, and one such effect is on heart rate (HR). The established methods to acquire HR is by electrocardiogram (ECG) and photoplethysmogram (PPG). ECG electrodes need to be placed on the chest, which can cause inconvenience and is not practical in daily life, while PPG signals are known to contain more noise than ECG. Thus, this work aims to investigate the efficacy of a piezoelectric-based sensory system in measuring HR and using the signal to differentiate stressed and relaxed conditions by means of statistical analysis. Two activities were conducted to achieve the goal. The first experiment involved collecting and analysing piezoelectric signals to measure the pulse rate (*bpm*) and compare this with the HR from PPG. For the second experiment, the piezoelectric-based HR was calculated from 20 subjects (male and female, age ranging between 20 and 25) in relaxed and stressed conditions. The stress condition was triggered using two stressors: the Stroop Colour Word Test and the Digit Span Test. Statistical analyses reveal a strong positive correlation between piezoelectric-based heart rate (HR) and oximeter readings ($r(12) = 0.993$, $p < 0.001$), despite the fact that the values are not precisely identical. In addition, the findings also indicate that there are significant effects by the mental states (stressed and relaxed) on the piezoelectric-based HR readings ($p<0.05$). Employing a within-subject design condition, the results further illustrated that piezoelectric readings are elevated during stressed conditions (*Mean*±*SD* = 72.395±0.097) and diminished during relaxed conditions (*Mean*±*SD* = 71.615±0.126). Therefore, the suggested piezoelectric-based sensory system has been validated as an effective means of categorizing stress and relaxation based on heart rate signals.

ABSTRAK: Tekanan mempunyai pelbagai kesan terhadap reaksi fisiologi manusia dan satu daripadanya adalah kadar denyut jantung (HR). Kaedah biasa bagi mengetahui HR adalah melalui elektrokardiogram (ECG) dan fotofetismogram (PPG). Elektrod ECG perlu dipasang pada bahagian dada di mana boleh menyebabkan ketidakselesaan dan tidak praktikal dalam kehidupan seharian, manakala PPG mengandungi lebih banyak gangguan isyarat berbanding ECG. Oleh itu, kajian ini bertujuan mengkaji kecekapan sistem deria berasaskan piezoeletrik bagi membaca HR dan menggunakan isyarat ini bagi membezakan keadaan tertekan atau tenang melalui kaedah analisa statistik. Untuk mencapai objektif, dua aktiviti telah dijalankan. Pertama eksperimen melibatkan pengumpulan dan analisis isyarat piezoeletrik bagi mengira kadar nadi (bpm) dan membandingkannya bacaan HR daripada PPG. Eksperimen kedua, HR berasaskan piezoelektrik dikira daripada 20 subjek (lelaki dan perempuan, berumur antara 20

dan 25) dalam keadaan tenang dan tertekan. Keadaan tertekan dibuat melalui dua kaedah: Stroop Color Word Test dan Digit Span Test. Analisis statistik mendedahkan hubungkait yang kuat antara kadar denyut jantung (HR) berasaskan piezoeletrik dan bacaan oksimeter ($r(12) = 0.993$, $p < 0.001$), walaupun bacaan tidak sama sepenuhnya. Tambahan, penemuan ini menunjukkan terdapat kesan penting terhadap tahap mental (tertekan dan tenang) pada bacaan denyut jantung (HR) berdasarkan piezoeletrik ($p<0.05$). Dengan mengaplikasi kaedah within-subject, hasil menunjukkan bacaan piezoeletrik meningkat dalam keadaan tertekan (Mean±SD = 72.395±0.097) dan kurang dalam keadaan tenang (Mean±SD = 71.615±0.126). Oleh itu, system deria berasaskan piezoeletrik yang dicadangkan telah diuji berkesan untuk mengenal pasti keadaan tertekan dan tenang berdasarkan isyarat denyut jantung.

KEYWORDS: *Piezoelectric Sensor, Heart Rate, Stress Detection*

1. INTRODUCTION

The issue of mental health has become a common and significant problem in modern society. One of the major contributors to mental illness is stress. Most of the time, people are unaware of being under stress, for example, when they are occupied with deadlines for certain projects or works. This deleterious effect of stress is usually defined as the short-term sympathetic nervous system activation caused by cognitive stressors [1]. However, long-term exposure to stress can be chronic, and people may not be aware of its existence. Chronic stress can result in severe problems, including increases in blood pressure, weakening of the body's immune system, muscle pain, increased risk of heart attack and stroke, depression, etc. [2].

Numerous methods and techniques are being introduced today to detect stress. A popular method uses human physiological signals to identify stress presence as it is proven that humans will be physiologically affected under stress [3]. The most frequently used parameters for stress detection are Blood Pressure (BP) [4], Respiration Rate (RR) [5], and Heart Rate (HR) [4], [6], [7], [8]. Others include muscle activity measurement [9], temperature measurement [10], and electrodermal activity [11]. Many studies report that heart rate increases significantly during states of stress, whereas a few others did not show any significant change [4], [6], [7], [8]. This method is the most straightforward and the most used, with the vast majority being a reliable measure for the arousal part of stress [12].

Among the popular sensors to acquire HR data used are the electrocardiogram (ECG) [13], [14], [15], and the photoplethysmogram (PPG) [16], [17]. ECG is a method where electrodes need to be placed on specific locations on the body (i.e. chest, forearms, and lower leg) to capture the electrical signals of the heart. ECG requires tedious preparation to ensure the quality of the data collected as the electrodes are placed onto the skin. The area of the skin will need to be prepared well, such as to shave excessive hair that might reduce the contact between the electrodes and skin, which consequently may cause discomfort to the subjects [18].

Meanwhile PPG employs the optical method for the identification of volumetric alterations in peripheral blood circulation. A significant challenge associated with the utilization of PPG-based monitoring techniques is their limited accuracy in tracking PPG signals during everyday activities and mild physical exertion due to the susceptibility of PPG signals to Motion Artifacts (MA) induced by hand movements. In addition, external factors such as environmental noise can further impact the acquisition of PPG signals, consequently affecting the precision of heart rate estimation [19].

Both ECG and PPG are explicitly measuring the physiological signal of a body to get the HR reading. ECG electrodes need to be placed on the chest, which can cause inconvenience and is not practical in daily life, while PPG signals are known to contain more noise as

compared to ECG. Another alternative that is gaining popularity is a mechanical-based sensory system, such as piezoelectric sensors.

Piezoelectric is a transducer that can assess electromechanical ability. It converts mechanical energy into electrical energy and vice versa [20]. Manipulating its electromechanical features, piezoelectric transducers are also exploited to measure one's HR. [21] analysed the efficacy of an in-ear piezoelectric-based sensory system to monitor HR and obtained promising results (i.e., 97.17% accuracy). Meanwhile, a study in [22] utilized piezoelectric ceramic sensors to measure HR and the findings showed a good accuracy as compared to an XD-58C HR pulse sensor. Another study reported in [23] measured HR via chest vibrations. These studies proved that piezoelectric sensors are reliable to be used as HR measurement. However, to the author's knowledge, none of the research conducted on investigating the HR captured by piezoelectric-based sensory systems was for stress detection application.

Therefore, this paper aims to investigate the efficacy of a piezoelectric-based sensory system in differentiating between stressed and relaxed conditions using the HR signal. The HR measurement was acquired using a piezoelectric-based sensory system, which was placed at the wrist of healthy young adult subjects. State Trait Anxiety Inventory (STAI) and psychological reactance questionnaire were used to benchmark the mental state of the subjects while performing the experiment. The data was then analysed for its consistency across the subjects and the correlation between the piezoelectric-based sensory system's data and mental health conditions (i.e., relax vs stress), which was attained via mental health score from the questionnaire.

2. METHODOLOGY

2.1. Piezoelectric-based sensory system data verification

In this study, a PVDF Piezoelectric Film with a length of 4cm (LDT1-028K) manufactured by TE CONNECTIVITY SENSOR Co., Ltd was used [24] because it has higher flexibility, higher sensitivity, and higher dynamic range. Conceptually, the heart rate can be measured by a piezoelectric sensory system by detecting the impulse that is caused by the expansion and contraction of the blood vessels near the surface of the skin. The mechanical pounding sensed by the piezoelectric sensor is then converted into an electrical signal and analysed to measure the HR [25]. In this study, the voltage signal from the sensory device was acquired using Arduino Uno.

The piezoelectric sensor measurement shown in Fig. 1 was compared with a PPG-based device (i.e., oximeter) to verify the reliability of the sensory system developed. Fig. 2 shows the experiment setup to check the reliability of the sensor. The piezoelectric sensory device is worn on the wrist, while the oximeter is worn on the same hand's finger. Both devices are worn on the non-dominant hand.

The voltage signal from the piezoelectric (Fig. 3) is acquired using the Arduino analog input and used in calculating the HR measurement using the formula in Eq. 1:

$$Heart\ Rate\ (in\ bpm) = \frac{60}{Time\ interval\ between\ two\ peaks\ (in\ seconds)} \quad (1)$$

https://doi.org/10.31436/iiumej.v25i2.3017

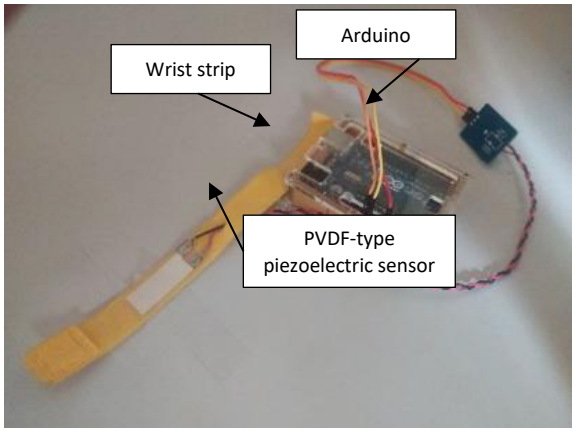


Figure 1. Piezoelectric-based HR measurement device used in this study.

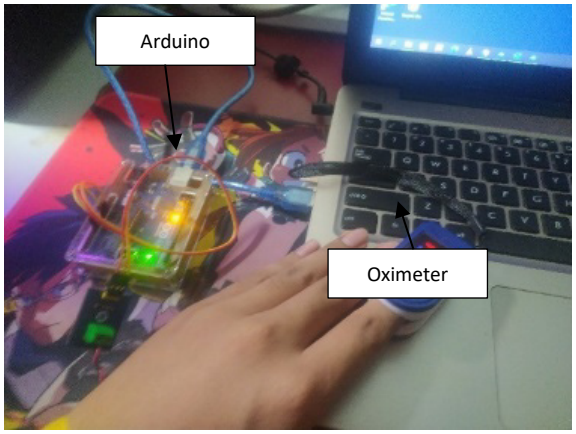


Figure 2. Set-ups to test the reliability of the sensors

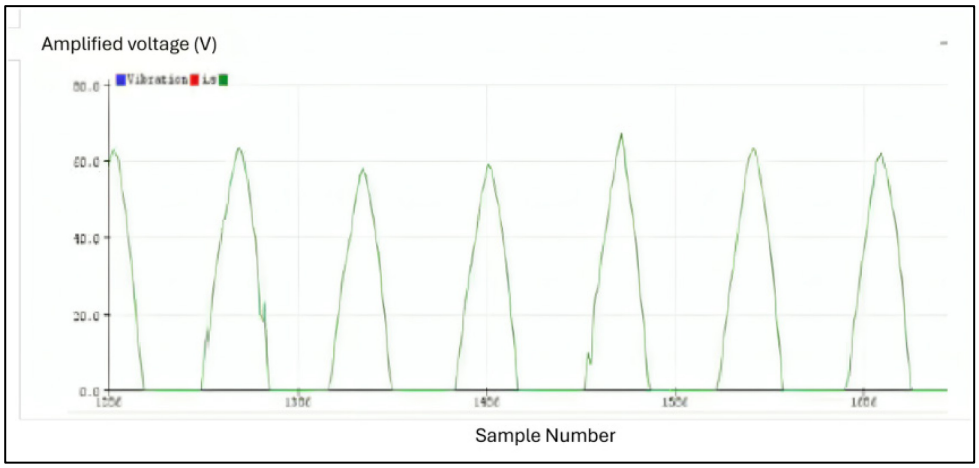


Figure 3. Example of the voltage signal produced by the piezoelectric sensor to measure pulse.

Since the aim of this experiment is to benchmark the developed sensory system with the established PPG-based sensor (i.e., oximeter), the HR readings were recorded from both sensors during resting and active conditions from two subjects. During the active state, the subject was asked to perform jump roping for one minute, and their heart rate was recorded using the two devices mentioned above. At the same time, for rest condition, the measurement was taken while the subjects were in a relaxed sitting position. Six measurements, each for relaxed and active conditions, respectively, made a total of twelve samples for relaxed and another twelve samples for active.

2.2. Piezoelectric-based sensory system efficacy in stress detection.

This section explains the procedures taken to investigate the developed sensory system efficacy in differentiating mental states (i.e., relaxation vs stress).

https://doi.org/10.31436/iiumej.v25i2.3017

2.2.1. Experiment Setup

For this work, 20 subjects were recruited, consisting of 10 males and 10 females, ages ranging from 20 to 25. All subjects were students at IIUM and physically healthy. Their participation in this study was voluntary, and consent forms were given and signed before starting the experiment. As this investigation primarily aimed to evaluate the capability of the implemented sensory system in distinguishing between stressed and relaxed mental states, the inclusion of diverse age groups and other variables was deemed unnecessary at this stage. Therefore, this study's selected cohort of subjects was regarded as adequate for achieving the specified objective.

The experiment was conducted in a controlled environment without outside interference. As illustrated in Fig. 4, it was conducted in a room that was covered by a black curtain to isolate the subject from others in the Biomechatronics Research Lab, IIUM.

White light was selected for the room to prevent the light from affecting the subject's mental conditions. Furthermore, the room was equipped with air conditioning to ensure the comfort of the subjects throughout the duration of the experiment. This research work has obtained ethics approval from the IIUM Research Ethics Committee (IREC 2022-044).

Figure 4. Setup of the experiment

2.2.2. Experiment Procedure

The experimental setups are designed to repeatedly place the subjects in states of emotional and cognitive discomfort. This is done to compare the differences in piezoelectric readings when the subjects are stressed or relaxed.

The data collection process for the mental condition involved a five-phase experiment (i.e., stress or relax). The phases are known as “First Relax Phase” (R1), “First Stress Phase” (S1), “Second Relax Phase” (R2), “Second Stress Phase” (S2), and “Last Relax Phase” (R3).

During the experiment, the subjects were placed in a quiet environment without any outside distractions. They were instructed to wear headphones and a piezoelectric sensor on their wrists. The music "Weightless" by the Macaroni Union was selected for every relax phase based on a research discovery [26] that indicated that it might lower anxiety and encourage

https://doi.org/10.31436/iiumej.v25i2.3017

relaxation. This procedure was critical for determining each subject's baseline since physiological measurements show a wide range of variation across all subjects [11].

In the stress phases, two stressors were used to induce stress among the subjects, namely the Stroop Colour Word Test (SCWT) and Digit Span Test (DST). The SCWT requires the subjects to concentrate between the words and the colour of the words while the DST requires the subjects to memorize a number or digit's arrangement and reproduce the same sequences back. During this phase, the subjects were requested to listen to a piece of hardcore background music throughout their experiments. Loud music can induce stress in humans [27]. Each subject was equipped with noise-canceling headphones to make sure they were fully immersed in the music provided during the experiment.

Each of the phases took five minutes duration. In total, each subject spent approximately 25 minutes to complete the whole experiment procedure and data was taken for only one cycle for each subject. The HR was calculated for every 30s for each subject. Therefore, the accumulated measurement for all 20 subjects comprised 600 for the relaxed condition and 400 for the stressed condition. For every completed phase, the subjects were asked to fill in the questionnaire detailed in Section 2.2.3.

2.2.3. Questionnaire

For the mental health score, subjects were asked to fill in questionnaires consisting of STAI (State-Trait Anxiety Inventory) [28] and psychological reactance [29] questions at the end of each phase. Both questionnaires were the standard quantitative measurement that is used to identify traits and states of anxiety [30].

2.3. Statistical Package for Social Science (SPSS) Analysis

The data accumulated from the questionnaires were analysed using the Statistical Package for Social Science (SPSS). A repeated measure of Analysis of Variance (ANOVA) was utilized to compare the mental pressure score and piezoelectric reading under rest and stress situations, and Cronbach's alpha measurement was used to test the reliability of the data. A correlation test was also conducted to find the correlation between the variables.

The analysis for this work was based on four hypotheses, which were:

- Hypothesis 1: There is a significant difference between phases (R1, S1, R2, S2 and R3) on the mental pressure score from the questionnaire.
- Hypothesis 2: There is a significant difference between conditions (Relax & Stress) on the piezoelectric reading.
- Hypothesis 3: There is a significant difference between phases (R1, S1, R2, S2 and R3) on the piezoelectric reading.
- Hypothesis 4: There are positive correlations between mental pressure scores and piezoelectric readings between phases.

3. RESULTS AND DISCUSSION

3.1. Piezoelectric-based reading vs oximeter HR measurement

The experiment in Section 2.1 was conducted to check the relationship between piezoelectric readings and the heart rates measured using an oximeter. The tabulated results are presented in Table 1.

https://doi.org/10.31436/iiumej.v25i2.3017

Based on the data, it can be observed that the value of the piezoelectric-based heart rate and the oximeter were not exactly within the same range, with the bpm obtained by the piezoelectric sensors being lower than those measured by the oximeter. Among the possible factors of the discrepancy are the resolution of the DAQ device used and the signal processing methods used in determining the peak-to-peak time intervals.

Table 1. The tabulated data for piezoelectric and oximeter data.

Sub	**N**	**Rest**		**Stress**	
		Piezo	*Oximeter*	*Piezo*	*Oximeter*
1	1	72	95	78	131
	2	72	97	76	119
	3	71	90	79	135
	4	72	97	78	130
	5	71	93	76	126
	6	70	89	79	137
2	1	72	94	78	130
	2	72	97	76	123
	3	71	91	79	134
	4	72	99	78	132
	5	71	93	76	126
	6	70	89	79	137

However, given that the primary objective of this study was to assess the viability of employing a piezoelectric sensor for classifying an individual's stress levels (i.e., stress versus relax), an exact match in heart rate values was not essential. What remained crucial was a strong correlation between all readings obtained from the same sensor device and the measurements derived from the oximeter.

Therefore, a statistical analysis was conducted using the collected data to quantify the strength of the linear relationship between the oximeter and piezoelectric-based HR measurement. The One-Tailed Pearson's Correlation Test was performed using SPSS, and the result showed that there was a high positive correlation between the piezoelectric reading and the oximeter reading, $r(12) = 0.993$, $p < 0.001$. This implies that as the oximeter value rose, the piezoelectric value similarly increased, thereby allowing the piezoelectric-based heart rate's current reading to be employed for mental condition analysis in a manner analogous to the oximeter.

3.2. Piezoelectric-based sensory system efficacy in stress detection

3.2.1. Mental Health Score via Questionnaire

Fig. 5 shows the mean and standard deviation of the STAI-6 score and psychological reactance questionnaire. Meanwhile, Fig. 6 illustrates the estimated marginal mean for each phase (i.e., R1, S1, R2, S2, R3) collected from 20 subjects. It can be observed from the graph that, in general, the data was consistent based on the low standard deviation value for all phases. Another outstanding observation that can be made from Fig. 5 and 6 is that both stress phases (S1 and S2) yielded a higher score for all subjects than the relax phases. The mean score for S1 and S2 is 2.76 ($SD = 0.78$) and 2.80 ($SD = 0.68$), respectively (Fig. 6); meanwhile, all relax phases yielded an average score below 2.0 (SD= 0.63), as demonstrated in Fig. 7.

https://doi.org/10.31436/iiumej.v25i2.3017

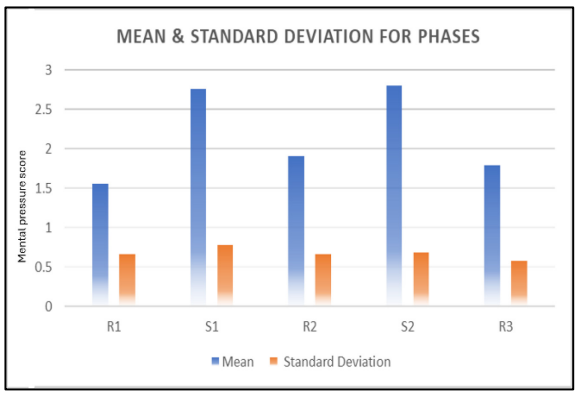


Figure 5. Mean & standard deviation for each phase of mental pressure score from STAI-6 and psychological questionnaire.

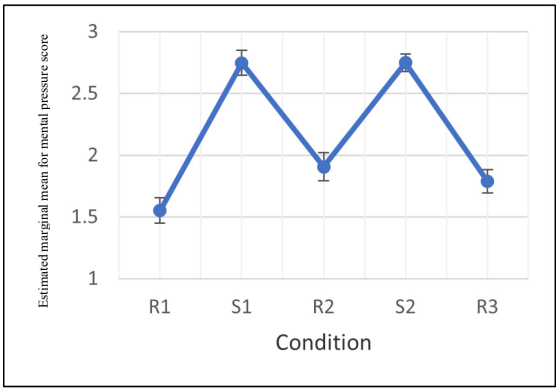


Figure 6. Estimated marginal mean of the effect of phases on mental pressure score from STAI-6 and psychological questionnaire.

An ANOVA test was performed to test Hypothesis 1 of this research. Results indicate that there were significant effects of phases (R1, S1, R2, S2, and R3) on the physiological signal, $F(4,16) = 44.856$, $p < 0.001$, partial $\eta 2 = 0.918$.

Hence Hypothesis 1 was accepted and therefore, the questionnaire score became the benchmark to indicate the real condition (i.e., relax or stress) of the subjects for the later experiments.

3.2.2. Piezoelectric Reading vs Mental Conditions

After the sensory system was developed, data from 20 volunteer subjects were analysed to investigate the reliability of the collected data and the efficacy of the piezoelectric-based HR measurement to differentiate relaxed and stressed conditions via statistical means. Fig. 7 illustrates the *Mean* ± *SD* of the piezoelectric-based HR measurement for all subjects with respect to time.

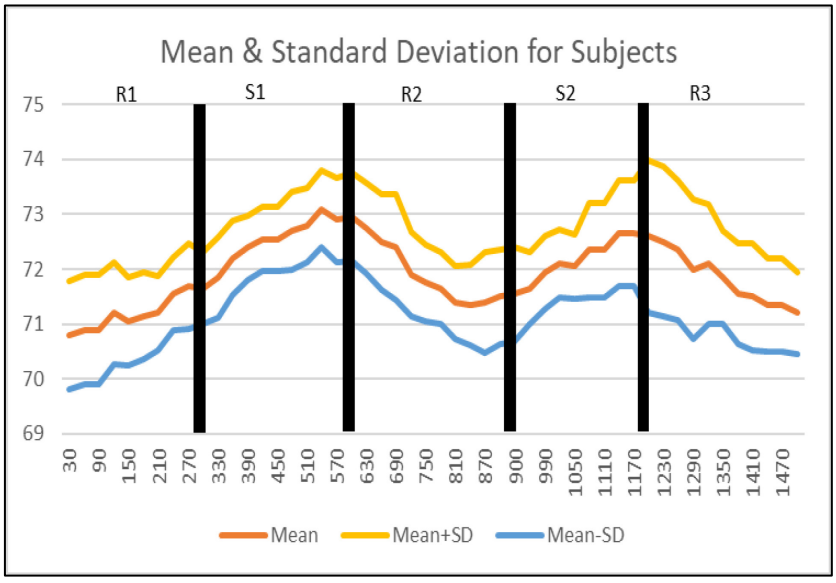


Figure 7. Mean ± SD of the piezoelectric-based HR measurement of 20 subjects.

https://doi.org/10.31436/iiumej.v25i2.3017

Cronbach's Alpha test was conducted to confirm the reliability of the data. It is the most common measure of internal consistency, and it is known as reliability. The findings indicate that data obtained from the piezoelectric sensor demonstrated a high level of reliability during both the relaxation (R1, R2, and R3) and stress (S1 and S2) phases, with coefficients of internal consistency of (30 readings; α = .942) and (20 readings; α = .871), respectively.

3.2.3. Piezoelectric readings and mental condition (R1, S1, R2, S2, and R3) correlation

ANOVA tests were performed to confirm Hypotheses 2 and 3 of this research.

Hypothesis 2: There is a significant difference between conditions (Relaxed vs. stressed) on the piezoelectric reading.

Repeated one-way measures of the ANOVA test were used for the effect toward the physiological signal with stress and relaxation as the independent variable and physiological signal as the dependent variable. The result shows that there is a significant effect of condition (relaxed and stressed) on the physiological signal, $F(1,19) = 80.953$, $p < 0.001$, partial $\eta^2 = 0.810$, thus proving Hypothesis 2 was accepted. Bonferroni Post Hoc Tests showed that subjects measured the lowest piezoelectric reading in the relax phase ($M = 71.615$, $SD = 0.126$) and the highest happened in the stress phase ($M = 72.395$, $SD = 0.097$). This outcome is in accordance with the findings of many other works summarized by [31], in which 18 out of 23 studies found that HR increases when a person is under stress.

To conclude this, when the subjects were in the relax phase, the sensor reading was lower, and when the subjects were in the stress phase, the sensor reading was higher. It shows that the piezoelectric sensor could distinguish between the relaxed and stressed conditions. Fig. 8 shows that the piezoelectric readings were lower in the relaxed condition and higher in the stressed condition. Hence, the results depict that the subjects were in the appropriate emotional and mental health situation as desired.

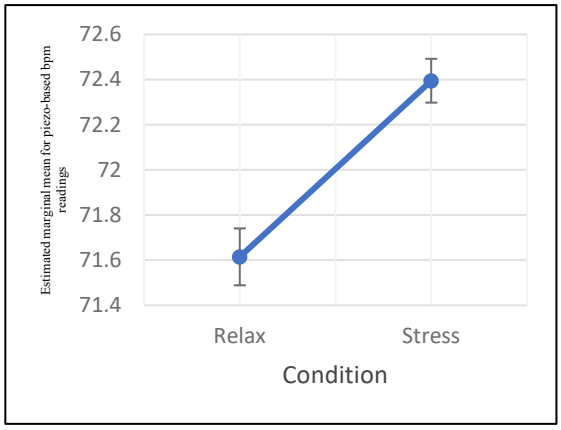


Figure 8. Estimated marginal mean of the effect of conditions on piezoelectric reading

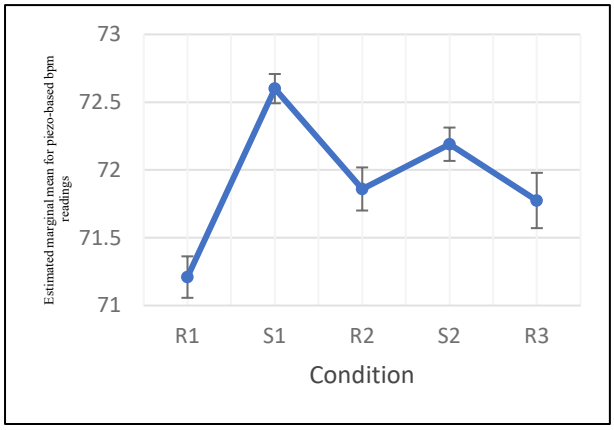


Figure 9. Estimated marginal mean of the effect of phases on piezoelectric reading

Hypothesis 3: There is a significant difference between phases on piezoelectric reading.

A repeated one-way measure of the ANOVA test was also used to measure the effect on the physiological signal, with phase as the independent variable and physiological signal as the

dependent variable. As results, there were significant effects of phase (R1, S1, R2, S2, and R3) on the physiological signal, $F(4,16) = 24.504$, $p < 0.001$, partial $\eta^2 = 0.860$. Hence Hypothesis 3 was accepted. The linear test of within-subject contrast indicates a significant relation between independent and dependent variables, $F(4,76) = 80.953$, $p < 0.001$, partial $\eta^2 = 0.810$. Fig. 9 illustrates that all relax phases (R1, R2, and R3) were relatively lower than stress phases (S1 and S2), and S1 ($M = 72.600$, $SD = 0.108$) displayed a higher marginal mean as compared to S2 ($M = 72.190$, $SD = 0.123$).

3.2.4. Piezoelectric readings and questionnaire findings correlation

Hypothesis 4: There are positive correlations between mental pressure score and piezoelectric reading between phases.

To confirm the hypothesis, a correlation test was run using SPSS to determine the correlation between the mental pressure scores for each relaxed and stressed condition. Table 2 tabulates the Pearson's correlation, *r*-value and significance, *p*-value by comparing piezoelectric readings with respect to phases (1st row) with questionnaire score with respect to phases (1st column).

Table 2: Correlation test results between piezoelectric readings and mental pressure score from questionnaire. (Note: 1st column refers to the labels for phases with respect to the questionnaire, while 1st row refers to phases with respect to piezoelectric reading)

	R1_piezo	**S1_piezo**	**R2_piezo**	**S2_piezo**	**R3_piezo**
R1_Ques	*r(20)=0.595, p<0.01*	Not significant	Not significant	*r(20)=0.436, p<0.05*	Not significant
S1_Ques	Not significant	Not significant	Not significant	Not significant	Not significant
R2_Ques	Not significant	Not significant	Not significant	Not significant	Not significant
S2_Ques	*r(20)=0.544, p<0.01*	*r(20)=0.401, p<0.05*	*r(20)=0.453, p<0.05*	Not significant	Not significant
R3_Ques	*r(20)=0.396, p<0.05*	Not significant	Not significant	*r(20)=0.404, p<0.05*	Not significant

From the table, only R1 shows a high correlation between the piezoelectric reading and the questionnaire's mental score ($r(20)=0.595$, $p<0.05$), while the rest do not show a high correlation between both methods. However, some correlation can be detected across relax-relax or stress-stress phases. S2_Ques shows significant correlation with S1_piezo ($r(20)=0.401$, $p<0.05$), and R3_Ques is correlated to R1_piezo ($r(20)=0.396$, $p<0.005$).

It can also be observed that there are significant correlations between stress-relax conditions in the table. For instance, S2_Ques is seen to be highly correlated with R1_piezo and R2_piezo. One viable factor is the carried-over effect from the previous phase, hence affecting the subject's answers in the questionnaire.

Nevertheless, these findings prove that piezoelectric readings are correlated to mental condition (i.e., stress and relax), but may not accurate enough to relate data from both methods of a specific phase (e.g., S1 to S1).

4. CONCLUSIONS

This study's goal is to show the feasibility of detecting stress by changes in HR using a piezoelectric-based sensor. Twenty subjects were recruited to investigate the feasibility of the proposed sensory system.

This project successfully designed a piezoelectric-based sensory system to detect the stress level of a human using a piezoelectric sensor. Although the data that was retrieved by the piezoelectric sensor was not the same as the reading from the oximeter (that used a different sensing method), there was nonetheless a high positive correlation between the two of them (p-value < 0.01).

In addition, the data obtained from SPSS analysis aligned remarkably with the phase involved. High mental pressure scores were found in the stress phase, which was the target of that phase, while low mental pressure scores were found in the relax phase, aligning the phase and piezoelectric readings of the phase involved. Hence, it can be concluded from the findings that a piezoelectric-based sensor is feasible to obtain heart rate projection and consequently distinguish between relaxed and stressed conditions.

Importantly, the sensory system developed in this study can be further improved to yield better performance. In the future, a classification algorithm shall be developed to categorize stressed vs relaxed based on piezoelectric data. A study to check the feasibility of the proposed sensory system in detecting different levels of stress is also recommended. In addition, the number of subjects with varying BMIs must also be added to increase the robustness of the data collected. Hence, there is a need for the introduction of newer, more compact, and ergonomically designed piezoelectric-based sensors in the future. This step will facilitate convenient comparisons with the outcomes presented in this project.

ACKNOWLEDGMENT

The research leading to these results has received funding from the UMP-IIUM-UiTM Sustainable Research Collaboration 2020 grant with a project titled "Analysis on the Correlation Between Piezoelectric Signal and Respiratory Rate (RR) and Heart Rate (HR) for Stress Detection Among Young Adults" under the Grant agreement number SRCG20-036-0036.

REFERENCES

[1] J. Chen, M. Abbod, and J. S. Shieh, "Pain and stress detection using wearable sensors and devices—a review," *Sensors (Switzerland)*, vol. 21, no. 4. 2021. doi: 10.3390/s21041030.

[2] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J Biomed Inform*, vol. 92, p. 103139, Apr. 2019, doi: 10.1016/J.JBI.2019.103139.

[3] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on Psychological Stress Detection Using Biosignals," *IEEE Trans Affect Comput*, vol. 13, no. 1, pp. 440–460, 2022, doi: 10.1109/TAFFC.2019.2927337.

[4] N. Widanti, B. Sumanto, P. Rosa, and M. Fathur Miftahudin, "Stress level detection using heart rate, blood pressure, and GSR and stress therapy by utilizing infrared," *2015 International Conference on Industrial Instrumentation and Control, ICIC 2015*, pp. 275–279, Jul. 2015, doi: 10.1109/IIC.2015.7150752.

[5] J. R. Machado Fernández and L. Anishchenko, "Mental stress detection using bioradar respiratory signals," *Biomed Signal Process Control*, vol. 43, pp. 244–249, May 2018, doi: 10.1016/J.BSPC.2018.03.006.

[6] U. Lundberg *et al.*, "Psychophysiological stress and emg activity of the trapezius muscle," *International Journal of Behavioral Medicine 1994 1:4*, vol. 1, no. 4, pp. 354–370, Dec. 1994, doi: 10.1207/S15327558IJBM0104_5.

[7] C. H. Vinkers *et al.*, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, Sep. 2013, doi: 10.3109/10253890.2013.807243.

[8] V. Engert, A. Merla, J. A. Grant, D. Cardone, A. Tusche, and T. Singer, "Exploring the Use of Thermal Infrared Imaging in Human Stress Research," *PLoS One*, vol. 9, no. 3, p. 90782, Mar. 2014, doi: 10.1371/JOURNAL.PONE.0090782.

[9] U. Lundberg *et al.*, "Psychophysiological stress and emg activity of the trapezius muscle," *Int J Behav Med*, vol. 1, no. 4, 1994, doi: 10.1207/s15327558ijbm0104_5.

[10] C. H. Vinkers *et al.*, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, 2013, doi: 10.3109/10253890.2013.807243.

[11] A. A. Amiruldin and A. S. Ghazali, "You Are Too Loud! Classification of Psychological Conditions for Stress Detection System Using Galvanic Skin Response," in *Lecture Notes in Electrical Engineering*, 2022. doi: 10.1007/978-981-19-2095-0_6.

[12] B. Szakonyi, I. Vassányi, E. Schumacher, and I. Kósa, "Efficient methods for acute stress detection using heart rate variability data from Ambient Assisted Living sensors," *Biomed Eng Online*, vol. 20, no. 1, 2021, doi: 10.1186/s12938-021-00911-6.

[13] D. Hu and L. Gao, "Psychological Stress Level Detection Based on Heartbeat Mode," *Applied Sciences (Switzerland)*, vol. 12, no. 3, 2022, doi: 10.3390/app12031409.

[14] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Analysis of stroop colorword test-based human stress detection using electrocardiography and heart rate variability signals," *Arab J Sci Eng*, vol. 39, no. 3, 2012, doi: 10.1007/s13369-013-0786-8.

[15] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear Heart Rate Variability features for real-life stress detection. Case study: Students under stress due to university examination," *Biomed Eng Online*, vol. 10, 2011, doi: 10.1186/1475-925X-10-96.

[16] S. N. M. Sayed Ismail, N. A. Nor, and S. Z. Ibrahim, "A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, 2022, doi: 10.1016/j.jksuci.2022.04.012.

[17] F. Riaz, M. A. Azad, J. Arshad, M. Imran, A. Hassan, and S. Rehman, "Pervasive blood pressure monitoring using Photoplethysmogram (PPG) sensor," *Future Generation Computer Systems*, vol. 98, pp. 120–130, Sep. 2019, doi: 10.1016/j.future.2019.02.032.

[18] Sattar Y and Chhabra L, "Electrocardiogram. [Updated 2023 Jun 5]," in *StatPearls [Internet]*, Treasure Island (FL): StatPearls Publishing, 2023.

[19] Castaneda D, Esparza A, Ghamari M, Soltanpur C, and Nazeran H, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int J Biosens Bioelectron*, vol. 4, no. 4, Aug. 2018.

[20] S. Das Mahapatra *et al.*, "Piezoelectric Materials for Energy Harvesting and Sensing Applications: Roadmap for Future Smart Materials," *Advanced Science*, vol. 8, no. 17. 2021. doi: 10.1002/advs.202100864.

[21] J. H. Park, D. G. Jang, J. W. Park, and S. K. Youm, "Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor," *Sensors (Switzerland)*, vol. 15, no. 9, 2015, doi: 10.3390/s150923402.

[22] E. Cepeda, D. H. Peluffo-Ordóñez, P. Rosero-Montalvo, M. A. Becerra, A. C. Umaquinga-Criollo, and L. Ramírez, "Heart Rate Detection using a Piezoelectric Ceramic Sensor: Preliminary results," *Bionatura*, vol. 7, no. 3, 2022, doi: 10.21931/RB/2022.07.03.30.

[23] Z. Ji and M. Zhang, "Highly sensitive and stretchable piezoelectric strain sensor enabled wearable devices for real-time monitoring of respiratory and heartbeat simultaneously," *Nanotechnology and Precision Engineering*, vol. 5, no. 1, 2022, doi: 10.1063/10.0009365.

[24] Y. Hamonangan and W. Purnamaningsih, "Heartbeat detection system using piezoelectric transducer," in *AIP Conference Proceedings*, American Institute of Physics Inc., Feb. 2017. doi: 10.1063/1.4976801.

[25] V. Setyowati, J. Muninggar, and M. R. S. Shantina, "Design of heart rate monitor based on piezoelectric sensor using an Arduino," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Feb. 2017. doi: 10.1088/1742-6596/795/1/012016.

[26] J. Ahmad Rather and Y. Shrivastava, "Effect of music therapy on pre-competition anxiety in college level soccer players of Kashmir," *~ 1176 ~ International Journal of Physiology*, vol. 4, no. 1, pp. 1176–1178, 2019, [Online]. Available: www.journalofsports.com

[27] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Comput Biol Med*, vol. 107, pp. 182–196, Apr. 2019, doi: 10.1016/J.COMPBIOMED.2019.02.015.

[28] L. J. Julian, "Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A)," *Arthritis Care Res (Hoboken)*, vol. 63, no. SUPPL. 11, Nov. 2011, doi: 10.1002/acr.20561.

[29] A. J. Donnell, A. Thomas, and W. C. Buboltz, "Psychological reactance: Factor structure and internal consistency of the questionnaire for the measurement of psychological reactance," *Journal of Social Psychology*, vol. 141, no. 5, 2001, doi: 10.1080/00224540109600581.

[30] A. A. Amiruldin and A. S. Ghazali, "You Are Too Loud! Classification of Psychological Conditions for Stress Detection System Using Galvanic Skin Response," pp. 51–60, 2022, doi: 10.1007/978-981-19-2095-0_6.

[31] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on Psychological Stress Detection Using Biosignals," *IEEE Trans Affect Comput*, vol. 13, no. 1, pp. 440–460, 2022, doi: 10.1109/TAFFC.2019.2927337.

https://doi.org/10.31436/iiumej.v25i2.3029

ENHANCING PROSTHETIC CONTROL: NEURAL NETWORK CLASSIFICATION OF THUMB MUSCLE CONTRACTION USING HD-SEMG SIGNALS

MUHAMMAD MUKLIS SUHAIMI, AIMI SHAZWANI GHAZALI*, AHMAD JAZLAN HAJA MOHIDEEN, MUHAMMAD HARIZ HAFIZALSHAH, SHAHRUL NA'IM SIDEK

Dept. of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia.

**Corresponding author: aimighazali@iium.edu.my*

(Received: 20 September 2023; Accepted: 31 October 2024; Published online: 14 July 2024)

ABSTRACT: The progression of prosthetic technology, enabling precise thumb control and movement, has reached a stage where noninvasive techniques for capturing bioelectrical signals from muscle activity are preferred over alternative methods. While electromyography's applications extend beyond just interfacing with prostheses, this initial investigation delves into evaluating various classifiers' accuracy in identifying rest and contraction states of the thumb muscles using extrinsic forearm readings. Employing a High-Density Surface Electromyogram (HD-sEMG) device, bioelectrical signals generated by muscle activity, detectable from the skin's surface, were transformed into contours. A training system for the thumb induced muscle activity in four postures: 0°, 30°, 60°, and 90°. The collection of HD-sEMG signals originating from both the anterior and posterior forearms of seventeen participants has been proficiently classified using a neural network with 100% accuracy and a mean square error (MSE) of 1.4923 x 10^{-5} based on the testing dataset. This accomplishment in classification was realized by employing the Bayesian regularization backpropagation (trainbr) training technique, integrating seven concealed layers, and adopting a training-validation-testing proportion of 70-15-15. In the realm of future research, an avenue worth exploring involves the potential integration of real-time feedback mechanisms predicated on the recognition of thumb muscle contraction states. This integration could offer an enhanced interaction experience between users and prosthetic devices.

ABSTRAK: Perkembangan teknologi prostetik mengguna pakai kaedah selamat iaitu isyarat bioelektrikal yang diperoleh dari pergerakan otot lebih digemari digunakan berbanding kaedah alternatif. Ini membolehkan kawalan dan pergerakan ibu jari dengan tepat. Sementara aplikasi elektromiografi telah melangkah jauh melebihi antara muka prostesis. Kajian awal ini mengkaji pelbagai ketepatan klasifikasi dalam mengenal pasti keadaan rehat dan kontraksi otot ibu jari menggunakan bacaan lengan bawah ekstrinsik. Dengan menggunakan peranti Elektromiogram Permukaan Kepadatan-Tinggi (HD-sEMG), isyarat bioelektrikal yang terhasil dari pergerakan otot, boleh ditanggalkan dari permukaan kulit, di ubah kepada kontur. Sistem latihan pada ibu jari menghasilkan pergerakan otot dalam empat postur iaitu: 0°, 30°, 60°, dan 90°. Isyarat terkumpul dari HD-sEMG berasal dari kedua-dua lengan tangan anterior dan posterior dari 17 peserta telah diklasifikasi dengan cekap menggunakan rangkaian neural dengan ketepatan 100% dan min kuasa dua ralat (MSE) sebanyak 1.4923 x 10-5 berdasarkan setdata yang diuji. Klasifikasi sempurna ini dicapai dengan menggunakan teknik latihan aturan rambatan-belakang Bayesian (trainbr), mengguna pakai tujuh lapisan tersembunyi dengan gabungan latihan-validasi-ujian mengikut kadar 70-15-15. Pada masa hadapan, pengkaji boleh menerokai potensi integrasi mekanisme tindak balas nyata dalam meramal dan

mengenali kontraksi otot ibu jari. Integrasi ini mungkin membolehkan pengalaman interaksi antara peranti prostetik dan pengguna.

KEYWORDS: *Thumb Posture, High-Density Surface Electromyogram (HD-sEMG), Forearm Muscle, Classification.*

1. INTRODUCTION

The human hand serves as the principal organ for tactile perception thanks to its high sensory receptor density. The hand performs a number of vital actions that are essential to human interaction and expression in everyday life, such as gripping for pick-up and holding mechanisms in addition to gesturing for nonverbal communication [1]. Executing complex motions that need coordination and precision like writing, relies on the intricate movements of hands, orchestrated by the muscles. For instance, gripping activities are created by the muscles in the hand, which include intrinsic muscles placed inside the hand and extrinsic muscles emerging from the forearm [2].

The absence of this vital biological component has a significant impact on an individual's freedom and general quality of life. As they adapt to their new surroundings, hand amputees frequently face physical, emotional, and mental challenges [3]. Significantly due to technological advancements, many hand amputees employ prosthetic devices to restore some level of hand function. Prosthetic technological advancements have resulted in the creation of sophisticated myoelectric prostheses that can be controlled by muscle impulses from the residual limb [4]. Amputees may use these prostheses to do a wide range of chores and activities, mimicking the capabilities of our natural hands.

There are different types of amputees, mainly transradial and transcarpal, distinguished by the extent of limb loss. Transradial amputees have forearms amputated between the elbow and wrist, losing the hand and wrist but keeping the elbow. This type of amputation affects the synergy between intrinsic and extrinsic hand muscles. On the other hand, transcarpal amputees lose part of the palm and wrist but retain some forearm and the wrist joint [5]. This type often preserves certain forearm muscles crucial for hand movements, making them the focus of this study since transcarpal amputees show potential for effective prosthetic use due to sustained forearm muscles capable of signaling specific hand movements [6].

In the context of prosthetic control, intrinsic and extrinsic muscles can be used to provide signals that direct the motions of a prosthetic hand, so increasing the device's utility and variety. Notably, the High-Density Surface Electromyogram (HD-sEMG) approach can detect muscle activity signals related to thumb movements using extrinsic muscles. HD-sEMG employs a larger number of electrodes to encompass a broader muscle area. The current state of prosthetic technology lacks precision, hindering its ability to address crucial aspects affecting users' daily lives [7]. This limitation hampers transcarpal amputees' ability to perform intricate movements essential for tasks involving grasping and manipulation. Moreover, HD-sEMG mitigates issues such as electrode misplacement that can occur with standard sEMG [8]. As a result, the imprecise thumb posture classification system from standard sEMG restricts transcarpal amputees from adapting to a wide range of tasks and activities. In general, there is a critical need for advancements in thumb posture classification from HD-EMG signals to revolutionize prosthetic control for transcarpal amputees, ensuring greater precision, functionality, and overall well-being for users [9].

A recent study has dedicated attention to optimizing the application of this technology to maximize the efficacy of HD-sEMG for hand gesture applications [10], while another

https://doi.org/10.31436/iiumej.v25i2.3029

combines the data collected with machine learning technique for classification purposes [11]. In line with this approach, a large dataset can be successfully handled by modern machine learning algorithms, combining with HD-sEMG's ability to collect large amounts of data. The first steps include feature extraction, which is a process of obtaining relevant information from the raw data provided by HD-sEMG. These features are classified into three types: time-domain (TD), frequency-domain (FD), and time-frequency domain (TFD). TFD analysis approaches find common ground in a variety of engineering fields, such as in gesture recognition purposes [12]. While frequency-based features (FD) have been employed due to their simple mathematical function (can be extracted using Fourier transformations) in earlier studies for healthcare purposes [13]. TD features have been widely used in previous research due to their superior performance in various fields such as in detecting sympathetic activity in post-acute COVID-19 patients [14], besides fault diagnosis of machines using vibration data [15]. This research is centered on time-domain (TD) features derived from the high-density surface electromyography (HD-sEMG) signals. This choice was informed by analyses conducted in a previous study [16], which advocated for the utilization of TD features, specifically emphasizing the inclusion of Root Mean Square (RMS) data, for effective biosignal feature extraction.

Based on the extracted features, the data are typically fed into machine learning for classification purposes. However, earlier studies have not treated the ratio of training, validation and testing in much detail. The research described in [17] used a 70-15-15 ratio in detecting healthy, myopathy, and neuropathy conditions based on the EMG signals in improving the evaluation of neuromuscular disorders. As per the performance assessment of the Convolutional Neural Network (CNN) approach, it attains a precision level of 98.57%. In another study as shown in [18], an 80-10-10 rule was employed to predict hand posture (pinch versus grip) and grasp force. The study utilized forearm sEMG and artificial neural networks (ANNs) during tasks that involved variations in repetition rate and duty cycle. The outcomes showed that overall accuracy for hand posture prediction was reported at 79%, while the overall accuracy for hand force prediction was recorded at 73%. In certain instances, a 60-20-20 split was utilized, as evidenced by a study with larger datasets consisting of 188 EMG signal data and 223 ECG signal data [19]. This approach was employed for muscle fatigue analysis and stress detection in the context of upper limb trauma rehabilitation. The outcomes revealed a remarkable 95% accuracy for multiclass muscle fatigue classification and a notable 97% accuracy during the binary classification of mental stress.

In this study, High-Density Surface Electromyogram (HD-sEMG) signals, derived from extrinsic forearm muscles situated on both the anterior and posterior sides, were utilized for classification purposes. During the experimental phase, participants were instructed to exert pressure on their thumb using a specifically designed platform. The thumb's posture was standardized at four different degrees. Employing a neural network developed within the Matlab environment, this study aims to reveal the highest accuracy achieved in classifying the extracted data.

Specifically, the study will assess the influence of forearm sides and thumb postures on HD-sEMG readings, considering three different training functions. Subsequent steps will involve exploring the optimal number of hidden layers for classifying the collected data, followed by selecting the best training-validation-testing ratios based on accuracy and mean square error (MSE).

https://doi.org/10.31436/iiumej.v25i2.3029

2. EXPERIMENTAL DESIGN

2.1 HD-sEMG Characteristics

HD-sEMG is a non-invasive measurement method for detecting muscle activity. The main difference between HD-sEMG and standard EMG is that HD-sEMG uses an array of closely spaced electrodes, which enables the detection of more detailed spatial muscle activity due to the large number of electrodes arranged in a dense manner. Standard EMG, on the other hand, typically uses a bipolar configuration consisting of two electrodes placed on the muscle of interest, enabling the detection of temporal muscle activity [23]. The ability to detect spatial muscle activity is advantageous in the development of prostheses. However, translating the HD-SEMG data into precise prosthetic hand movements remains a challenge. This section describes an experimental procedure for data collection to investigate the relationship between thumb movements/force exertion and forearm muscle activity.

2.2 Participants and Design

Seventeen students from the IIUM Gombak campus, aged 24 to 30 (with an average age of 26.35 and a standard deviation of 1.64), were chosen at random to participate in a within-subject experimental arrangement involving two conditions (relaxed vs. contracted) x four thumb postures (zero-degree vs. thirty-degree vs. sixty-degree vs. ninety-degree) for both forearm sides (anterior and posterior) across three trials based on RMS values. Prior to the start of the experiment, all participants were briefed on the experimental methods, after which they were granted their informed permission. The International Islamic University Malaysia (IIUM) approved the study protocol with the reference number 2020-080. There was no history of surgery, nerve damage, or accidents involving any of the individuals' dominant hands.

2.3 Task

Throughout the experimental session, all participants were instructed to engage their Maximum Voluntary Contraction (MVC) of thumb force using a portable thumb training device that had been developed (depicted in Fig. 1) in a neutral placement (at zero-degree position).

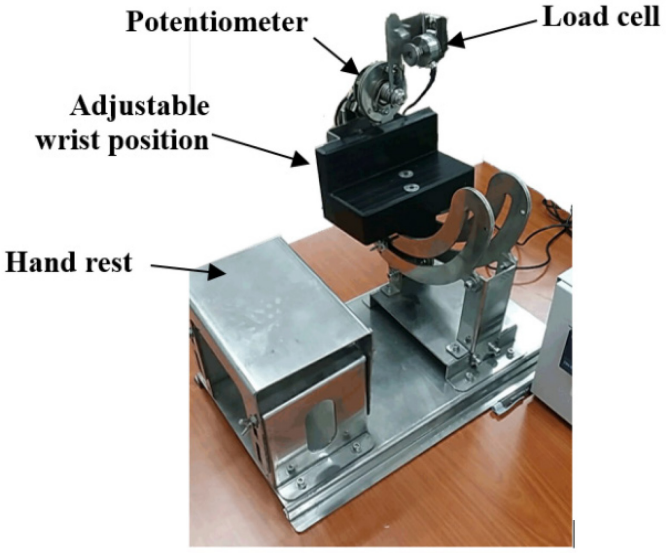


Figure 1. Thumb training platform (adopted from [5])

https://doi.org/10.31436/iiumej.v25i2.3029

The device included a potentiometer for detecting thumb posture arrangements, accompanied by a load cell for quantifying the force applied by participants. In addition, an adaptive wrist positioning system with a robust locking mechanism was designed to ensure accurate thumb position angles. A level hand rest was intended for participants to comfortably rest their arms during the experiment, effectively minimizing the potential onset of undue fatigue.

The thumb training platform was previously employed in a preceding study [16], albeit with slight adjustments made to establish a fixed angle configuration. For this experiment, angles of the thumb were set at zero-degree, thirty-degree, sixty-degree, and ninety-degree as shown in Fig. 2. Before the experiment, the experimenter set the dedicated angle by measuring the angle between the black base of the device and the tip of the load cell.

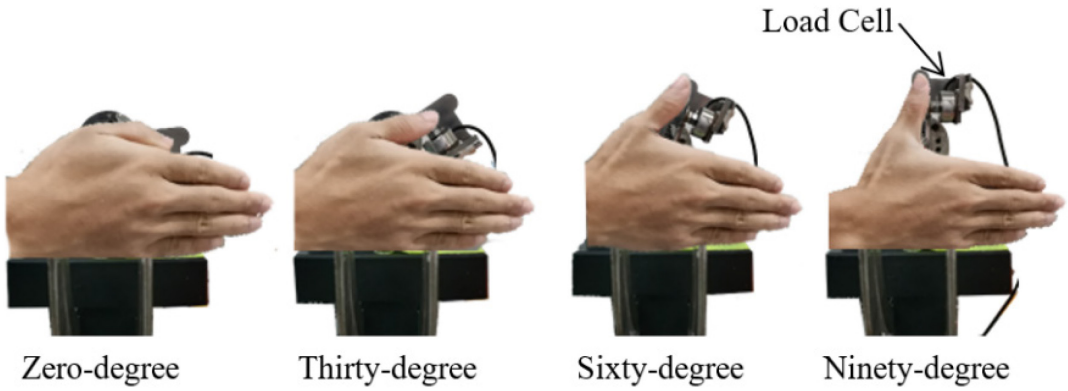


Figure 2. Thumb postures (adopted from [5])

Based on the integration of the Matlab program within the thumb training system, an automated process was employed to ascertain 30% of the MVC values for each thumb posture and subject, subsequently generating a corresponding graphical trajectory shown in Fig. 3.

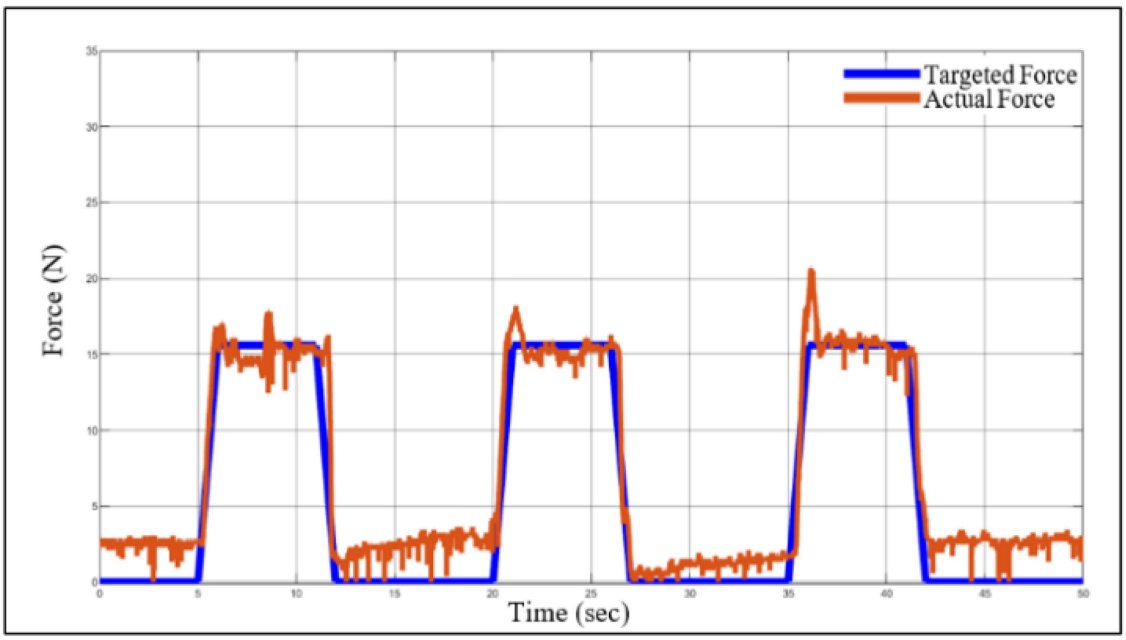


Figure 3. The trajectory of rest (force = 0 N) and contract (force= 30% MVC) conditions (adopted from [5])

https://doi.org/10.31436/iiumej.v25i2.3029

Utilizing the graphical user interface, participants engaged in repetitive cycles of resting and contracting positions for each trial. The blue line represents the targeted force, indicating the force that participants are required to exert, while the red line represents the actual force exerted by participants, measured by the load cell. This protocol encompassed three trials for every thumb posture, resulting in a total of nine contractile force measurements for each distinct thumb posture.

2.4 HD-sEMG Setting

HD-sEMG signals were captured using a portable biomedical signal amplifier (referred to as "Sessantaquattro") developed by OT-Bioelettronica . The electrode pad, organized in a grid pattern, consisted of 64 electrodes arranged in 5 columns and 13 rows, spaced 8mm apart. Following the methodology outlined in [20], the electrode pads were positioned on the anterior and posterior forearm, specifically at 25% of the previously measured full forearm length, counted from the ulnar head to the elbow crease. The positioning of the electrode in this study was also in line with a previous research [21], which found that the activation of the thumb is concentrated in the mid-forearm, adjacent to the radius bone.

Prior to affixing the patch onto participants' dominant forearm, a conductive gel was applied as a medium to ensure a secure connection between the skin and electrodes. This facilitated the direct transmission of electric impulses to the underlying tissues.

2.5 Procedure

Before beginning any experimental procedures, participants were given a consent form and instructed to read and complete it. After measuring each participant's forearm for patch placement purposes, the experimenter attached the electrodes to the forearm. Participants were then advised to sit up straight in a comfortable position before the data-collecting process began, as shown in Fig. 4.

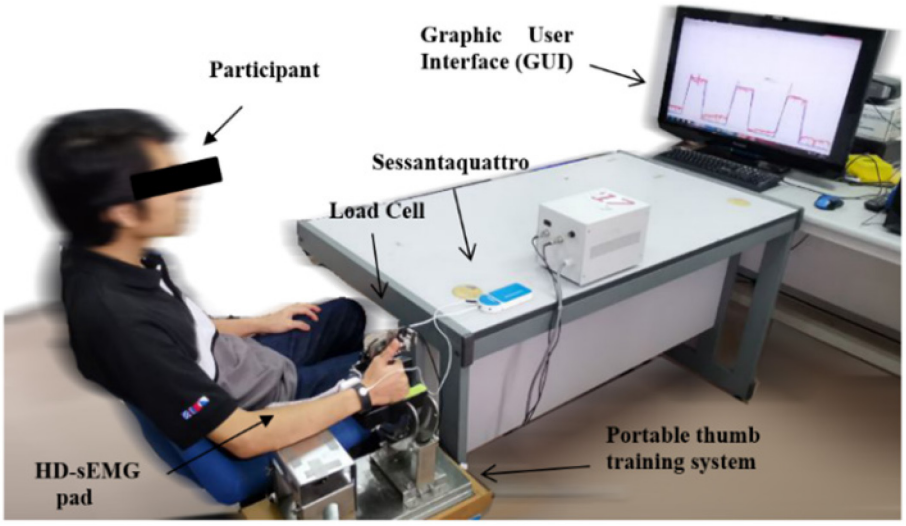


Figure 4. Experimental setup (adopted from [5])

Following that, instructions were given to the participants on how to place their arms on the portable thumb training device in a neutral posture (i.e., at a zero-degree angle). Participants were then directed to initiate their Maximum Voluntary Contraction (MVC) in order for MATLAB to calculate 30% of the MVC and generate the corresponding trajectory. This trajectory represented the force that participants had to apply over a certain period of time.

https://doi.org/10.31436/iiumej.v25i2.3029

Each trial required participants to match the anticipated trajectory by applying thumb force for five seconds, followed by an eight-second rest interval meant to reduce muscle strain. The experiment took around an hour to complete for all four thumb postures, following which the participants were debriefed by the experimenter. In recognition of their involvement, participants were provided with a monetary incentive at the end of the experiment.

2.6 Data processing

The collected data was subsequently extracted. A time-domain feature, which is RMS, was derived using software developed by OT-Bioelettronica. This feature computed through equation (1), encapsulating important aspects of the data.

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k)^2} \quad (1)$$

where N is the number of samples per window, and x_k is the amplitude of the signal of the amplifier input measured in mV. Note that, the samples per window were predefined at 1000 samples.

As each dataset was acquired from individual subjects exhibiting distinct characteristics, the process of normalization became imperative to address inter-subject variability. Normalization also helped in preventing bias due to domination of certain variables because of their greater magnitudes than others. The HD-sEMG signal data was subjected to normalization as per the formula depicted in equation (2).

$$\frac{n}{\bar{x}} \times 100\% \quad (2)$$

where $\bar{x}$ is the average of all normalized values, and n is n^{th} electrode's data out of the 64 electrodes.

3. RESULTS AND DISCUSSION

A total of 1632 data points were gathered for this study, comprising 2 conditions (contract vs. rest) x 4 angles (zero-degree vs. thirty-degree vs. sixty-degree vs. ninety-degree) x 3 trials for each condition x 4 readings (for each trial) x 17 participants. However, certain data from specific participants were excluded from the analysis due to grounding issues. Consequently, the dataset used for results encompassed 960 entries with 128 readings from the electrodes (64 electrodes each for posterior and anterior).

The data underwent classification using Matlab's neural network training. The selection of the training function was limited to Levenberg-Marquardt backpropagation (trainlm), Bayesian regularization backpropagation (trainbr), and Scaled conjugate gradient backpropagation (trainscg), depending on factors like classification speed and data size. Additionally, the neural network's hidden layer count was adjustable for better accuracy. Notably, the training, validation, and testing ratio was manually configured as suggested by earlier studies [22-24]. Lastly, customization of the classified data's performance was done, in which the neural network's performance was evaluated using the mean squared error value (MSE) [25].

To facilitate analysis, the data was divided into random segments (dividerand). 128 input nodes (64 electrodes x two forearm sides) were designed to classify the data into eight different classes (four thumb postures x two conditions), as detailed in Table 1.

https://doi.org/10.31436/iiumej.v25i2.3029

Table 1. Classes for neural network

Class	Condition	Thumb posture
1	Contract	Zero-degree
2		Thirty-degree
3		Sixty-degree
4		Ninety-degree
5	Relax	Zero-degree
6		Thirty-degree
7		Sixty-degree
8		Ninety-degree

In the first step, the number of hidden layers was temporarily set to one due to the small data size and the most popular 70-15-15 ratio was used for training-validation-testing. The outcomes indicated fluctuations in the Mean Squared Error (MSE) value for this setup. The configuration employing trainbr achieved the lowest error, which is at 0.0011, followed by trainlm at 0.0016, while trainscg exhibited the highest error at 0.1736.

For trainbr, the best training performance was 2.512×10^{-11} recorded at epoch 112. Further investigation was done to observe the receiver operating characteristic (ROC) of the classified data. The ROC can be interpreted as the closer each curve aligns with the left and upper boundaries of the plot, the more accurate the classification. While a true positive corresponds to a situation where the model accurately predicts the positive class, the false positive denotes an instance where the model mistakenly predicts the positive class. The findings depicted in Fig. 5 indicate elevated false positive rates in the testing ROC for two specific classes: Class 5 and Class 6. Conversely, the remaining classes exhibited no errors in this regard.

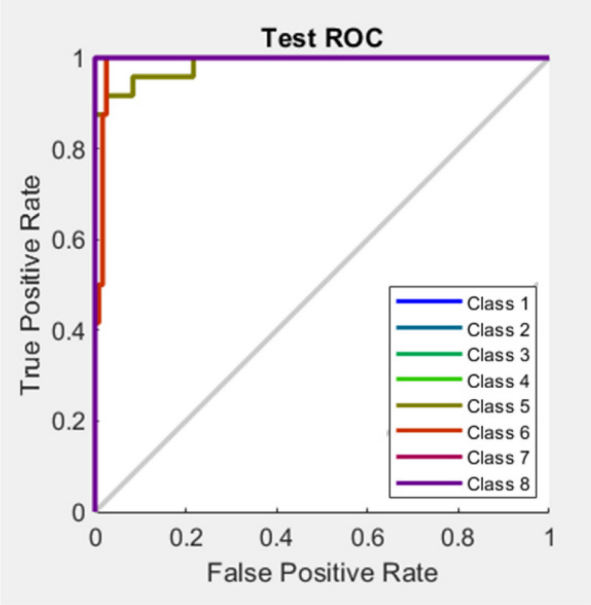


Figure 5. ROC for trainbr with 70-15-15 ratio

Derived from the confusion matrix of the testing subset, approximately 96.5% of the data was accurately classified. Amongst these, those two specific classes exhibited comparatively lower accuracy: 91.3% for Class 5 and 88.0% for Class 6. Taking into account the comprehensive confusion matrix, the collective rate of accurately classified data reached 99.5%.

https://doi.org/10.31436/iiumej.v25i2.3029

The second step of neural network analysis proceeded by varying the hidden layer size value, using trainbr function and still, a 70-15-15 ratio. Previous research presents divergent viewpoints regarding the optimal count of hidden layers. In [26], the recommendation leaned towards employing just one or two hidden layers to minimize neural network errors. Conversely, [27] proposed that utilizing five hidden layers adequately facilitated the classification of handwritten numerals. Consistent with the suggestions in [28] that advocated for trial-and-error approaches in determining hidden layer counts, [27] accentuated that augmenting the number of hidden layers led to a corresponding enhancement in the performance of the neural network under development.

Drawing from existing literature, alterations to the number of hidden layers were implemented in accordance with the details presented in Table 2. Furthermore, in alignment with the trial and error methodologies advocated in [28], the hidden layer configurations were diversified across multiple numbers to align with the classification's performance.

Table 2. MSE and % accuracy of the data based on the number of hidden layers

Hidden layer	MSE	% Correctly classified data (testing subset)
1	0.0011	96.5
2	0.0021	93.8
3	1.1943×10^{-4}	99.3
4	1.3852×10^{-4}	99.3
5	1.0235×10^{-4}	99.3
6	3.0434×10^{-5}	100
7	1.4923×10^{-5}	100
8	5.2345×10^{-4}	98.6

The escalation of hidden layers was capped at eight due to a decline in the neural network's performance. Specifically, when the hidden layer count was set at eight, both the MSE demonstrated an increase and the percentage of correctly classified data exhibited a decrease. Drawing insights from the outcomes presented in Table 2, a discernment can be drawn that the optimal hidden layer count for the given dataset amounted to seven. Additionally, it's notable that the computational time expanded with each increase in the hidden layer count. For instance, when the hidden layer count was set at five, the neural network required a mere 17 seconds to classify the data, whereas this time extended to a minute and 51 seconds when the hidden layer count was raised to eight.

Consequently, in the third phase of analysis, the training function was configured as 'trainbr', with the hidden layer count fixed at seven. Concurrently, variations were applied to the distribution of the training-validation-testing dataset ratios, encompassing both 80-10-10 and 60-20-20 settings. The findings indicated that with the 80-10-10 ratio, the neural network achieved a testing dataset classification accuracy of 97.9%, accompanied by an MSE of 5.2433x10-4. On the other hand, the 60-20-20 configuration yielded a classification accuracy of 99.0% and an MSE of 5.7381×10^{-4}.

As the 70-15-15 training-validation-testing ratio demonstrated superior classification accuracy in comparison to other ratios, additional analysis was conducted to delve into the finer aspects of this classification performance. The best training performance recorded was

https://doi.org/10.31436/iiumej.v25i2.3029

$1.3231x10^{-10}$ at epoch 105. Details of the performance for neural network training state can be seen in Fig. 6.

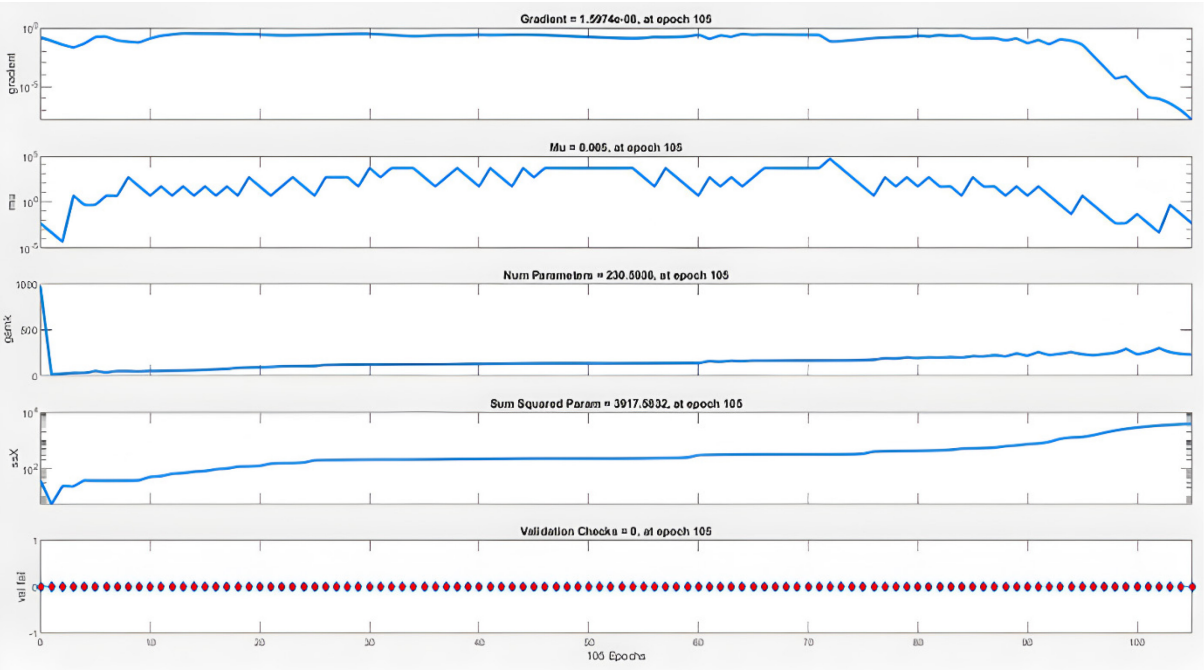


Figure 6. Training state performance for trainbr with seven hidden layers and 70-15-15 ratio

Furthermore, the comprehensive confusion matrix revealed a remarkable outcome, indicating that all data points, irrespective of their classes, were impeccably classified with a flawless accuracy of 100%. This exceptional classification achievement is corroborated by the ROC outputs depicted in Fig. 7.

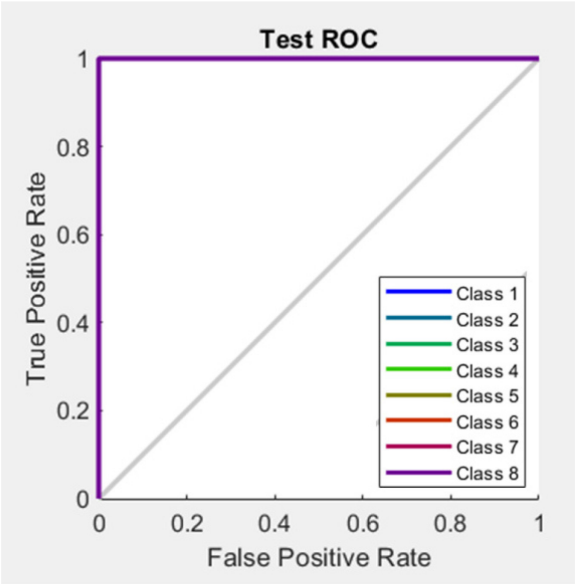


Figure 7. ROC for the configuration of trainbr with seven hidden layers and 70-15-15 ratio

4. CONCLUSION

The dataset comprising HD-sEMG signals obtained from the anterior and posterior forearm can indeed be effectively classified utilizing a neural network. This successful classification was achieved through the implementation of the Bayesian regularization backpropagation (trainbr) training function, incorporating seven hidden layers, and employing a training-validation-testing ratio of 70-15-15 with 100% accuracy classified data and 1.4923 $x10^{-5}$ MSE. Subsequent investigations could delve into assessing the adaptability and resilience of the established neural network classification model when confronted with variances in electrode positioning and user-specific variables. Such inquiries would further augment its pragmatic applicability. Moreover, broadening the scope of analysis to encompass more intricate hand movements and a wider array of muscle groups could provide valuable insights regarding the extended applicability of the HD-sEMG-based methodology.

REFERENCES

[1] A. R. Brown, W. Pouw, D. Brentari, and S. J. P. S. Goldin-Meadow, "People are less susceptible to illusion when they use their hands to communicate rather than estimate," vol. 32, no. 8, pp. 1227-1237, 2021.

[2] N. J. Jarque-Bou, J. L. Sancho-Bru, and M. J. S. Vergara, "A systematic review of emg applications for the characterization of forearm and hand muscle activity during activities of daily living: results, challenges, and open issues," vol. 21, no. 9, p. 3035, 2021.

[3] F. Kristjansdottir, L. B. Dahlin, H.-E. Rosberg, and I. K. J. J. o. H. T. Carlsson, "Social participation in persons with upper limb amputation receiving an esthetic prosthesis," vol. 33, no. 4, pp. 520-527, 2020.

[4] Z. Chen, H. Min, D. Wang, Z. Xia, F. Sun, and B. J. B. Fang, "A Review of Myoelectric Control for Prosthetic Hand Manipulation," vol. 8, no. 3, p. 328, 2023.

[5] M. M. Suhaimi, A. S. Ghazali, A. Jazlan, and N. J. C. E. Sidek, "Analysis of High-Density Surface Electromyogram (HD-sEMG) signal for thumb posture classification from extrinsic forearm muscles," vol. 9, no. 1, p. 2055445, 2022.

[6] E. L. Le, M. L. Iorio, M. A. J. E. J. o. O. S. Greyson, and Traumatology, "Targeted muscle reinnervation in upper extremity amputations," pp. 1-9, 2023.

[7] N. Parajuli *et al.*, "Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," vol. 19, no. 20, p. 4596, 2019.

[8] P. M. Kloskowska, "The biomechanical determinants of sports related groin pain in athletes," Queen Mary University of London, 2016.

[9] A. Asghar, S. Jawaid Khan, F. Azim, C. S. Shakeel, A. Hussain, and I. K. J. P. o. t. I. o. M. E. Niazi, Part H: Journal of Engineering in Medicine, "Review on electromyography based intention for upper limb control using pattern recognition for human-machine interaction," vol. 236, no. 5, pp. 628-645, 2022.

[10] X. Jiang *et al.*, "Optimization of HD-sEMG-based cross-day hand gesture classification by optimal feature extraction and data augmentation," vol. 52, no. 6, pp. 1281-1291, 2022.

[11] H. Duan, C. Dai, and W. J. M. P. i. E. Chen, "The Evaluation of Classifier Performance during Fitting Wrist and Finger Movement Task Based on Forearm HD-sEMG," vol. 2022, 2022.

[12] Y. Li, J. Li, P. Tu, H. Wang, and K. J. I. S. J. Wang, "Gesture Recognition Based on EEMD and Cosine Laplacian Eigenmap," 2023.

[13] S. Siecinski, P. S. Kostka, and E. J. Tkacz, "Time domain and frequency domain heart rate variability analysis on electrocardiograms and mechanocardiograms from patients with valvular diseases," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 653-656: IEEE.

[14] F. Liviero *et al.*, "Persistent Increase of Sympathetic Activity in Post-Acute COVID-19 of Paucisymptomatic Healthcare Workers," vol. 20, no. 1, p. 830, 2023.

[15] S.-i. Kim, Y. Noh, Y.-j. Kang, S. Park, and B. J. J. o. t. C. S. E. I. o. K. Ahn, "Fault classification model based on time domain feature extraction of vibration data," vol. 34, no. 1, pp. 25-33, 2021.

[16] S. N. Sidek, M. R. Roslan, S. Sidek, and M. S. M. J. I. J. o. C. I. S. Khalid, "Thumb-tip force prediction based on Hill's muscle model using electromyogram and ultrasound signal," vol. 11, no. 1, pp. 238-247, 2018.

[17] E. Aguiar-Salazar *et al.*, "Intelligent Electromyograph for Early Detection of Myopathy and Neuropathy Using EMG Signals and Neural Network Model," in *Conference on Information and Communication Technologies of Ecuador*, 2022, pp. 32-45: Springer.

[18] M. Wang *et al.*, "Hand posture and force estimation using surface electromyography and an artificial neural network," vol. 65, no. 3, pp. 382-402, 2023.

[19] E. Kassaw, E. Worassa, K. Fetene, and G. Aboye, "Muscle Fatigue Analysis and Stress Detection from Surface EMG and ECG Using Deep Learning for Upper-Limb Trauma Rehabilitation," 2023.

[20] A. Aranceta-Garza and B. A. Conway, "Differentiating variations in thumb position from recordings of the surface electromyogram in adults performing static grips, a proof of concept study," vol. 7, p. 123, 2019.

[21] X. Bao, Y. Zhou, Y. Wang, J. Zhang, X. Lü, and Z. J. P. o. Wang, "Electrode placement on the forearm for selective stimulation of finger extension/flexion," vol. 13, no. 1, p. e0190936, 2018.

[22] M.-C. Lu *et al.*, "Automatic Classification of Slit-Lamp Photographs by Imaging Illumination," p. 10.1097, 2022.

[23] H. Gupta *et al.*, "A hybrid convolutional neural network model to detect COVID-19 and pneumonia using chest X-ray images," vol. 33, no. 1, pp. 39-52, 2023.

[24] W. Setiawan and R. Rulaningtyas, "Visual explanation of maize leaf diaseases classification using squeezenet and gradient-weighted class activation map," in *AIP Conference Proceedings*, 2023, vol. 2679, no. 1: AIP Publishing.

[25] S. M. Alardhi *et al.*, "Artificial neural network model for predicting the desulfurization efficiency of Al-Ahdab crude oil," in *AIP Conference Proceedings*, 2022, vol. 2443, no. 1: AIP Publishing.

[26] G. Panchal, A. Ganatra, Y. Kosta, D. J. I. J. o. C. T. Panchal, and Engineering, "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers," vol. 3, no. 2, pp. 332-337, 2011.

[27] S. Asthana, A. Pandit, A. J. I. J. o. S. Bahrdwaj, and Technology, "Analysis of Multiple Hidden layers vs. Accuracy in Performance using Back Propagation Neural network," vol. 10, no. 4, pp. 1-4, 2017.

[28] F. S. Panchal, M. J. I. J. o. C. S. Panchal, and M. Computing, "Review on methods of selecting number of hidden nodes in artificial neural network," vol. 3, no. 11, pp. 455-464, 2014.

https://doi.org/10.31436/iiumej.v25i2.3094

PERFORMANCE ANALYSIS OF AN INDUCTION MOTOR COUPLED VIF WITH MR FLUID DAMPER

SYED MUNIMUS SALAM, MUHAMMAD MAHBUBUR RASHID*

Department of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

**Corresponding author: mahbub@iium.edu.my*

(Received: 3 November 2023; Accepted: 6 March 2024; Published online: 15 July 2024)

ABSTRACT: The flywheel is a classic mechanical device that is often used to enhance the rotational motion of engines. Electrical machines often face fluctuations in speed that hamper speed stability and cause extra power consumption. Recently, a few publications have analyzed the impact of the flywheel to reduce fluctuation and energy consumption in electric motors. This study proposes the use of a flywheel with a changeable moment of inertia, which can be manipulated to boost both speed stability and energy efficiency. The objective is to improve the speed stability of industrial motors in the presence of the loading effect. This study introduces a magneto-rheological variable inertia flywheel (MRVIF) to control rotational speed and reduce power usage. The purpose of analytical development is to assess the influence of rotational speed and excitation current on the MR damper's moment of inertia for control purposes. The investigation focuses on the analysis of power usage and stability across different power inputs and rotating speeds. The effectiveness of the suggested MRVIF was evaluated via the development of a prototype. Experiments were undertaken to validate the effectiveness and stability of the system. The findings illustrate the potential use of MRVIF in reducing energy consumption and enhancing speed stability.

ABSTRAK: 'Flywheel' atau roda tenaga adalah peranti mekanikal klasik yang sering digunakan bagi meningkatkan gerakan putaran enjin. Mesin elektrik sering menghadapi turun naik kelajuan yang menghalang kestabilan kelajuan dan menyebabkan penggunaan kuasa tambahan. Baru-baru ini, terdapat beberapa kajian terdahulu yang menganalisis kesan roda tenaga bagi mengurangkan turun naik dan penggunaan tenaga dalam motor elektrik. Kajian ini mencadangkan penggunaan roda tenaga dengan momen inersia boleh ubah, di mana ia boleh dimanipulasi bagi meningkatkan kestabilan kelajuan dan kecekapan tenaga. Objektif kajian adalah bagi meningkatkan kestabilan kelajuan motor industri bersama kesan muatan. Kajian ini memperkenalkan pembolehubah magnetorheologikal roda tenaga inersia (MRVIF) bagi mengawal kelajuan putaran dan mengurangkan penggunaan kuasa. Tujuan pembangunan analitikal ini adalah bagi menilai pengaruh kelajuan putaran dan arus pengujaan pada momen inersia peredam MR bagi tujuan kawalan. Kajian memfokuskan pada analisis penggunaan kuasa dan kestabilan merentas pelbagai input kuasa dan kelajuan putaran. Keberkesanan MRVIF yang dicadangkan telah diuji melalui pembangunan prototaip. Eksperimen dijalankan bagi mengesahkan keberkesanan dan kestabilan sistem. Penemuan ini menggambarkan potensi MRVIF dalam mengurangkan penggunaan tenaga dan meningkatkan kestabilan kelajuan.

KEYWORDS: *Variable Inertia Flywheel; Energy Saving; Magneto Rheological Fluid.*

https://doi.org/10.31436/iiumej.v25i2.3094

1. INTRODUCTION

A flywheel with variable inertia requires a more complex design compared to one with constant inertia. For a flywheel to possess a controlled moment of inertia along its rotating axis, it is necessary for the flywheel to have a flexible geometry around it. Furthermore, any changes made to the geometry of the flywheel need the movement of parts inside it. Because of its higher level of complexity, a variable inertia flywheel (VIF) may make more efficient use of energy saved per unit than a conventional fixed inertia flywheel. Nevertheless, when the flywheel of a machine is substituted with the VIF, the additional mass of the VIF might still lead to a configuration that has a higher energy density in comparison to its fixed inertia equivalent [1].

The coefficient of variance is a measure used to characterize the variation in angular velocity of a flywheel across a cycle relative to its average angular velocity. A system using a conventional flywheel is used to minimize fluctuations in angular velocity. A high moment of inertia is necessary to get a low coefficient of fluctuation. An adjustable inertia flywheel might be quite advantageous in this situation [2].

The energy-saving capacity of the flywheel has gained significant interest in recent years. Jauch developed a flywheel energy storage system that was integrated into the rotor of a wind turbine [3]. Another study developed a flywheel that was directly coupled to an engine and had a changeable inertia due to the use of springs. The flywheel offers a wide range of energy storage choices while simultaneously preserving its rotational speed [4]. The double-mass flywheel incorporated a centrifugal pendulum devised by Ishida et al. [5]. The variable inertia developed by the fluid of a flywheel proposed by James et al. had the capability to maintain a consistent rotational speed across a wide range for energy storing [6]. According to Yuan L et al., the variable inertia flywheel may take the role of the conventional fixed inertia flywheel [7]. In a similar study, a flywheel was considered without any fixed adjustment to the system's input shaft, along with a variable moment of inertia of the mass. By utilizing a fluid that behaved like a series of inertia masses, Matsuoka presented new vibration reduction systems. On the other hand, the usage of magnetorheological fluid in flywheels has not been very common [8]. The energy-saving potential of VIMRF (variable inertia magnetic rheological flywheel), in contrast to competitors, has not yet been fully explored. Alternative methods are developing with respect to research progress. Without taking into account the impact of high system stability, the majority of the research in [9] concentrated on the design and implementation of VIF in the field of power savings and stability.

Improving overall operating efficiency and sustainability in industrial motors requires addressing the problems of speed stability and energy usage. Increased speed stability lowers the possibility of interruptions and defective products in industrial operations by ensuring steady and dependable performance. Simultaneously, reducing energy use reduces operating expenses and supports international environmental responsibility programs. Improved motor efficiency supports regulatory compliance and places companies positively in the context of changing environmental requirements by helping to minimize carbon emissions and energy waste. Overall, putting energy efficiency and speed stability first in industrial motors not only makes them more economically viable but also shows that the company is committed to using sustainable and ethical manufacturing techniques.

The inertia control of a VIF may be unpredictable under conditions of load disruptions, which can cause instability and system failure. The possible weakness of the strategies is the lack of concentration on inertia control. All previous variable inertia control designs and methodologies do not even provide stable control without considering the consequences of the

applied device. Lack of control on a variable inertia flywheel's moment of inertia causes additional energy loss. Modeling a system with proper control of variable inertia is challenging. Moreover, implementing a strong adaptive control approach for any unexpected changes in the applied machine with significant stability and energy savings is more crucial.

The main objective of this study is to analyze the performance of a developed MR VIF coupled induction motor prototype to find out the impact of the MR fluid properties on energy consumption and speed stability of the induction motor.

2. THE METHOD OF VARIABLE INERTIA CONTROL OF A FLYWHEEL

A flywheel with variable inertia inherently requires a significantly intricate construction compared to a flywheel with static inertia. The change in velocity of an object's mass produces a force of inertia that counteracts the object's flexibility, leading to the generation of preserving forces. Consequently, when external disturbances affect a body, its inertia and inner preserving forces result in unwanted oscillation. The kind of vibration (linear, torsional, or bending) depends on the design of the machine. Vibration poses a substantial issue in all mechanical systems, as it may have adverse effects on the machine's durability, product excellence, and overall safety. In addition, it has the potential to cause structural disruption [10]. Hence, the act of reducing and regulating vibrating is an essential effort. The essential elements of vibratory systems consist of the spring, damper, and mass. The efficient design, optimization, and exact production of these three components are crucial to the development of an efficient vibration isolation system [11].

Vibration is a frequent motion phenomenon that often oscillates around a balancing position in both the flywheel and inertia mass. In some normal scenarios, vibration can be required when it comes to anything from little devices like loudspeakers to huge ones like tamping machines, vibrating conveyors, and sieves. High levels of vibration may sometimes be advantageous and desired. However, vibration in VIF is generally undesirable because it makes annoying noise, consumes energy, leads to mechanical wear and structural stress, and increases risk. As a result, experts have devoted a lot of time to studying vibration control. Here, the term "vibration control" refers to the reduction of vibration to lessen unwelcome oscillations in a protected structure. Typically, a compact connecting element made up of a spring, damper, or mass accomplishes this [12].

Various ideas for reducing vibration exist for managing vibrations of the specified sort, depending on the need for external power or information. The exploration of three vibration-controlling mechanisms, namely passive, semi-active, and active systems, has taken place in recent decades. These systems are often known as vibration control systems [13]. Vibration is typically undesirable when it causes annoying commotion, results in peril, or dissipates energy. As a result, researchers have always been interested in vibration control. Research in recent decades has mostly focused on active, passive, and semi-active vibration control systems. The stiffness of these components greatly influences their reaction, including responding time and operational frequency range. Semi-active control is characterized by a reduced reliance on outside force as control energy, resulting in increased stability compared to active control. Additionally, it has a broader range and greater adaptability compared to passive control. Hence, semi-active control systems are very intriguing and find application in several technical fields. For semi-active control, variables can be mass, rigidity, or damping [14].

The earliest vibration control technique is passive control, which receives no energy or information from the outside world. A passive system has the benefits of simple construction,

cheap cost, stability, and high dependability. Still, it also falls short since it can't adapt its structural parameters to changes in vibration over time and can't adapt its design to different excitation frequencies. Active control uses an actuator, driven by external energy and control information, to create an additional vibration channel. This enables the exact and quick modification of the external force of an outside interruption for the purpose of reducing vibration. This gadget, under an active control system, has the capacity to augment or diminish mechanical energy immediately. Specifically, the main vibration is paired with an input that has the same amplitude but the opposite phase, canceling both vibrations [15].

Semi-active devices are unable to input energy into a system that they are controlling directly. This crucial requirement makes sure that changes in the system parameters do not result in parametric excitation. When employing the word "semi-active" in literature, this is often not taken into appropriate consideration. It has been shown to function better than a system with passive control and is utilized as an effective method to reduce unwelcome oscillations in many applications. Active vibration control is, however, overshadowed by the system's relative expense, complexity, and instability [16].

A Semi-active damper can be called a hydro-mechanical controlling system that could modulate energy dissipation with little electrical power consumption. For instance, hydraulic dampers are constructed to exert a force that is directly proportionate to the velocity. These devices enhance the range of damping characteristics that may be achieved with a regular (passive) damper. Various energy-dissipating gadgets have been used to provide the necessary damping force for semi-active damping control.

A magnetorheological (MR) or electrorheological (ER) damper is composed of a hydraulic cylinder that contains polarizable particles in an MR or ER fluid, often oil, with particle sizes in the micron range. Non-Newtonian fluids, such as MR and ER fluids, undergo changes in their properties when subjected to electric or magnetic fields. Micron-sized iron particles, when suspended in a carrier fluid such as water, petroleum-based oil, or silicon-based oil, may change the flow properties of the fluid by arranging themselves in chain-like patterns along the lines of magnetic or electrical fields. Both MR and ER materials may rapidly transition from a fluid state to a semi-solid state within milliseconds when exposed to a magnetic or electric field. Since they don't have any moving components, MR and ER dampers are mechanically dependable [17].

The moment of inertia grows when the flywheel spins rapidly because the variable masses have shifted to the flywheel's radius. However, if the flywheel's varying masses are allowed to oscillate freely, the system will cause vibration and instability. Spring forces are applied as a dampening force to prevent this from happening. There is not enough resistance here to maintain a steady state. Low-speed operation further complicates mass position control. By introducing MR fluid into the cavity, the viscosity may be modulated by external stimulation, allowing for precise manipulation of the mass's location.

3. VIF MODELLING WITH MR FLUID

When the slider is moved along the slot, the flywheel's inertia changes accordingly. Figure 1 is a diagrammatic representation of the variable inertial flywheel. When the flywheel is spinning at 0 RPM, the springs are at full extension and contacting the hub. In this case, we ignore the pull of gravity. The springs are fixed in their length and can only be compressed. The VIF is composed of a compact circular disc, including four sliders and four springs. The spinning of the flywheel induces the displacement of slides along the slots. The compression

https://doi.org/10.31436/iiumej.v25i2.3094

of the springs up to their maximum length is determined by the speed of the flywheel. The movement of the sliders inside the slots causes the moment of inertia of the flywheel to vary.

Considering the slider's moment of inertia, it can be expressed for the slider's own mass center, while the flywheel fixed structural part's moment of inertia around the shaft center is

$$J_{slide} = \frac{1}{4} m_{slide} \left(\frac{D_{slide}^2}{4} + \frac{L_{slide}^2}{3} \right) \quad (1)$$

$$J_{flywheel} = J_{solid} + n_{slide} * J_{slide} \quad (2)$$

where m_{slide}, D_{slide} and L_{slide} are the mass, diameter, and length of the slider. n_{slide} represents the number of slots available in the flywheel.

The following expression describes the calculation procedure for the moments of inertia of the solid part of the flywheel, considering a fan-shaped body.

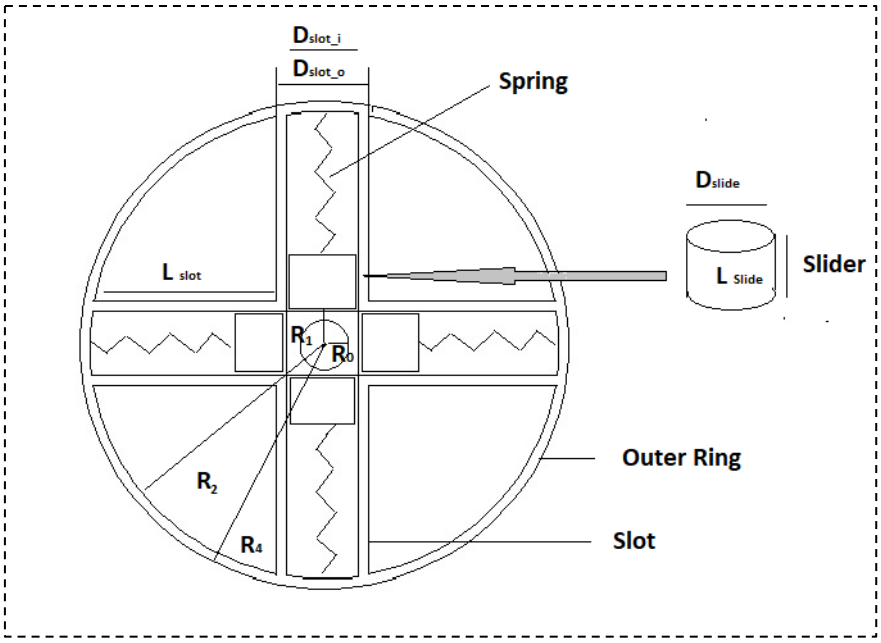


Figure 1. Schematic diagram of VIF system.

$$J_{full} = \int_{R_0}^{R_4} r^2 dm = \int_{R_0}^{R_4} 2\pi H r^3 \rho dr = \frac{1}{2}\pi H \rho (R_4^2 - R_0^2) \quad (3)$$

$$J_{hole} = \int_{R_1}^{R_2} r^2 dm = \int_{R_1}^{R_2} 2\pi H \left(r - \frac{\sqrt{2}}{2} D_{slot_O} \right) r^2 \rho dr$$

$$= \pi H \rho \left(\frac{1}{8} (R_2^2 - R_1^2) - \frac{\sqrt{2}}{12} D_{slot_O} (R_2^2 - R_1^2) \right) \quad (4)$$

Again, here moment of inertia of the slot is calculated as

$$J_{slot} = \frac{1}{12} \rho (L_{slot}^2 + D_{slotI}^2)^2 \quad (5)$$

Finally, the value of J_solid can be obtained by,

$$J_{solid} = J_{full} - n_{slide} * (J_{hole} + J_{slot}) \quad (6)$$

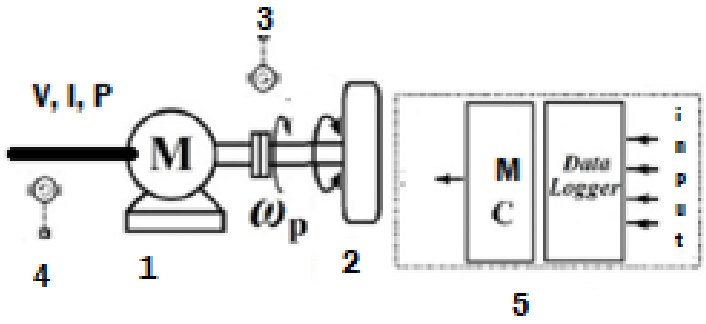
As the flywheel spins, the slider slides along the slot.

https://doi.org/10.31436/iiumej.v25i2.3094

4. Simulation Modelling of The System

In the second step, a simulation system has been developed for testing the proposed theory. To analyze the complex changing aspects of the induction motor with the MR VIF and the VIF, the system model is constructed utilizing the systematic bond graph approach [18].

This method is a self-contained construction of a conventional method that relies especially on a constant type of energy interchange in an integrated manner. The sliders' maximum displacement is limited by the fixed outside boundary ring of the constructed flywheel, which is closely attached to the main frame of this flywheel. The sliders are housed inside the cylindrical slots. The sliders are encircled by a coil at the outer periphery. The condition of the MR fluid inside the slot undergoes a change when it is stimulated by external power. During the normal VIF operation, the cylindrical springs are extracted while the sliders inside the cylinder are securely immobilized with the frame of the flywheel. The flywheel rpm is detected using a special non-contact rpm sensor and is communicated to the data logger.



System Componenets

1 Induction Motor
2 Flywheel
3 Speed Sendor
4 Voltage, Current, Power, Temperature Sensor
5 MR Exciter

Figure 2. Schematic diagram of IM coupled VIF system.

In the Figure 2, the schematic diagram of the system utilized for the IM coupled VIF is illustrated. The basic system model consists of the induction motor (1) and the flywheel (2). The system parameters that are considered as inputs are voltage current and power, and the output parameters that are observed for control are the speed and temperature. The Magneto rheological cylinder excitation can change the viscosity of the MR fluid, which can control the VIF semi-actively. In addition to the structural characteristics of the flywheels, the model considers the qualities of MR fluid, variations in viscosity, and the induction motor. The following assumptions are made:

- The fluid inertia is ignored.
- The fluid is classified as Newtonian.
- The spring's properties are assumed to be linear, and the model does not consider the dynamics of the springs.
- The gravitational force acting on the slider is not considered.

https://doi.org/10.31436/iiumej.v25i2.3094

The effect of excitation current and rotational speed on the moment of inertia of a Magnetorheological (MR) damper is typically investigated using both computational and experimental techniques. Dynamic testing sensors, which measure the moments of inertia produced by exposing the damper to a range of rotational speeds and excitation currents, are a common component of experimental approaches. Computational techniques replicate the behavior of the MR damper under various operating situations using simulation modeling and finite element analysis. Through the optimization and design of MR damper systems, these analytical techniques assist in understanding how variations in rotational speed and excitation current affect the moment of inertia.

Notably, for modeling electromagnetic machines and their rotating magnetic fields, the support for vector variables is crucial. Existing bond graph simulation tools only accommodate scalar variables, prompting our proposal for an innovative approach. We suggest introducing support for complex variables and vectors in bond graphs to facilitate the modeling of polyphase electromagnetic and spatial vector systems. To address this, we have introduced a rotary gyrator element and incorporated it with a switched junction to formulate a toolbox supporting complex and vector variables. This approach is implemented by creating a complex S-function toolbox in Simscape within the MATLAB environment, chosen to leverage the speed of S-functions, the user-friendly nature of Simulink, and the widespread popularity of MATLAB. The spring remains undeformed at its minimal length ($lmin$).

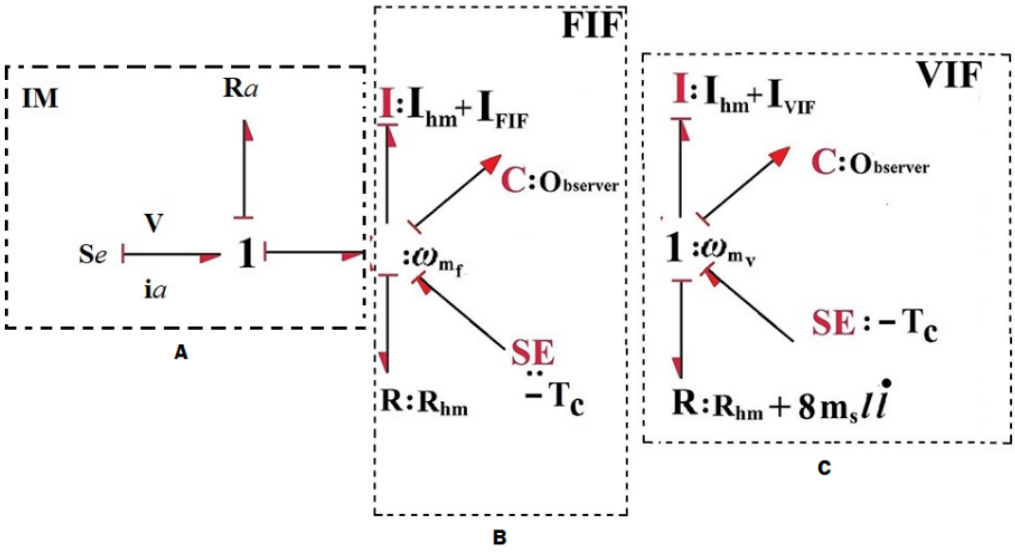


Figure 3. Bond graph representation

Figure 3 depicts the bond graph concept model of the induction motor drive system, including the flywheel. The sub-models of the induction motor, flywheel inertia factor (FIF), and voltage inverter factor (VIF) are labeled as A, B, and C, respectively, to enhance clarity.

The source of flow (SF) element is the electrical power that develops the rotary velocity of the motor (ωmf). A motor-coupled shaft converts electrical energy into mechanical energy or vice versa. Ultimately, a Simulink simulation has been created to assess the MR VIF system using the bond graph correlation and equation. Figure 4 depicts the system, and simulations have been conducted under various scenarios. The data has been compared to the authentic data generated using a similar configuration, and subsequently, an error analysis has been conducted to enhance the results further.

https://doi.org/10.31436/iiumej.v25i2.3094

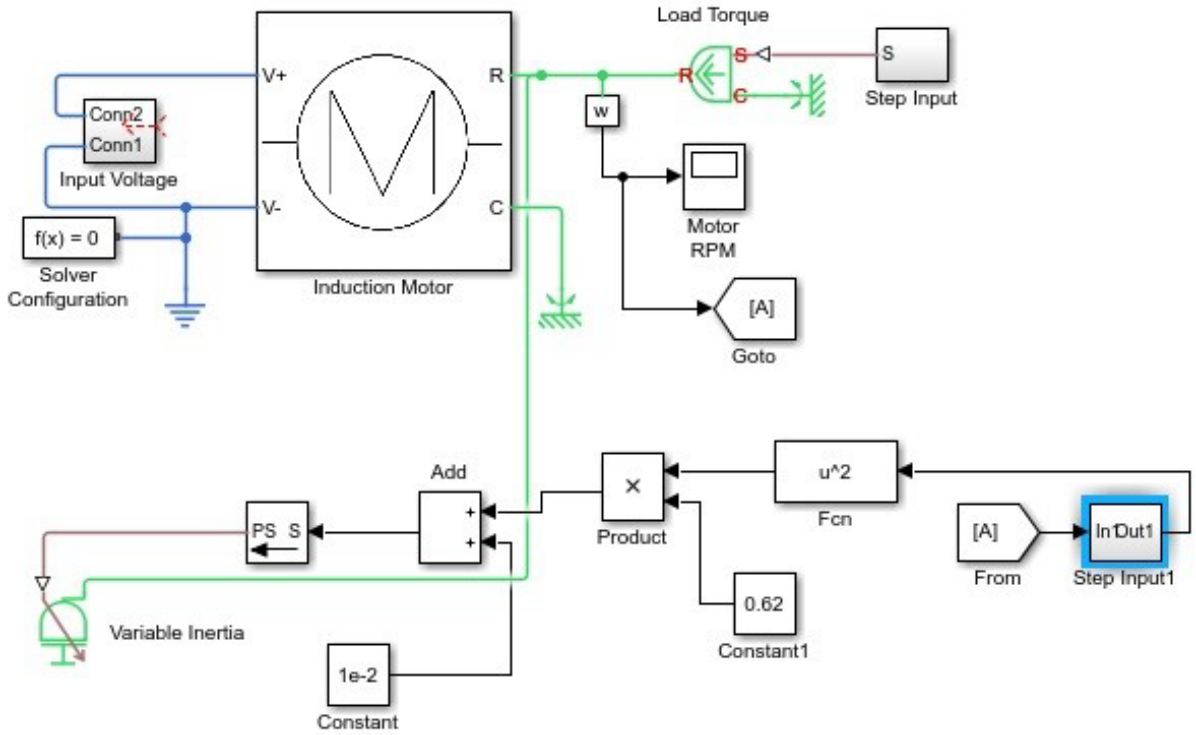


Figure 4. Simulink simulation representation

5. PROTOTYPING A MR VIF SYSTEM COUPLED WITH IM

A prototype made of aluminum was created in the beginning for testing purposes. The primary motivation for using aluminum was to prevent magnetic field induction. The fly-wheel cage was designed by the criteria given in Table 1. In certain circumstances, the value of some parameters was modified for analysis purposes. An induction motor applies the flywheel system.

Table 1. Construction details of MRVIF

Parts and Accessories	Properties
Frame	Aluminium Alloy
Slider	Iron casting
Cylinder	Stainless Steel
Spring	Steel
MR Fluid	Black Iron Dust with Silica Oil
Coil	Copper
Cable	DC cable
Brush	Carbon Brush
Base	Steel sheet

The characteristics of the electrical system have been tested using an electric multi-meter, and the system's speed is controlled by a converter. The converter primarily adjusts current and power to bring about the desired change in velocity. Figure 5 displays the physical system that is being examined in this research. A variable-speed electric motor controller is used to operate an induction motor in the system.

The induction motor has been connected to the flywheel constructed from an aluminum alloy (1), which is positioned on the base (5). The changing inertia mass is contained within the cylinder housing the MR (Magnetorheological) fluid (2). To excite the MR fluid, a slip ring

https://doi.org/10.31436/iiumej.v25i2.3094

(3) is attached to the carbon brush to supply power to the coil within the cylinder, thereby modifying the properties of the MR fluid (4).

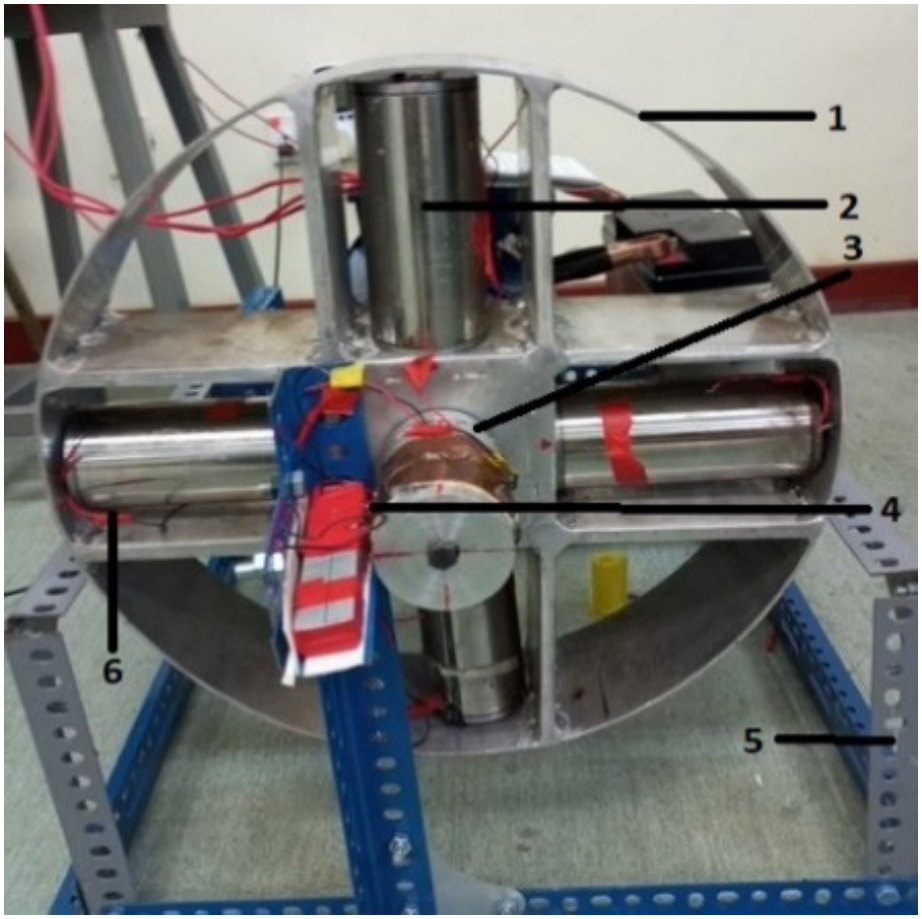


Figure 5. Prototype model representation

To ensure the maintenance of the MR fluid, each cylinder is equipped with an additional cover (6). The controller adjusts the current consumption to modulate and change the speed of the motor. It can both reverse and forward the direction of the motor's speed. The induction motor is connected to a flywheel.

The induction motor speed is regulated by supplying variable power flow via regulation of the command signal applied to the controller. The primary function of the induction motor is to power the inertia load. An analysis is conducted on the system's performance at different speeds, including both constant and changing inertia loads. Figure 6 shows the working procedure of the MR VIF with the flow chart.

To achieve the desired configuration of a cylinder with variable inertia, a hollow cylinder is equipped with a spring and a mass. The flow chart in Figure 6 shows the specific working process applied to the flywheel and controlling requirement of the system. The cylindrical structure is comprised of an internal volume containing magnetorheological (MR) fluid. The MR fluid is synthesised in a laboratory setting by combining black iron oxide (Fe_3O_4) dust with silica oil as a carrier. The resulting mixture utilises a semi-active substance, which exhibits a viscosity ranging from 0.21 to 0.27 Pas at a temperature of 30 degrees Celsius. To manipulate the MR characteristics of the fluid, an externally generated magnetic field is used as a means of exciting the fluid. Additionally, the viscosity was assessed inside the laboratory setting,

revealing a variation of 0.06 Pas when a magnetic field was applied. The table includes all the components used for manipulating inertia.

For data collection, a lab set up was installed with different data collection equipment. As discussed earlier, the electrical readings were taken by integrated multi electric meter. The speed of the system was measured using a digital tachometer. For avoiding any confusion and error, similar data was taken several times

6. PERFORMANCE ANALYSIS OF THE MODEL

To validate the model with three different viscosities of the MR fluid, the rpm of the induction motor was varied from 300 to 900 rpm. The simulated result and the experimental data were analyzed using the rpm change at the start of the induction motor. Three cases were considered for the system model:

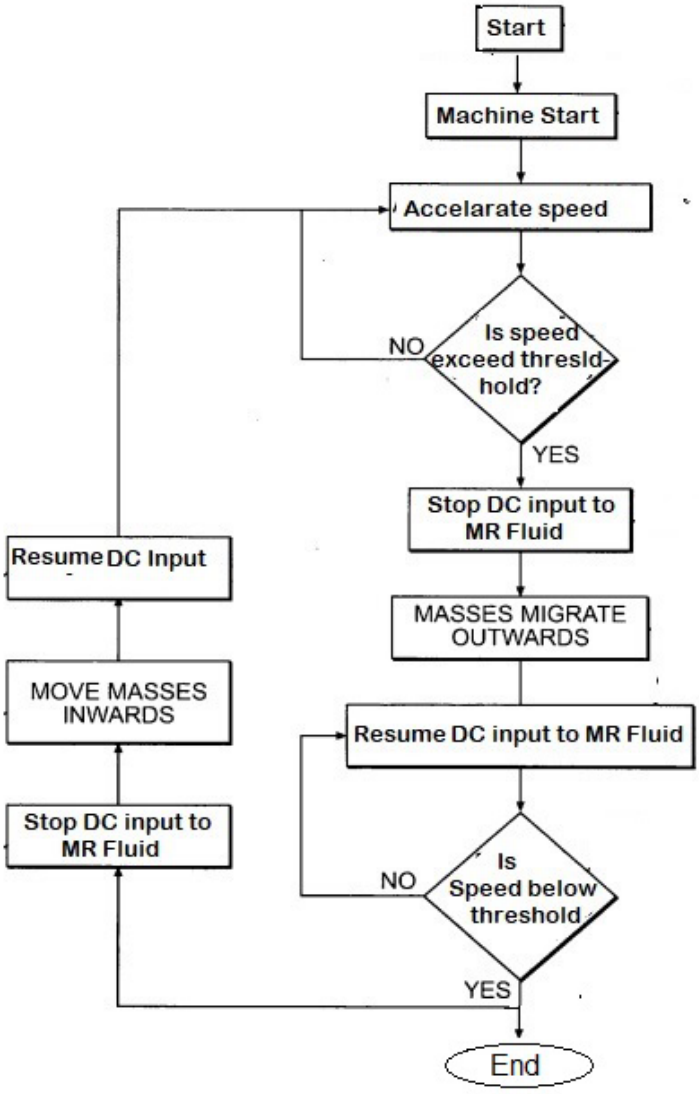


Figure 6. Flow chart used with the system model

Case 1: The motor started and the speed increased up to the highest possible changes in the supplied power to the induction motor driving the VIF with low excitation current with viscosity 0.21 to 0.23 Pas. This was achieved by giving the input power from the controller applied to the induction motor when the speed increment was done.

https://doi.org/10.31436/iiumej.v25i2.3094

Case 2: In this case, the speed of the motor driving the VIF with regular temperature and with medium viscosity is 0.24 to 0.25 Pas. It was achieved by giving the input power from the controller applied to the induction motor, then the speed increment was similarly done.

Case 3: In this case, the rpm of the induction motor is altered during running the VIF with a relatively high viscosity of 0.25 to 0.27 Pas at a regular temperature. This is achieved by applying the input power from the controller to the induction motor and then incrementing the speed similarly.

6.1. Speed Stability Analysis of the Model

A comparative analysis was conducted between the MR VIF and VIF setups. When the VIF was set up with MR fluid, the simulation results showed that the inclusion of MR fluid reduced oscillation, leading to improved speed stability and reduced overshooting. The viscosity value was augmented by the incremental damping factor for simulation purposes. The variation in the speed buildup characteristics is illustrated in Figure 7. The excitation amplitude for three separate situations was chosen to remain constant during the observation period. The induction motor was maintained at an optimal level, neither too low nor too high, as the slides could not overcome friction if the speed were low.

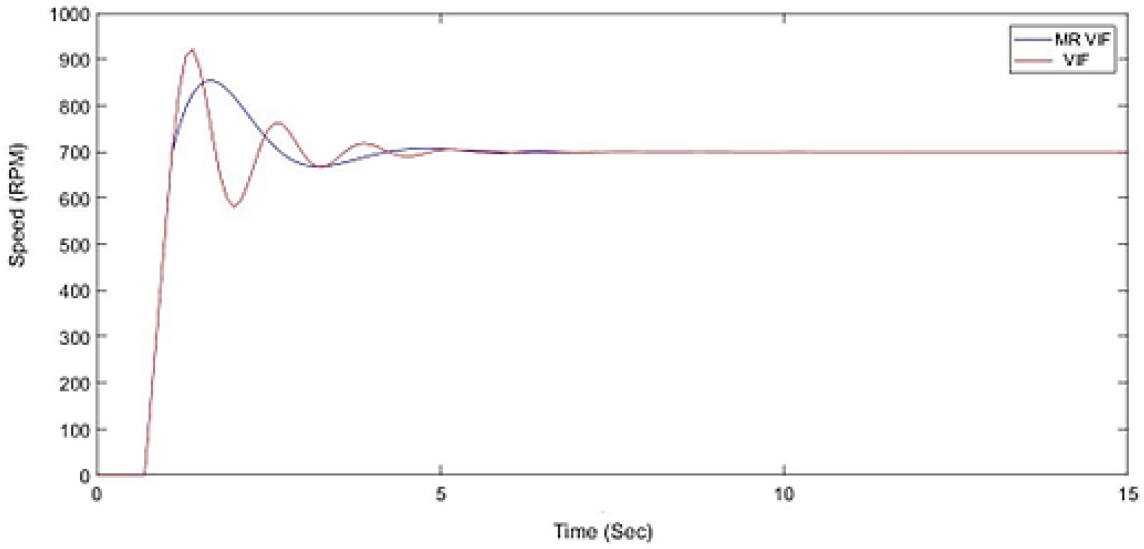


Figure 7. Speed stability analysis for MR VIF and VIF

If the speed was excessively high, there may not be sufficient viscosity to keep the sliders in the VIF and the entire flywheel would function as FIF. Hence, the maximum and minimum speed values were determined through multiple iterations of the experiment. The relative investigation of the output rpm of the induction motor driving the MRVIF and the VIF was conducted using the same process. Numerical solutions for the model equations provided in the section are obtained using software [19].

The validation of the system model was achieved through a direct comparison between the measured responses and the expected outcomes. During the testing, the input power remained constant at the early stage. Figure 7–9 illustrates the contrasting responses of the induction motor speed when driven by VIF and FIF in various scenarios. When the induction motor operated with the MR VIF, the inertia of the flywheel fluctuated in response to the motor's changing speed. Thus, when comparing the FIF to the MR VIF, the induction motor's top speed was decreased due to the growing inertial load.

https://doi.org/10.31436/iiumej.v25i2.3094

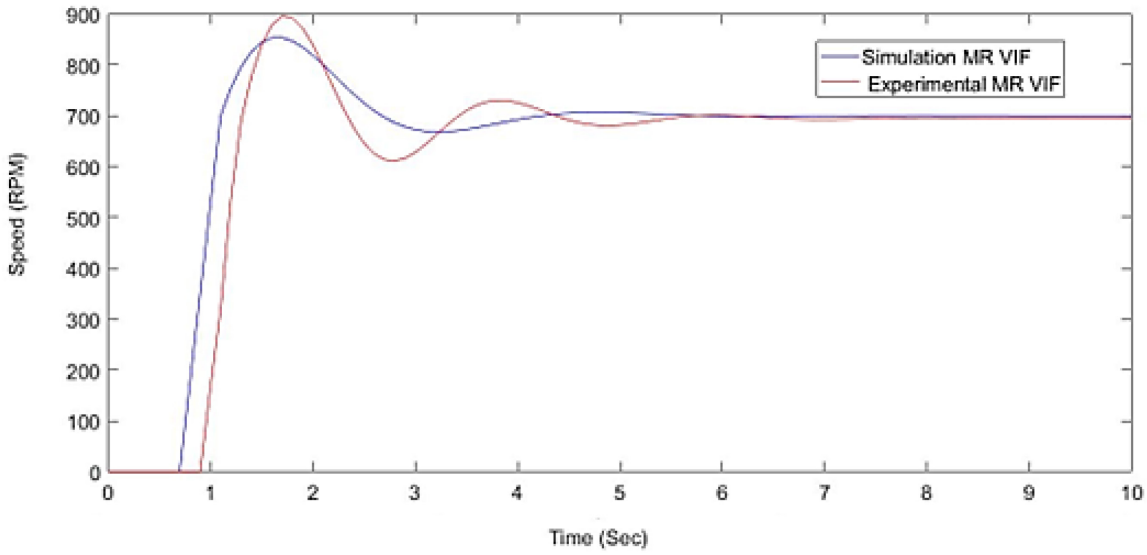


Figure 8. Speed stability analysis for simulation and experimental MRVIF

Figure 8 illustrates the comparison between the expected speed and the experimental speed of the induction motor while it was driving the MR VIF during the starting phase. The inertia of the flywheel changed in response to the varying speed of the induction motor when it was being driven using the VIF. Consequently, the peak speed of the hydraulic motor that drove the VIF was reduced due to the growing inertial load compared to the FIF. Figure 7 presents a comparison between the projected speeds using MR VIF and VIF. The Russell error approach was employed to approximate the discrepancy between the experimental and simulated responses. Equations (7) and (8) were utilized to calculate the magnitude, phase, and comprehensive errors for the speed response of the VIF and the MR VIF, respectively [20]. The magnitude error (e_m) of the experimental and simulation data was determined by,

$$\mathrm{e_m} = \mathrm{Sgn}(\mathrm{e_{rm}})\log_{10}(1 + |\mathrm{e_{rm}}|) \tag{7}$$

and

$$\mathrm{e_{rm}} = \frac{\sum_{k=1}^{n} \mathrm{M_k} - \sum_{k=1}^{n} \mathrm{D_k}}{\sum_{k=1}^{n} \mathrm{M_k}\ \mathrm{X} \sum_{k=1}^{n} \mathrm{D_k}} \tag{8}$$

where, $\mathrm{e_{rm}}$, M, and D are the relative magnitude of error and simulation data magnitude, and experimental data magnitude. The Simulink simulation and the real-time experiment were performed considering a sample time of 10ms, and a minimum sample rate was maintained at 1000. Equations (7) and (8) were used to estimate the magnitude error (e_m) as 1.411 and 0.0352 for the MR VIF and the VIF speed response, respectively.

https://doi.org/10.31436/iiumej.v25i2.3094

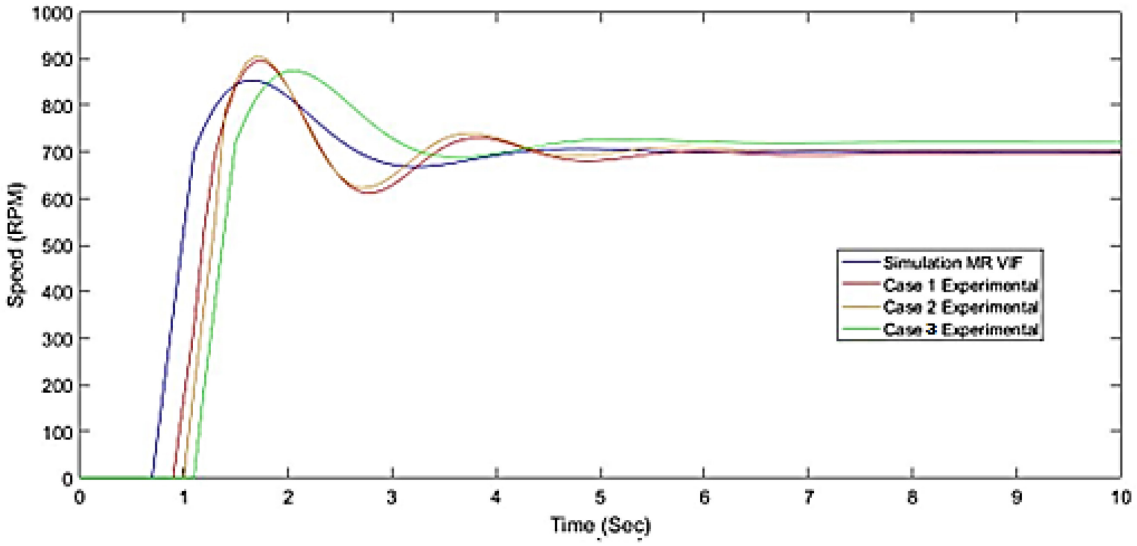


Figure 9. Speed stability analysis for simulation and experimental MRVIF for different case studies

6.2. Power Consumption Analysis of the Model

Two distinct methodologies were employed to determine energy consumption: a short-term approach and a long-term approach, as shown in Figure 10. The short-term difference computation included capturing and documenting readings using camera recording for a duration of 15 to 30 seconds. Subsequently, the recorded values were manually documented. The system operated continuously for an extended duration of around thirty minutes, after which the aggregate power consumption was computed. To examine long-term energy usage, a motor linked with a Fixed Inertia Flywheel (FIF) was operated at various speeds for a duration of thirty minutes. The digital energy meter collected all power consumption data, and the resulting difference is shown in Table 2. The data clearly indicates that there was a significant reduction in power consumption for all scenarios when the magnetorheological (MR) fluid was used to alter the mass position in the variable inertial flywheel (VIF).

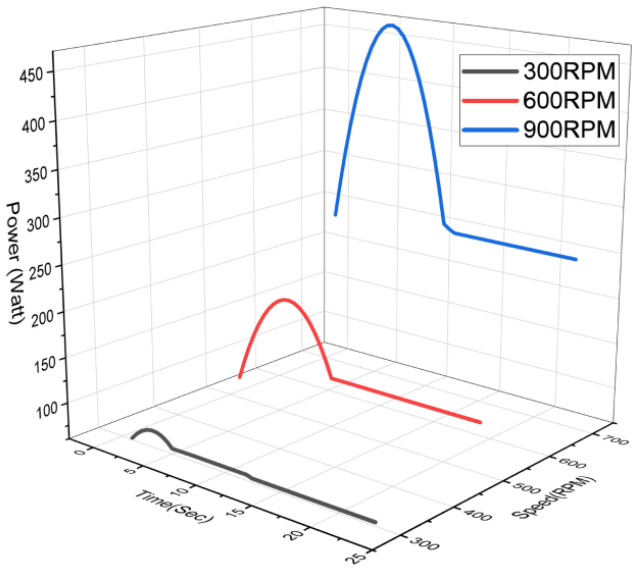


Figure 10. Power consumption representation for different speed of MR VIF

https://doi.org/10.31436/iiumej.v25i2.3094

Table 2. Power consumption analogy with and without MRVIF

Speed RPM	Energy	VIF type			
		VIF	MR Fluid VIF Case 1	MR Fluid VIF Case 2	MR Fluid VIF Case 3
300	Consumed energy in half an hour (kwhr)	44.21	42.37	42.13	42.07
	Maximum power (Watt)	89.9	91.6	92.1	92.3
	Minimum Power (Watt)	86.1	81.2	82.2	81.7
	Average Power (Watt)	88.3	82.85	81.22	81.05
600	Consumed energy in half an hour (kwhr)	112.31	103.64	103.01	101.14
	Maximum power (Watt)	227.6	239.4	239.7	240.1
	Minimum Power (Watt)	217.16	197.1	196.5	196.1
	Average Power (Watt)	220.48	201.47	199.07	198.97
900	Consumed energy in half an hour (kwhr)	237.11	215.55	213.95	211.7
	Maximum power (Watt)	540.13	566.1	569.4	571.2
	Minimum Power (Watt)	436.48	397.21	396.01	395.91
	Average Power (Watt)	464.5	426.2	419.12	416.7

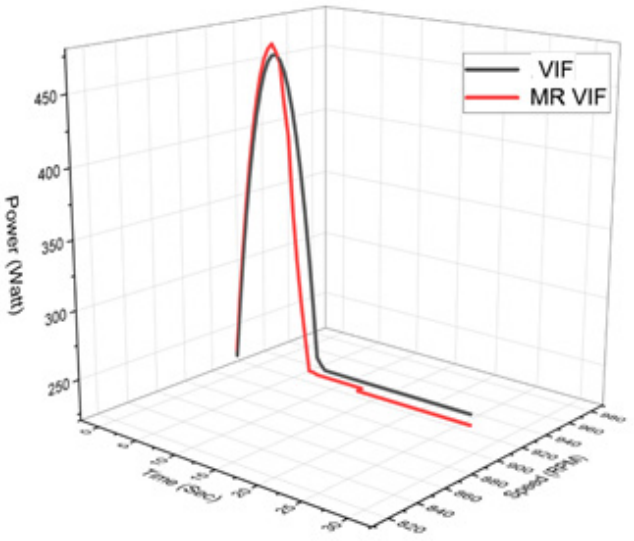


Figure 11. Power consumption representation for 900 RPM speed of MR VIF and VIF

The MR fluid requires activation by an external power source, with an energy consumption of 0.012 kilowatt-hours (kWh) per hour, which may be deemed negligible. Moreover, the power-saving capability exhibited an upward trend as the rotating speed rose.

https://doi.org/10.31436/iiumej.v25i2.3094

6.3. Discussion

The usefulness and stability of the Magneto-Rheological Variable Inertia Flywheel (MRVIF) in regulating rotational speed and limiting power consumption were demonstrated by the results of its experiments. The MR fluid damper in the MRVIF showed accurate corrections to the moment of inertia to changing magnetic fields during a series of dynamic tests. The system's capacity to effectively manage kinetic energy was confirmed by the dynamic control mechanism, which translated into steady and regulated rotating speeds. The trial results also showed a significant decrease in power usage compared to conventional flywheel systems, highlighting the useful advantages of using magnetorheological technology. Overall, the experimental findings confirm the MRVIF's potential as a novel approach to energy-efficient rotational control by offering empirical proof for its effectiveness and stability in practical applications. The close correspondence between the expected and test outcomes is a clear indication of a strong agreement. The factors contributing to minor disparities between them are as follows:

1. During a state of equilibrium, the test results exhibited more variations than their anticipated values. The model did not consider the dynamics of the multi-piston induction motor, the controller's non-linear behaviour, or the forces of MR fluid flow.

2. The model did not consider the MR fluid element's non-linearity. Nevertheless, the disparities between the average values of the steady-state portion of the anticipated and experimental outcomes were minimal.

Their strong relationships validate the assumptions established throughout the model's development. The accuracy of the projected outcomes may have been enhanced by considering the controller detail and the MR fluid. Because of constraints in the test setup design, it was not possible to carry out the experiment using a flywheel with a big diameter. Alternatively, a larger range of operating conditions would have provided improved visibility of the speed fluctuations. The study demonstrates the impact of inertia variation on speed response by simulating models and adjusting the slider position from 0.07 to 0.14m.

Table 3. Comparison of parameters for different case studies

Sl No	Case	Maximum Overshot (%)	Rise Time	Settling Time
1	Simulation	21.9220	0.3200	3.8389
2	Case 1	28.8274	0.3163	4.9763
3	Case 2	28.4227	0.2996	4.8041
4	Case 3	21.3131	0.2719	4.2290

However, the speed response became slower as the viscosity of the MR Fluid increased due to the varying inertia of the flywheel. The system answers were likewise acquired by utilizing the initial power delivered to the controller in all instances. The simulation investigation was also conducted by varying the damping coefficient with different viscosities. Figure 9 displays the temporal responses of the slider position and the variation of the damping coefficient of MR fluid in relation to the slider position for all three situations. The slider rapidly stabilizes as the spring stiffness increases, as depicted in the figure. Nevertheless, the speed of the slider diminished as the viscosity increased, as evidenced by the slope at 1.5 seconds. This is because the slider's change in position was slower when the spring became firmer. The spring force response has a similar pattern to that of the slider position response. As viscosity increases, more power is required to move the sliders. Hence, the speed range of the induction motor expands in proportion to the viscosity of the MR fluid.

The induction motor speeds exhibited lower maximum overshoots (%) when using the MR VIF compared to the VIF shown in Table 2. This signifies that the MR VIF generated reduced levels of vibration, hence enhancing the longevity of the system's components. Thus, it is concluded that by utilizing the MR VIF instead of the VIF, it is possible to decrease the maximum amplitude of vibration and attain a higher steady-state speed. Finally, also for the long running condition, the model consumed less energy if the excitation for the MR VIF was applied. The possible reason for the overall inertia increments for MR fluid addition and the variable inertia mass stayed close to the central position for the higher damping force.

Among the many essential parts of the Magneto-Rheological Variable Inertia Flywheel (MRVIF) is the MR fluid damper. A magnetizable fluid that changes viscosity in reaction to an applied magnetic field powered the MR fluid damper. To dynamically modify the moment of inertia, this damper was carefully included in the flywheel system. The flywheel's rotating speed may be actively controlled by the system by varying the strength of the magnetic field inside the MR fluid damper. The effective management of rotational energy made possible by this dynamic adjustment of inertia promoted the best possible energy storage and release. As a result, the MRVIF is a potential technology for applications as it provides a way to lower power consumption by cleverly adjusting the flywheel's inertial characteristics in real time.

7. CONCLUSION

This article examined and analyzed the dynamic reactions of the induction motor speed driving VIF and MR VIF while maintaining the same maximum inertia values. The system model was constructed utilizing the bond graph simulation methodology, and the resulting equations produced from the model were solved using numerical methods. The experimental validation confirmed the model's transitory behaviour. The MR VIF method demonstrated a significant decrease in peak values of the induction motor speed compared to the VIF method, following a sudden shift in speed. The viscosity of the magnetorheological (MR) fluid in the VIF system directly affected the stability of the speed of the induction motor. These responses are also validated by comparing them with the test data. Minor disparities identified between the projected and actual data necessitate additional improvement of the model and the utilization of sensors with a higher sampling rate and more advanced data-gathering technology. Additionally, the system displayed a significant decrease in power consumption. While the construction of MR VIF was more intricate than VIF, it effectively decreased the maximum amplitude of vibration and thus reduced the system's speed.

Also, energy storage and rotational control systems might be completely transformed by implementing energy-saving and speed stability in Magneto-Rheological Variable Inertia Flywheels (MRVIF) in industrial settings. Using MR fluid dampers to dynamically modify the moment of inertia, the MRVIF provides accurate control over rotational speed, guaranteeing stability in industrial operations. This flexibility improves operational dependability while simultaneously optimizing energy usage. This technique might be applied in the real world to a variety of industrial systems, including production lines, grid stability, and installations for renewable energy. Significant power consumption savings, increased efficiency, and increased stability in the face of fluctuating loads or disturbances are among the implications.

REFERENCES

[1] V. Kartašovas, V. Barzdaitis, and P. Mažeika, “Modeling and simulation of variable inertia rotor,” *J. Vibroengineering*, vol. 14, no. 4, pp. 1745–1750, 2012.

[2] X. Dong, J. Xi, P. Chen, and W. Li, “Magneto-rheological variable inertia flywheel,” *Smart Mater. Struct.*, vol. 27, no. 11, p. 115015, Nov. 2018, doi: 10.1088/1361-665X/aad42b.

[3] C. Jauch, "Controls of a flywheel in a wind turbine rotor," *Wind Eng.*, vol. 40, no. 2, pp. 173–185, 2016, doi: 10.1177/0309524X16641577.

[4] M. Huang, "Optimization of powered wheels for commercial aircraft and design of test scheme," *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 237, no. 7, pp. 1751–1764, Jun. 2023, doi: 10.1177/09544070221093182.

[5] M. Malekan, A. Khosravi, and X. Zhao, "The influence of magnetic field on heat transfer of magnetic nanofluid in a double pipe heat exchanger proposed in a small-scale CAES system," *Appl. Therm. Eng.*, vol. 146, pp. 146–159, 2019, doi: 10.1016/j.applthermaleng.2018.09.117.

[6] L. Islam, M. M. Rashid, and M. A. Faysal, "Investigation of the Energy Saving Capability of a Variable Inertia Magneto-Rheological (MR) Flywheel," vol. 2, no. 1, pp. 25–31, 2022.

[7] S. M. Salam and M. M. Rashid, "A new approach to analysis and simulation of flywheel energy storage system," in *8th International Conference on Mechatronics Engineering (ICOM 2022)*, 2022, pp. 90–94, doi: 10.1049/icp.2022.2271.

[8] W. Lu, Y. Luo, L. L. Kang, and D. Wei, "Characteristics of magnetorheological fluids under new formulation," *J. Test. Eval.*, vol. 47, no. 4, 2019, doi: 10.1520/JTE20170477.

[9] T. Liu, J.; Miura, Y.; Ise, "Comparison of dynamic characteristics between virtual synchronous generator and droop control in inverter-based distributed generators," *IEEE Trans. Power Electron.*, vol. 31, pp. 3600–3611, 2015.

[10] C. Li, M. Liang, and T. Wang, "Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals," *Mech. Syst. Signal Process.*, vol. 64–65, pp. 132–148, Dec. 2015, doi: 10.1016/j.ymssp.2015.04.004.

[11] D. Richiedei, A. Trevisani, and G. Zanardo, "A Constrained Convex Approach to Modal Design Optimization of Vibrating Systems," *J. Mech. Des.*, vol. 133, no. 6, Jun. 2011, doi: 10.1115/1.4004221.

[12] D. J. Inman, *Engineering Vibration*, Internatio. Prentice-Hall International, Inc, 1994.

[13] T. E. Saaed, G. Nikolakopoulos, J. E. Jonasson, and H. Hedlund, "A state-of-the-art review of structural control systems," *JVC/Journal Vib. Control*, vol. 21, no. 5, pp. 919–937, 2015, doi: 10.1177/1077546313478294.

[14] M. Trikande, N. Karve, R. Anand Raj, V. Jagirdar, and R. Vasudevan, "Semi-active vibration control of an 8x8 armored wheeled platform," *J. Vib. Control*, vol. 24, no. 2, pp. 283–302, Jan. 2018, doi: 10.1177/1077546316638199.

[15] G. W. Housner *et al.*, "Structural Control: Past, Present, and Future," *J. Eng. Mech.*, vol. 123, no. 9, pp. 897–971, 1997, doi: 10.1061/(asce)0733-9399(1997)123:9(897).

[16] S.-G. Luca, F. Chira, and V.-O. Rosca, "Passive Active and Semi-Active Control Systems in Civil Engineeering," *Constr. Arhit.*, vol. 3, p. 4, 2005.

[17] N. Eslaminasab, "Development of a Semi-active Intelligent Suspension System for Heavy Vehicles," p. 181, 2008.

[18] Thoma JU., "Simulation by bondgraphs: Introduction to a graphical method," *Berlin Springer Sci. Bus. Media*, 2012.

[19] M. T. Fan Y, Mu A, "Study on the application of energy storage system in offshore wind turbine with hydraulic transmission," *Energ Convers Manag. 2016*, pp. 338–346, 2016.

[20] E. P. Kat C-J, "Validation metric based on relative error," *Math Comp Model Dyn*, vol. 18(5), pp. 487–520, 2012.

https://doi.org/10.31436/iiumej.v25i2.3188

BEE ALGORITHM BASED CONTROL DESIGN FOR TWO-LINKS ROBOT ARM SYSTEMS

RUSSUL A. KADHIM, MINA Q. KADHIM, HUTHAIFA AL-KHAZRAJI*, AMJAD J. HUMAIDI

Control and System Engineering Department, University of Technology-Iraq, Baghdad-Iraq

**Corresponding author: 60141@uotechnology.edu.iq*

(Received: 29 January 2024; Accepted: 24 April 2024; Published online: 15 July 2024)

ABSTRACT: This paper presents a comparative study between two advanced versions of the classical Proportional-Integral-Derivative (PID) controller including the Proportional Integral minus Proportional Derivative (PI-PD) controller and the Nonlinear Proportional Derivative (NPD) to manipulate the angular position of the two-link robot arm system and eliminate the effects of the load disturbances. The dynamic equations of the two-link robot arm system were obtained based on the Lagrange approach. To determine the best value of the adjustable coefficients of each controller, the tuning process was converted to an optimization problem. Then, the Bee Algorithm (BA) optimization technique was employed to find the best value of the adjustable coefficients of each controller. The computer simulation results based on MATLAB show the NPD-BA controller outperformed the PI-PD-BA controller in normal conditions. Furthermore, the NPD-BA demonstrated a substantial enhancement when a load disturbance was applied.

ABSTRAK: Kajian ini membentangkan perbandingan antara dua versi terkini pengawal klasik Keseimbangan-Pengkamiran-Terbitan (PID) termasuk pengawal Keseimbangan Pengkamiran tolak Keseimbangan Terbitan (PI-PD) dan Keseimbangan Terbitan Tak Linear (NPD) bagi memanipulasi kedudukan sudut sistem lengan robot dua pautan dan menghapuskan kesan gangguan beban. Persamaan dinamik sistem lengan robot dua pautan ini diperoleh berdasarkan pendekatan Lagrange. Bagi menentukan nilai terbaik pekali boleh laras setiap pengawal, proses penalaan ditukar kepada masalah pengoptimuman. Kemudian, teknik pengoptimuman Algoritma Lebah (BA) digunakan bagi mencari nilai terbaik bagi pekali boleh laras setiap pengawal. Dapatan simulasi komputer berdasarkan MATLAB menunjukkan pengawal NPD-BA mengatasi prestasi pengawal PI-PD-BA dalam keadaan normal. Tambahan, NPD-BA menunjukkan peningkatan yang ketara apabila gangguan beban digunakan.

KEYWORDS: *Control Design, Two-link Robot Arm System, Proportional Integral minus Proportional Derivative Controller, Nonlinear Proportional Derivative Controller, Bee Algorithm*

1. INTRODUCTION

The two-link robot arm system is a major component in the modern manufacturing industry [1]. Additionally, robotic arms can be used as medical solutions for persons that face difficulty in performing physical activities [2]. It is a nonlinear, high coupled, multi-input multi-output (MIMO) and time varying dynamics system. One of the major challenges in controlling the two-link robot arm system is the uncertainties due to the unknown loads that must be handled by the robot arm (i.e. pick and place tasks) [3]. In the context of control design, the two-link robot arm system can be modeled as a double pendulum system with two degrees of freedom where the equation of motion may be established by Lagrange equation

https://doi.org/10.31436/iiumej.v25i2.3188

[4]. Due to its complex dynamic structure, the two-link robot arm system can be considered as a benchmark system for testing and evaluation of different control approaches [5]. For example, Guechi et al. [6] presented a comparative study between Model Predictive Control (MPC) and the Linear Quadratic (LQ) control based on the feedback linearization of the two-link robot arm system. It was observed that the performance of the MPC control outperforms the performance of the LQ control approach. In the same way, Mohammed and Eltayeb [5] compared the performance of the Proportional-Integral-Derivative (PID) controller and Sliding Mode Control (SMC). The outcomes of this study revealed that the performance of the SMC has a faster and more robust response compared to the performance of the PID controller. However, better control signal was observed by SMC. As an alternative control strategy, the Fuzzy Logic Controller (FLC) was implemented by [2]. Baccouch and Dodds [1] proposed a robust PID controller. Recently, Bendimrad [7] introduced a SMC approach for controlling the two-link robot arm system. Taking advantage of the simple structure of the PID and the robustness of the SMC, Long et al. [8] proposed a variable structure PID control method for the two-link robot arm system. The outcomes show that the proposed approach enhanced the speed of the convergence by more than 80% compared with the classical PID control method, while maintaining the same steady-state accuracy. In terms of intelligence controller, Shen [9] proposed a Fuzzy Neural Network (FNN) controller. The design parameters of the proposed control are optimized by combining the Particle Swarm Optimization (PSO) and backpropagation (BP) algorithm. The findings of the study indicate that the system has good tracking performance, good adaptability, and stability by applying the control scheme.

Unlike previous studies, this paper presents a comparative study between two versions of the classical PID controller named Proportional-Integral minus Proportional-Derivative (PI-PD) controller and Nonlinear Proportional-Derivative (NPD) controller for controlling the two-link robot arm system. These two control approaches can be considered as an improved version of the classical PID controller. As opposed to the trial and error method to find the right value of the adjusted parameters of each controller, various swarm optimization algorithms have been proposed in the literature to achieve an optimal performance of the controllers. Swarm optimization algorithms have provided a substantial improvement in the capabilities of solving multivariate, high dimensional engineering problems, and at the same time, it is easy to implement [10-12]. This paper introduces the Bee Algorithm (BA) to tune the two controllers based on the error performance index.

The rest of the paper is organized as follows: the mathematical model of the two-link robot arm system is presented in Section 2. In Section 3 and Section 4, the proposed controllers are introduced and the bee algorithm is given, respectively. The simulation results and discussions are given in Section 5. Section 6 contains the conclusion.

2. MATHEMATICAL MODEL

This section describes in detailed the mathematical model of the two-link robot arm system. For simplicity, the system can be represented as a double pendulum with two masses M_1 and M_2 connected by two weightless rigid rods of lengths L_1 and L_2 as shown in Fig. 1 [1].

The two-link robot arm system has two degrees of freedom represented by the angle that rotates around the origin (θ_1) and the angle that rotates at the endpoint of the first pendulum (θ_2). The two angles $(\theta_1 \text{ and } \theta_2)$ are outputs of the system and they are manipulated by the two input torques $(\tau_1 \text{ and } \tau_2)$ [1].

https://doi.org/10.31436/iiumej.v25i2.3188

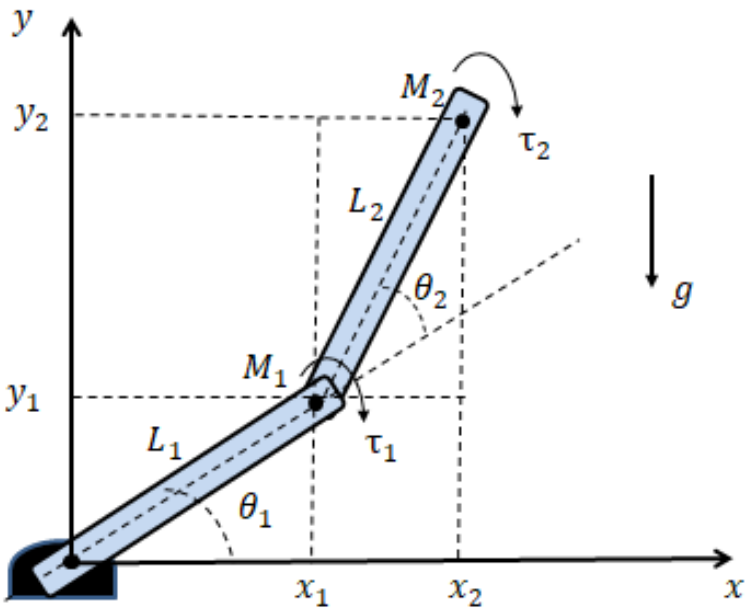


Figure 1. Two-link robot arm system

According to the Lagrangian approach to drive the equations of motion of the system, the Kinetic Energy (KE) and the Potential Energy (PE) of the system have to be obtained.

For the first mass, the equation of the mass in x direction and y direction is given by:

$$x_1 = L_1 cos\ (\theta_1) \quad (1)$$

$$y_1 = L_1 sin\ (\theta_1) \quad (2)$$

For the second mass, the equation of the mass in x direction and y direction is given by:

$$x_2 = L_1\ cos(\theta_1) + L_2 cos\ (\theta_2) \quad (3)$$

$$y_2 = L_1\ sin(\theta_1) + L_2 sin\ (\theta_2) \quad (4)$$

The velocity of the two masses is given by:

$$v_1 = \sqrt{\dot{x}_1{}^2 + \dot{y}_1{}^2} \quad (5)$$

$$v_2 = \sqrt{\dot{x}_2{}^2 + \dot{y}_2{}^2} \quad (6)$$

where

$$\dot{x}_1 = -L_1 \dot{\theta}_1 sin\ (\theta_1) \quad (7)$$

$$\dot{y}_1 = L_1 \dot{\theta}_1 cos\ (\theta_1) \quad (8)$$

$$\dot{x}_2 = -L_1 \dot{\theta}_1\ sin(\theta_1) - L_2 \dot{\theta}_2 sin\ (\theta_2) \quad (9)$$

$$\dot{y}_2 = L_1 \dot{\theta}_1\ cos(\theta_1) + L_2 \dot{\theta}_2 cos\ (\theta_2) \quad (10)$$

The Lagrangian equation is given by:

$$L = KE - PE \quad (11$$

The kinematic energy of the system can be obtained as follows:

$$KE = \frac{1}{2}\ M_1\ \dot{v}_1 + \frac{1}{2}\ M_2 \dot{v}_2 \quad (12)$$

https://doi.org/10.31436/iiumej.v25i2.3188

Substitute $\dot{v}_1$ and $\dot{v}_2$ as given in Eq. (5) and Eq. (6) respectively yields:

$$KE = \frac{1}{2} M_1 (\dot{x}_1^2 + \dot{y}_1^2) + \frac{1}{2} M_2 (\dot{x}_2^2 + \dot{y}_2^2 \quad (13)$$

Substitute $\dot{x}_1, \dot{y}_1, \dot{x}_2$ and $\dot{y}_2$ as given in Eq. (7), Eq. (8), Eq. (9) and Eq. (10), the KE can be rewritten as:

$$KE = \frac{1}{2} M_1 \left(\left(-L_1\dot{\theta}_1 sin(\theta_1)\right)^2 + \left(L_1\dot{\theta}_1 cos(\theta_1)\right)^2 \right) + \frac{1}{2} M_2 \left(\left(-L_1\dot{\theta}_1 sin(\theta_1) - L_2\dot{\theta}_2 sin(\theta_2)\right)^2 + \left(L_1\dot{\theta}_1 cos(\theta_1) + L_2\dot{\theta}_2 cos(\theta_2)\right)^2 \right) \quad (14)$$

Eq. (14) can be rearranged as follows:

$$KE = \frac{1}{2} (M_1 + M_2) {L_1}^2 {\dot{\theta}_1}^2 + \frac{1}{2} M_2 {L_2}^2 {\dot{\theta}_2}^2 + M_2 L_1 L_2 \dot{\theta}_1 \dot{\theta}_2 cos(\theta_1 - \theta_2) \quad (15)$$

The potential energy of the system can be obtained as follows:

$$PE = M_1 g y_1 + M_2 g y_2 \quad (16)$$

Substitute y_1 and y_1 as given in Eq. (2) and Eq. (4) respectively obtains:

$$PE = M_1 g (L_1 sin(\theta_1)) + M_2 g (L_1 sin(\theta_1) + L_2 sin(\theta_2)) \quad (17)$$

Eq. (17) can be simplified as:

$$PE = (M_1 + M_2) g L_1 sin(\theta_1) + M_2 g L_2 sin(\theta_2) \quad (18)$$

Substitute Eq. (15) and Eq. (18) into Eq. (11)

$$L = \left(\frac{1}{2} M_1 \left(\left(-L_1\dot{\theta}_1 sin(\theta_1)\right)^2 + \left(L_1\dot{\theta}_1 cos(\theta_1)\right)^2 \right) + \frac{1}{2} M_2 \left(\left(-L_1\dot{\theta}_1 sin(\theta_1) - L_2\dot{\theta}_2 sin(\theta_2)\right)^2 + \left(L_1\dot{\theta}_1 cos(\theta_1) + L_2\dot{\theta}_2 cos(\theta_2)\right)^2 \right)\right) - ((M_1 + M_2) g L_1 sin(\theta_1) + M_2 g L_2 sin(\theta_2)) \quad (19)$$

The Euler-Lagrange equation is determined as:

$$\frac{d}{dt} \left[\partial L/\partial \dot{\theta}_i\right] - \partial L/\partial \theta_i = \tau_i \qquad i = 1,2 \quad (20)$$

The partial derivatives of Eq. (20) $w.r.t \ to \ i = 1$ obtains:

$$\frac{\partial L}{\partial \dot{\theta}_1} = (M_1 + M_2) L_1^2 \dot{\theta}_1 + M_2 L_1 L_2 \dot{\theta}_2 cos(\theta_1 - \theta_2 \quad (21)$$

$$\frac{\partial L}{\partial \theta_1} = - M_2 L_1 L_2 \dot{\theta}_1 \dot{\theta}_2 sin(\theta_1 - \theta_2) - (M_1 + M_2) g L_1 cos(\theta_1) \quad (22)$$

Then:

$$\frac{d}{dt}\left[\partial L/\partial \dot{\theta}_1\right] = (M_1 + M_2) L_1^2 \ddot{\theta}_1 + M_2 L_1 L_2 \ddot{\theta}_2 cos(\theta_1 - \theta_2) - M_2 L_1 L_2 \dot{\theta}_2 \left(\dot{\theta}_1 - \dot{\theta}_2\right) sin(\theta_1 - \theta_2) \quad (23)$$

Substitute Eq. (22) and Eq. (23) into Eq. (20) $w.r.t \ i = 1$ obtains:

$$\left((M_1 + M_2) L_1^2 \ddot{\theta}_1 + M_2 L_1 L_2 \ddot{\theta}_2 cos(\theta_1 - \theta_2) - M_2 L_1 L_2 \dot{\theta}_2 \left(\dot{\theta}_1 - \dot{\theta}_2\right) sin(\theta_1 - \theta_2)\right) - \left((M_1 + M_2) L_1^2 \dot{\theta}_1 + M_2 L_1 L_2 \dot{\theta}_2 cos(\theta_1 - \theta_2)\right) = \tau_1 \quad (24)$$

In the same way, the partial derivatives of Eq. (20) $w.r.t \ to \ i = 2$ obtains:

$$\frac{\partial L}{\partial \dot{\theta}_2} = M_2 L_2^2 \dot{\theta}_2 + M_2 L_1 L_2 \dot{\theta}_1 cos(\theta_1 - \theta_2) \quad (25)$$

https://doi.org/10.31436/iiumej.v25i2.3188

$$\frac{\partial L}{\partial \theta_2} = M_2 L_1 L_2 \dot{\theta}_1 \dot{\theta}_2 sin(\theta_1 - \theta_2) - M_2 g L_2 cos(\theta_2) \quad (26)$$

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{\theta}_2}\right] = M_2 L_2^2 \ddot{\theta}_2 + M_2 L_1 L_2 \ddot{\theta}_1 cos(\theta_1 - \theta_2) - M_2 L_1 L_2 \dot{\theta}_1 (\dot{\theta}_1 - \dot{\theta}_2) sin(\theta_1 - \theta_2) \quad (27)$$

Substitute Eq. (26) and Eq. (26) into Eq. (20) $w.r.t\ i = 2$ obtains:

$$\left(M_2 L_2^2 \ddot{\theta}_2 + M_2 L_1 L_2 \ddot{\theta}_1 cos(\theta_1 - \theta_2) - M_2 L_1 L_2 \dot{\theta}_1 (\dot{\theta}_1 - \dot{\theta}_2) sin(\theta_1 - \theta_2)\right) - \left(M_2 L_1 L_2 \dot{\theta}_1 \dot{\theta}_2 sin(\theta_1 - \theta_2) - M_2 g L_2 cos(\theta_2)\right) = \tau_2 \quad (28)$$

Solving Eq. (24) and Eq. (28) for $\ddot{\theta}_1$ and $\ddot{\theta}_2$ respectively yields:

$$\ddot{\theta}_1 = g_1(t, \theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2, \tau_1, \tau_2) \quad (29)$$

$$\ddot{\theta}_2 = g_2(t, \theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2, \tau_1, \tau_2) \quad (30)$$

where

$$g_1 = \frac{\frac{M\ \tau_1}{M_2 L_1} - M L_2 \dot{\theta}_2^2 sin(\theta_1 - \theta_2) - g cos(\theta_1) - M\ cos(\theta_1 - \theta_2)\left[\frac{\tau_2}{M_2 L_2} + L_1 \dot{\theta}_1^2 sin(\theta_1 - \theta_2) - g cos(\theta_2)\right]}{L_1(1 - M\ cos^2(\theta_1 - \theta_2))} \quad (31)$$

$$g_2 = \frac{\frac{\tau_2}{M_2 L_2} + L_1 \dot{\theta}_1^2 sin(\theta_1 - \theta_2) - g cos(\theta_1) - cos(\theta_1 - \theta_2)\left[\frac{M\ \tau_1}{M_2 L_1} + M L_2 \dot{\theta}_2^2 sin(\theta_1 - \theta_2) - g cos(\theta_1)\right]}{L_2(1 - M\ cos^2(\theta_1 - \theta_2))} \quad (32)$$

$$M = \frac{M_2}{M_1 + M_2} \quad (33)$$

Let x_1 represents θ_1, x_2 represents θ_2, x_3 represents $\dot{\theta}_1$ and x_4 represents $\dot{\theta}_2$. The dynamics of the two-link robot arm system are given by following differential equations:

$$\dot{x}_1 = x_3 \quad (34)$$

$$\dot{x}_2 = x_3 \quad (35)$$

$$\dot{x}_3 = g_1(t, x_1, x_2, x_3, x_4, \tau_1, \tau_2) \quad (36)$$

$$\dot{x}_4 = g_2(t, x_1, x_2, x_3, x_4, \tau_1, \tau_2) \quad (37)$$

3. CONTROLLER DESIGN

In this paper, two modified versions of the classical PID controller including the Proportional Integral minus Proportional Derivative (PI-PD) controller and Nonlinear Proportional Derivative (NPD) are used to manipulate the two angular positions of the robot arm and eliminate the effects of the load disturbances to fulfill the application of pick and place tasks of the robot arm. The determination of the design variables of both controllers is essential and it required a suitable cost function to enable the system to reach a stable mode [13] [14]. For this purpose, in Section 4, BA is utilized to find the optimum setting of the adjustable parameters of the controllers.

3.1. PI-PD Controller

The classical Proportional-Integral-Derivative (PID) controller is the most common control strategy that is used in control system design due to its robustness and simplicity [15] [16]. Much research and practice represent a considerable effort to propose a different structure of the classical PID controller. In this direction, this paper utilizes a modified

https://doi.org/10.31436/iiumej.v25i2.3188

version of the classical PID controller named a Proportional-Integral minus Proportional-Derivative (PI-PD) controller. The general block diagram of the PI-PD controller structure is illustrated in Fig. 2 [17]. The control law (u) of the PI-PD controller is given by [18]:

$$u = K_{p1}e + K_i \int e - K_{p2}y - K_d \frac{dy}{dt} \tag{38}$$

where e and y are the error and the output of the process respectively, K_{p1} and K_{p2} are the proportional gains, K_i is the integrals gain, and K_d is the derivative gain.

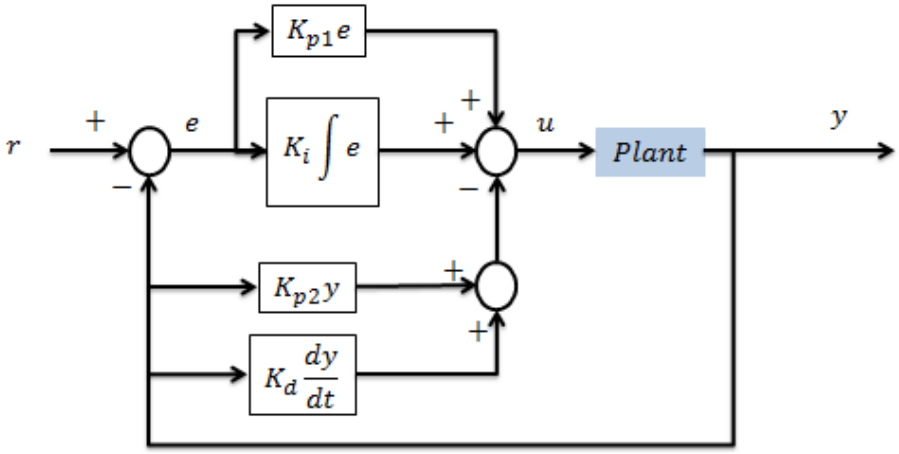


Figure 2. Block diagram of PI-PD controller.

In spite of the simplicity of the PI-PD controller's structure, the tuning process plays a key role in the performance of the PI-PD controller. In order to achieve the optimal performance of the PI-PD controller, BA is employed to find the best value of the design variables of the PI-PD controller.

3.2. Nonlinear PD Controller

To overcome the deficiencies of the classical PID controller for nonlinear systems, a Nonlinear Proportional Derivative (NPD) is introduced. NPD control has been implemented successfully on various process control problems. The control law (u) of the NPD controller is given by [19]:

$$u = K_p f_p(e, \alpha_e, \delta_e) + K_d f_d(\dot{e}, \alpha_{\dot{e}}, \delta_{\dot{e}}) \tag{39}$$

where f_p and f_d are nonlinear function given by [20]:

$$f_p = \begin{cases} |e|^{\alpha_e} sign(e), & for\ |e| > \delta_e \\ \frac{e}{{\delta_e}^{1-\alpha_e}}, & for\ |e| \le \delta_e \end{cases} \tag{40}$$

$$f_d = \begin{cases} |\dot{e}|^{\alpha_{\dot{e}}} sign(\dot{e}), & for\ |\dot{e}| > \delta_{\dot{e}} \\ \frac{\dot{e}}{{\delta_{\dot{e}}}^{1-\alpha_{\dot{e}}}}, & for\ |\dot{e}| \le \delta_{\dot{e}} \end{cases} \tag{41}$$

where $\alpha_e, \delta_e, \alpha_{\dot{e}},$ and $\delta_{\dot{e}}$ are additional parameters that describe the behavior of the nonlinear function of the controller. The parameters α_e and $\alpha_{\dot{e}}$ determine the nonlinearity of the nonlinear function. The parameters δ_e and $\delta_{\dot{e}}$ act as a threshold [19]. The Block diagram of the NPD controller is shown in Fig. 3. The selection of the parameter design of the nonlinear PD controller is the major concerns to ensure the stability and to achieve the best

https://doi.org/10.31436/iiumej.v25i2.3188

performance. To address this issue, BA is introduced to find the optimal value for the adjusted design variables in the NPD controller.

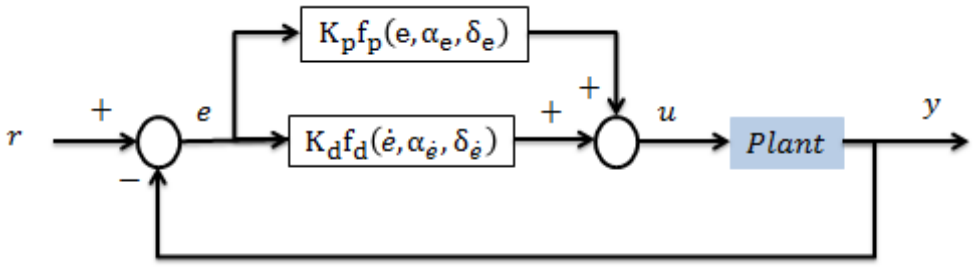


Figure 3. Block diagram of NPD controller.

4. BEE ALGORITHM

Swarm optimization algorithms, in general, are algorithms motivated by the natural behavior of animals, plants, insects… etc. [21]. The Bee Algorithm (BA) is one of the swarm optimization algorithms. The algorithm is introduced by [22]. The algorithm simulates foraging behavior for food of the honey bees. The algorithm starts by initializing the population of bees randomly within the lower and upper boundaries of the search space of the problem as given:

$$p_i = p_{min} + Rand * (p_{max} - p_{min}), \quad i = 1,2,3, \dots, N_{pop} \tag{42}$$

where $i, N_{pop}, and\ p_i$ refer to the index of population, total number of the population, and individual solution respectively, p_{min} and p_{max} are the lower bound and upper bound of the search space, and $Rand$ is a random value between 0 and 1. The BA uses two mechanisms named exploration (i.e. random search) and exploitation (i.e. neighborhood search) to search for the best solution of the optimization problem. In the exploration phase, the algorithm discovers new territory within the search space. On other words, it attempts to find new solutions that haven't been explored previously. In the exploitation phase, the algorithm searches around promising solutions that have showed favorable objective values. In other words, this search converges to the solutions that have already exhibited strong performance. To balance between the exploration search and exploitation search, the BA generates a group of the population with the size of N_m $(i.e. N_m < N_{pop})$ and updates the position of each bee in this group based on the location of the elite bee (p_{ilt}) as given in Eq. (43). The elite bee is the bee that has the best solution that is found by the algorithm.

$$p_i(k+1) = p_i(k) + \sigma(p_i(k) - p_{ilt}) \tag{43}$$

where k is the index of iteration, $p_i(k+1)$ and $p_i(k)$ are the new and current solutions respectively, and σ is the step size. The remaining bees $(N_{pop} - m)$ are assigned to search randomly around the search space as given in Eq. (42). The new population is evaluated and the elite bee is updated. The pseudo code of the BA is given in Algorithm 1.

Algorithm 1 Pseudo code of BA

1. Input
- Objective function, Population size (N_{pop}), Number of iterations (T_{max}), Number of sites (N_s), Step size (σ)

2. Initialization
- Initialize population N_{pop}
- Evaluate objective function
- Assign p_{ilt}

3. Loop:
- **while** ($itr < T_{max}$)
  - **For** each bee in the sites (N_s)
    - ✓ Update the location of bees using Eq. (3.2)
  - **End for**
  - **For** the remaining bees ($N_{pop} - m$)
    - ✓ Update the location Gorilla using Eq. (4.1)
  - **End for**
  - Update p_{ilt}
  - $itr = itr + 1$
- **End while**

4. Print the Optimal Solution

5. SIMULATION RESULTS AND DISCUSSION

To evaluate the performance of the PI-PD controller and NPD controller to control the two-link robot system, the simulation results using MATLAB software are presented in this section. The objective of the controller is to make the two-link in the system follow a desired angular position. The dynamics of the two-link robot arm system, as described by Eq. (34), Eq. (35), Eq. (36) and Eq. (37), are used to conduct the computer simulation. The system's parameters used in the system are listed in Table 1 [1,5].

Table 1. Parameters of two-link robot arm system

Parameters	Values
Mass of first link (M_1)	1 k_g
Mass of second link (M_2)	1 k_g
Length of first link (L_1)	1 m
Length of second link (L_2)	1 m
Acceleration of gravity (g)	9.81 m/s^2

The BA is used for tweaking each controller's design settings in order to guarantee optimal performance. The PI-PD controller is optimized by regulating the adjusted design variables (K_{p1}, K_i, K_{p2} and K_d) of the control action that is introduced in Eq. (38). Similarly, the NPD controller is optimized by regulating the adjusted design variables ($K_p, \alpha_e, \delta_e, K_d, \alpha_{\dot{e}}$ and $\delta_{\dot{e}}$) of the control action that is introduced in Eq. (39). The Integral Time of Absolute Errors (ITAE) index as provided in Eq. (44) [23] is employed as a cost function in the optimization process.

$$ITAE = \int_{tt=0}^{tt=t_{sim}} tt|e(t)|dt \tag{44}$$

where tt is the time and t_{sim} is the total simulation time. Table 2 lists the parameters of the BA. Fig. 4 shows the convergence of BA for tuning the two controllers. Table 3 provides the

https://doi.org/10.31436/iiumej.v25i2.3188

best values of the designed coefficients of the PI-PD and NPD controllers. Fig. 5 illustrates the time response of the angular angles θ_1 and θ_2 when the system is subjected to a unit step input. The evaluation of the response is performed by measuring the settling time (t_s), the steady state error (e_{ss}), the maximum overshoot, and the ITAE index. These specifications of the two controlled systems are reported in Table 4.

Table 2. Algorithm parameters of BA

Parameters	**Value**
Population Size (N_{pop})	25
Number of Iterations (T_{max})	50
Number of sites (N_s)	15
Step size (σ)	2

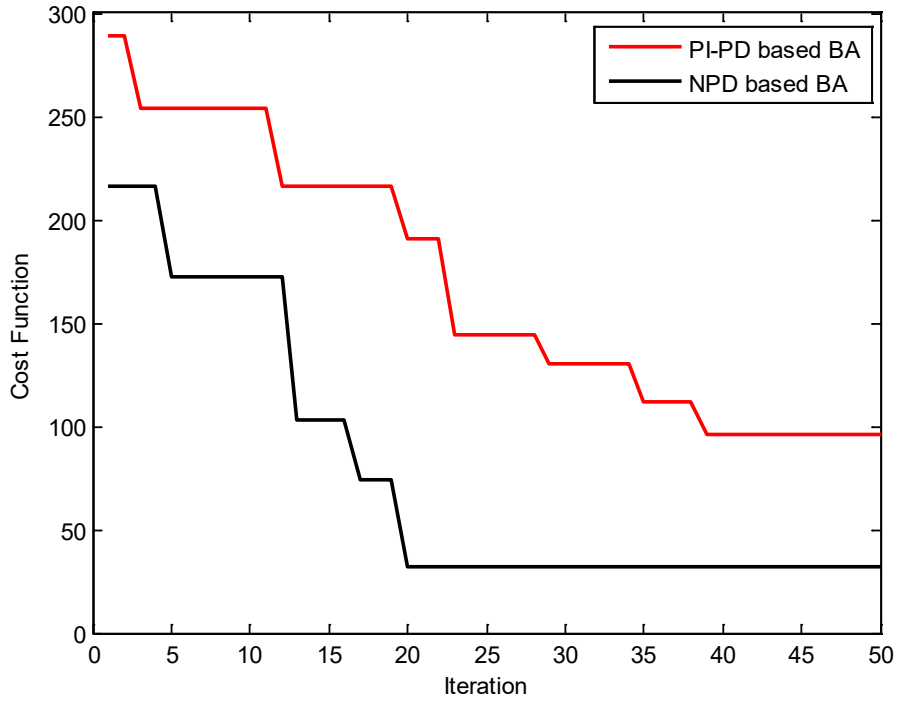


Figure 4. Convergence of BA for proposed controllers

It is evident from Fig. 5 that the two controllers are capable of successfully stabilizing and controlling the system with zero e_{ss}, and zero overshoot response. Regarding t_s and the ITAE index, the dynamics of the NPD controller performs better than the dynamics of the PI-PD controller. For instance, as shown in Table 4, it can be noticed that the value of the t_s is reduced from 2.4 sec and 2 sec for θ_1 and θ_2 response respectively in the case of the PI-PD controller to 1.3 sec and 0.8 sec for θ_1 and θ_2 response respectively in the case of the NPD. This means that the value of t_s is improved by 45.834% and 60% for θ_1 and θ_2 respectively. Furthermore, the value of the ITAE index is reduced from 61.9 and 34.36 for θ_1 and θ_2 response respectively in the case of PI-PD controller to 15.96 and 15.69 for θ_1 and θ_2 response respectively in the case of NPD. This means that the value of the ITAE index is improved by 74.22% and 54.37% for θ_1 and θ_2 respectively. Fig. 6 shows the control signals τ_1 and τ_2.

https://doi.org/10.31436/iiumej.v25i2.3188

Table 3. Optimal value of design parameters based on BA

Controller	Parameter	Value
PI-PD	K_{p1}	30
	K_i	20
	K_{p1}	5
	K_d	10
NPD	K_p	60
	α_e	0.2
	δ_e	0.05
	K_d	10
	$\alpha_{\dot{e}}$	0.6
	$\delta_{\dot{e}}$	0.05

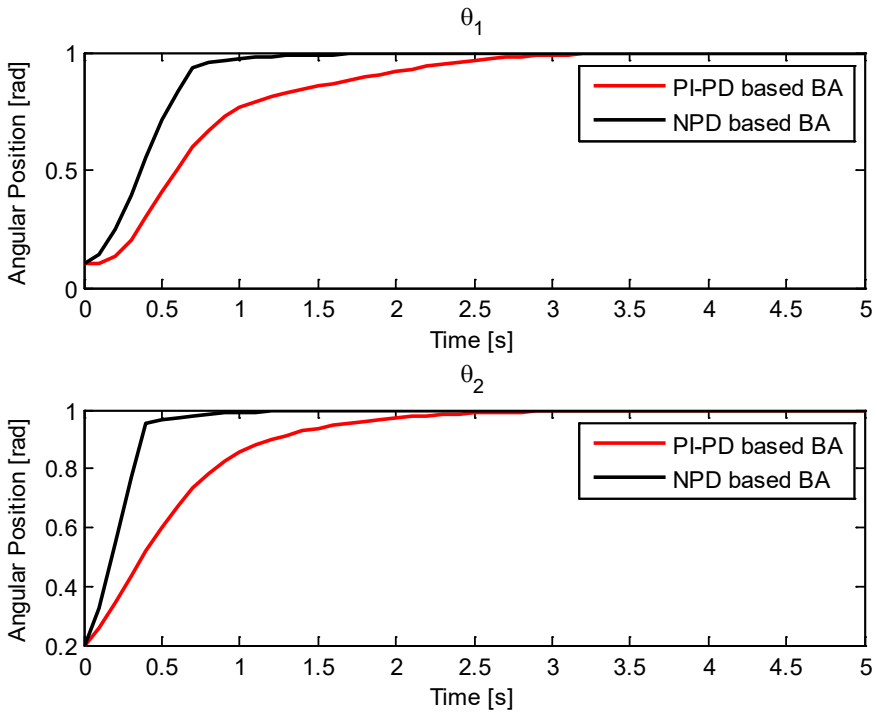


Figure 5. Response of θ_1 and θ_2 for unit step input

Table 4. Specification performances of system without disturbance

Controller	θ	Settling Time (s)	Error Steady State (rad)	Maximum Overshoot (%)	ITAE
PI-PD	θ_1	2.4	0	0	61.9
	θ_2	2	0	0	34.36
NPD	θ_1	1.3	0	0	15.97
	θ_2	0.8	0	0	15.97

https://doi.org/10.31436/iiumej.v25i2.3188

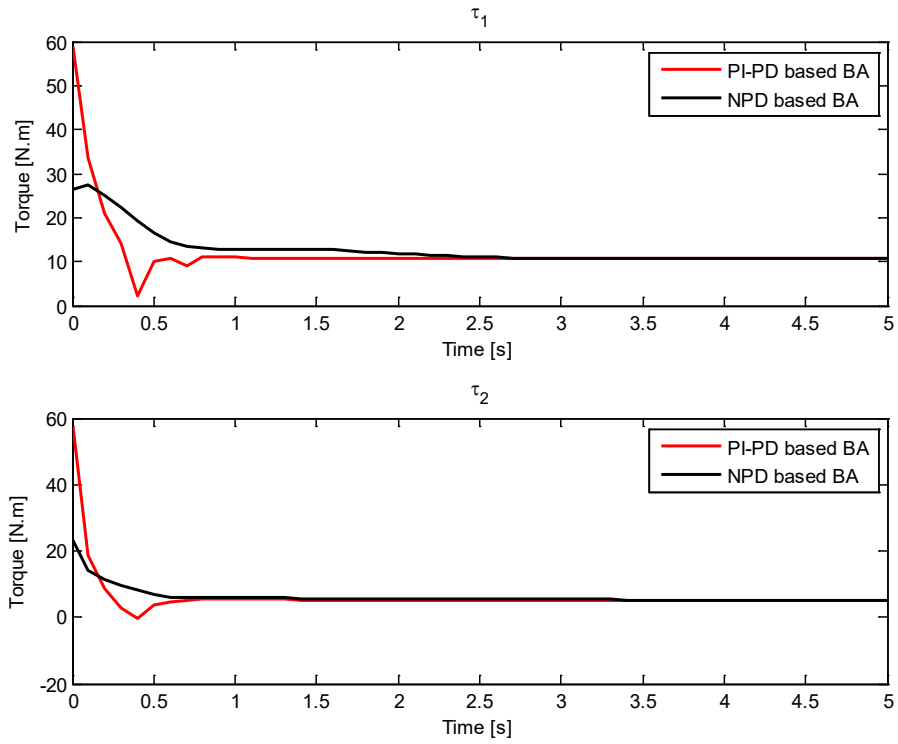


Figure 6. Control signals for unit step input

An external load disturbance has been added to the system based on each controller structure after 4 seconds of simulation time to guarantee the proposed controllers' resilience to load uncertainty. The same designed variables of the controllers that are obtained in Table 3 are used in the simulation. The time response for unit step input is shown in Fig. 7. The system's recovery time and undershoots have been used to evaluate the performance of the system as given in Table 5.

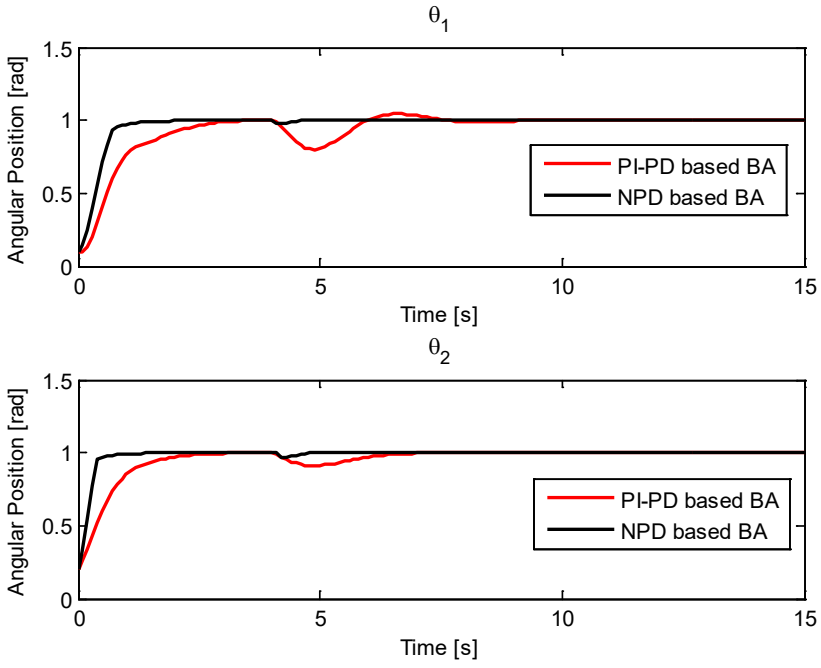


Figure 7. Response of θ_1 and θ_2 for unit step input with disturbance

https://doi.org/10.31436/iiumej.v25i2.3188

It is evident from Fig. 7 that in the load disturbance scenario, the NPD controller performs better for recovery form the disturbance than the PI-PD controller. For example, Table 5 shows that the settling time is reduced from 3.3 sec and 2.7 sec for θ_1 and θ_2 response respectively in the case of the PI-PD controller to 0.5 sec and 0.7 sec for θ_1 and θ_2 response respectively in the case of the NPD controller. This means that the value of t_s is improved by 84.5% and 74.1% for θ_1 and θ_2 respectively. Additionally, the maximum undershoot index is reduced from 20% and 9% for θ_1 and θ_2 response respectively in the case of the PI-PD controller to 0.25% and 0.29% for θ_1 and θ_2 response respectively in the case of the NPD. This means that the maximum undershoot is improved by 98.75 and 96.78 for θ_1 and θ_2 respectively. Furthermore, the value of the ITAE index is reduced from 217.96 and 106.87 for θ_1 and θ_2 response respectively in the case of PI-PD controller to 21.188 and 21.188 for θ_1 and θ_2 response respectively in the case of NPD. This means that the value of ITAE index is improved by 89.96% and 80.17% for θ_1 and θ_2 respectively. Fig. 8 shows the control signals τ_1 and τ_2.

Table 5. Specification performances of the system with disturbance

Controller	$\boldsymbol{\theta}$	Recovery Time (s)	Maximum Undershoot (%)	ITAE
PI-PD	θ_1	3.3	20%	217.96
	θ_2	2.7	9%	106.87
NPD	θ_1	0.5	0.25%	21.188
	θ_2	0.7	0.29%	21.188

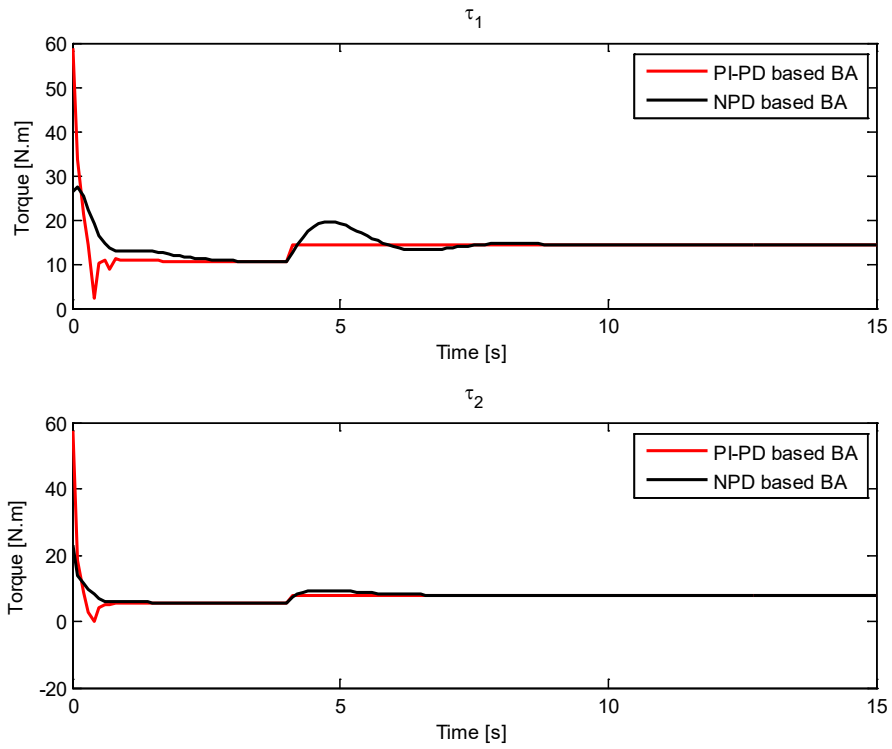


Figure 8. Control signals for unit step input with disturbance

The comprehensive overview of the performance of both controller structures indicates that the NPD outperforms of the PI-PD across the two considered scenarios.

6. CONCLUSION

Controlling the two-link robot arm system was studied based on two control structures named PI-PD and NPD controllers. The dynamics of the system were modeled using Lagrange mechanics. BA was used to adjust the controllers' design parameters in order to guarantee that each controller operated at its optimal performance. According to the simulation results, which were obtained using a MATLAB program, the two controllers optimized by the BA were able to successfully stabilize and regulate the two angular positions of the robot arm system with a 0% error steady state. The outcome also shows that, in terms of reducing the settling time and ITAE index, the NPD-BA controller outperforms the PI-PD-BA controller. Based on the numerical results, the settling time has been improved by 45.834% and 60% for θ_1 and θ_2 respectively whereas the ITAE index has been improved by 74.22% and 54.37% for θ_1 and θ_2 respectively. Furthermore, the NPD-BA controller shows a notable improvement in mitigating the impact of external load disturbances. Based on the numerical results, the settling time has been improved by 84.5% and 74.1% for θ_1 and θ_2 respectively whereas the ITAE index has been improved by 89.96% and 80.17% for θ_1 and θ_2 respectively. This work can be extended further by using another optimization technique to find the adjustment to the controllers' design parameters.

REFERENCES

[1] Baccouch M, Dodds S. (2020) A two-link robot manipulator: Simulation and control design. International Journal of Robotic Engineering, 5(2): 1-17.

[2] Dersarkissian N, Jia R, Feitosa, DL. (2018) Control of a two-link robotic arm using fuzzy logic. In 2018 IEEE International Conference on Information and Automation (ICIA), pp: 481-486.

[3] Velagic J, Hebibovic M, Lacevic B. (2005) On-Line Identification of a Robot Manipulator Using Neural Network with an Adaptive Learning Rate. IFAC Proceedings, 38(1):403-408.

[4] Li-jie C. (2011) Research on the nonlinear dynamical behavior of double pendulum. In 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp. 1637-1640.

[5] Mohammed AA, Eltayeb A. (2018) Dynamics and control of a two-link manipulator using PID and sliding mode control. In 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), pp. 1-5.

[6] Guechi EH, Bouzoualegh S. Zennir Y, Blažič S. (2018) MPC control and LQ optimal control of a two-link robot arm: A comparative study. Machines, 6(3): p.37.

[7] Bendimrad A, El Amrani A, El Amrani B. (2022) Optimization of the Performances of a Two-Joint Robotic Arm using Sliding Mode Control. International Journal of Automotive and Mechanical Engineering, 19(2): 9634-9646.

[8] Long Y, You J, Huang Y, Chen Z, Shi M. (2023) Robust Variable Structure PID Controller for Two-Joint Flexible Manipulator. In International Conference on Computing, Control and Industrial Engineering, Singapore: Springer Nature Singapore, pp. 133-144.

[9] Shen Y. (2024) Robotic trajectory tracking control system based on fuzzy neural network. Measurement: Sensors, 31:101006.

[10] Al-Khazraji H. (2022) Comparative study of whale optimization algorithm and flower pollination algorithm to solve workers assignment problem. International Journal of Production Management and Engineering, 10(1):91-98.

[11] Al-Khazraji H, Nasser AR, Khlil S, (2022) An intelligent demand forecasting model using a hybrid of metaheuristic optimization and deep learning algorithm for predicting concrete block production. IAES International Journal of Artificial Intelligence, 11(2): 649.

[12] AL-Khazraji H, Cole C, Guo W. (2021) Optimization and simulation of dynamic performance of production–inventory systems with multivariable controls. Mathematics, 9(5): 568.

[13] Mustafa N, Hashim FH. (2020) Design of a predictive pid controller using particle swarm optimization. International Journal of Electronics and Telecommunications, 66(4): 737-743.

[14] Mahmood ZN, Al-Khazraji H, Mahdi SM. (2023) PID-Based Enhanced Flower Pollination Algorithm Controller for Drilling Process in a Composite Material. In Annales de Chimie Science des Matériaux, 47(2): 91-96.

[15] Mahdi SM, Raafat SM. (2021) Robust Interactive PID Controller Design for Medical Robot System. International Journal of Intelligent Engineering and Systems, 15 (1): 370-382.

[16] Al-Khazraji H. Cole C, Guo W. (2018) Analysing the impact of different classical controller strategies on the dynamics performance of production-inventory systems using state space approach. Journal of Modelling in Management, 13(1): 211-235.

[17] Ozyetkin MM, Onat C, Tan N. (2020) PI-PD controller design for time delay systems via the weighted geometrical center method. Asian Journal of Control, 22(5): 1811-1826.

[18] Peker F, Kaya I, (2017) Identification and real time control of an inverted pendulum using PI-PD controller. In 2017 21st International Conference on System Theory, Control and Computing (ICSTCC), pp. 771-776.

[19] Ouyang PR, Pano V, Tang J, Yue WH. (2018) Position domain nonlinear PD control for contour tracking of robotic manipulator. Robotics and Computer-Integrated Manufacturing, 51:14-24.

[20] Huang H. (2002) Nonlinear PID controller and its applications in power plants. In Proceedings. International Conference on Power System Technology, 3, pp. 1513-1517.

[21] Mahmood ZN, Al-Khazraji H, Mahdi SM. (2023) Adaptive Control and Enhanced Algorithm for Efficient Drilling in Composite Materials. Journal Européen des Systèmes Automatisés, 56(3): 507-512.

[22] Pham DT, Ghanbarzadeh A, Koc E, Otri S, Rahim S, Zaidi M. (2005) Bee algorithm. Technical Note: MEC 0501. Manufacturing Engineering Centre, Cardiff University, UK

[23] Ahmed AK, Al-Khazraji H. (2023) Optimal control design for propeller pendulum systems using gorilla troops optimization. Journal Européen des Systèmes Automatisés, 56(4): 575-582.

https://doi.org/10.31436/iiumej.v25i2.3207

INTEGRATION OF MFCCS AND CNN FOR MULTI-CLASS STRESS SPEECH CLASSIFICATION ON UNSCRIPTED DATASET

NUR AISHAH ZAINAL, ANI LIZA ASNAWI*, AHMAD ZAMANI JUSOH, SITI NOORJANNAH IBRAHIM, HUDA ADIBAH MOHD. RAMLI

Department of Electrical & Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

**Corresponding author: aniliza@iium.edu.my*

(Received: 3 March 2024; Accepted: 29 May 2024; Published online: 15 July 2024)

ABSTRACT: Stress is an interaction between individuals and their environment, where perceived threats can lead to serious consequences if prolonged and consistently linked to adverse physical and mental health outcomes. Our study explores methods for stress classification via speech, utilizing an unscripted dataset from an experimental study that was able to show the spontaneous reactions of stressed individuals. Mel-Frequency Cepstral Coefficients (MFCCs) emerge as promising speech features, adept at representing the power spectrum crucial to human auditory perception, especially in stress speech recognition. Leveraging deep learning technology, specifically Convolutional Neural Network (CNN), our research optimally combines speech features and CNN algorithms for stress classification. Despite the scarcity of publications on unscripted datasets and multi-class stress classifications, our study advocates their adoption, aiming to enhance performance metrics and contribute to research expansion. The proposed system shows that MFCCs achieve an accuracy of 95.67% in distinguishing among three stress classes (low-stress, medium-stress, and high-stress), surpassing the prior unscripted dataset study by 81.86%. This highlights the efficacy of the proposed MFCCs-CNN system in stress classification.

ABSTRAK: Tekanan merupakan interaksi antara individu dan persekitaran, di mana ancaman akan membawa kepada akibat serius jika berlarutan, dan secara konsisten dikaitkan dengan kesan kesihatan fizikal dan mental yang buruk. Kajian ini mengkaji kaedah pengelasan tekanan melalui pertuturan, menggunakan set data tanpa skrip yang diperoleh daripada kajian eksperimen, iaitu mampu menunjukkan tindak balas spontan individu tertekan. Pekali Septral Frekuensi-Mel (MFCCs) muncul sebagai ciri pertuturan berpotensi, iaitu mahir dalam menunjukkan secara ringkas spektrum kuasa penting bagi persepsi pendengaran manusia, terutama ketika pengecaman pertuturan bertekanan. Memanfaatkan teknologi pembelajaran mendalam, khususnya Rangkaian Neural Lingkaran (CNN), kajian ini menggabungkan ciri pertuturan dan algoritma CNN secara optimum bagi pengelasan tekanan. Walau terdapat kekurangan penerbitan pada set data tanpa skrip dan klasifikasi tekanan pelbagai kelas, kajian ini meningkatkan penggunaannya, bertujuan bagi meningkatkan metrik prestasi dan menyumbang kepada keluasan penyelidikan. Sistem yang dicadangkan ini menunjukkan bahawa MFCC mencapai ketepatan 95.67% dalam membezakan antara tiga kelas tekanan (tekanan rendah, tekanan sederhana dan tekanan tinggi), mengatasi kajian dataset tanpa skrip terdahulu sebanyak 81.86%. Ini menunjukkan keberkesanan sistem MFCCs-CNN dalam pengelasan tekanan.

KEYWORDS: *Multi-class stress classification; Unscripted dataset; Speech stress detection; MFCCs; CNN*

https://doi.org/10.31436/iiumej.v25i2.3207

1. INTRODUCTION

Stress is recognized as an interaction between individuals and their environment, where they perceive something as potentially threatening to their lives [1]. When prolonged, it can escalate to more serious consequences, consistently linked to adverse physical and mental health outcomes [2]. Therefore, the exploration of methods to detect and classify stress is of paramount importance for human well-being. Such endeavors hold the potential to mitigate the risk of severe physical and mental health issues in the future.

In the domain of detection and classification of stress using speech features, Mel-Frequency Cepstral Coefficients (MFCCs) produced promising performance in a few studies [3-8]. As a subset of cepstral features, MFCCs exhibit the capability to succinctly represent the power spectrum, accentuating frequency components crucial to human auditory perception. This attribute is particularly advantageous for stress speech recognition tasks, facilitating the extraction and emphasis of key characteristics across various levels of stress speech productions.

The application of deep learning (DL) technology involves the utilization of a Convolutional Neural Network (CNN) to aid in the classification task. As evidenced by prior research [6, 9-11], CNN has demonstrated superior performance in classifying speech features for binary classification tasks, such as distinguishing between stressed and non-stressed speech. Consequently, our study leverages the advantages of the CNN architecture to determine the optimal combination of speech features and the CNN algorithm for the task of stress classification.

However, the scarcity of publications of unscripted or freestyle datasets, as well as multi-class stress classifications, underscores the necessity of undertaking our study. The motivation behind introducing multi-class stress levels is rooted in its potential to portray the nuanced spectrum of stress perceived by individuals, surpassing the binary classification of stress and no stress. Consequently, the resulting stress classification model is deemed more reliable, given that the provided data is authentic and reflects the actual stress conditions of individuals.

On the other hand, the imperative to generate an unscripted dataset stems from its ability to capture the natural and authentic states of individuals, contrasting with scripted or staged datasets [3]. To the best of our knowledge, only six studies [3,8-12] have utilized unscripted datasets, and one study has employed multi-class stress classification [11]. As a result, our study endeavors to advocate for the adoption of unscripted datasets and multi-class stress classification, enhance existing speech stress performance metrics (e.g., accuracy, F1-score), and ultimately contribute to the expansion of publications in this research domain.

Our study conducted an experimental stress analysis involving 50 tertiary education students who are native speakers of the Malay language in their daily conversations. The rationale behind focusing on native speakers stems from the findings of previous studies [13, 14], which revealed that many students experience apprehension when required to speak in public, particularly in a language other than their native one. This emphasis on native speakers is further justified by concerns raised in the literature, indicating that Malaysian graduates entering the workforce may encounter challenges related to insufficient presentation skills and limited proficiency in the English language. This challenge often arises from a fear of public speaking. Therefore, all these factors contribute to the elicitation of stress among students in the context of the experimental stress analysis.

In addition, the selection of the procedure in the methodology referred to the previous papers that produced the stress speech classification using an unscripted dataset which started

with data acquisition, data pre-processing, feature extraction, and lastly, stress classification [3, 8-12]. The detailed procedure was further discussed in Section 3.

The structure of this paper unfolds as follows: the second section delves into an analysis of related works employing unscripted datasets for stress classification. Following that, the third section outlines the methodology adopted to fulfill the research objectives. Subsequently, the fourth section scrutinizes the results and engages in a discussion of the findings derived from the conducted study. Finally, the fifth section encapsulates the conclusion and outlines potential avenues for future research in this domain.

2. RELATED WORKS

In this section, recent papers utilizing unscripted datasets with speech features, machine learning (ML), and DL technology are presented. A review of prior works is undertaken to enhance the current performance of speech stress detection and classification by examining the methodology of each study. Subsequent subsections provide detailed insights into the integration of the dataset, speech features, and ML/DL architectures utilized in recent studies.

2.1. 40th Mel-filterbank with LSTM-SVM (40M-LSTM-SVM) and Short-term Fourier Transform with Convolutional Recurrent Neural Network (STFT-CRNN)

A study conducted by [10] served as an extension of the research carried out by [12], who initially utilized the same unscripted dataset, known as The Multimodal dataset. This dataset encompassed both audio and visual recordings and involved 56 participants aged between 20 and 30 years, all native speakers of Korean. Participants were subjected to stress-inducing activities, including an English interview, watching an irritating video, and engaging in a repetitive activity (crossing out the same letter). Notably, both studies exclusively utilized audio data, neglecting video recordings.

Despite using the same dataset, the two studies employed different speech features and deep learning (DL) architectures. The rationale behind this deviation lies in the suboptimal accuracy of the earlier work by [12], which stood at 66.40%. Consequently, [10] aimed to enhance the model's accuracy. [10] extracted the 40th Mel-filterbank coefficients during feature extraction, employing the Hann window for audio signal analysis. The normalization procedure included zero mean and unit variance extraction. Their DL architecture incorporated the fusion of Long-Short Term Memory (LSTM) and Support Vector Machine (SVM), known as LSTM-SVM.

Meanwhile, [10] utilized Short-term Fourier Transform (STFT) as the speech feature and adopted a fused architecture of CNN and Recurrent Neural Network (RNN), referred to as CRNN, for classification. Additionally, multi-head attention and a gradient reversal layer were integrated into their DL architecture. Despite achieving a notable performance boost with a 10.90% increase, reaching a 77.30% accuracy, the outcome remains unsatisfactory due to numerous misclassifications of features into incorrect classes. This issue renders the DL model unreliable and necessitates further improvement.

2.2. Spectrogram images with CNN (Spec-CNN)

In a related study within the same field, [9] employed the Distress Analysis Interview Corpus (DAIC) dataset, curated for clinical interviews tailored to diagnose psychological distress issues. The study utilized spectrogram images derived from the raw audio speech obtained during these interviews, and Convolutional Neural Network (CNN) architecture was selected for the classification task. However, it was observed that the spectrogram feature fell

https://doi.org/10.31436/iiumej.v25i2.3207

short of being the optimal choice, as it failed to effectively emphasize the crucial characteristics inherent in distressed speech production.

2.3. Spectral, Cepstral and Prosodic with LSTM (SCP-LSTM)

Stress speech data was collected from recorded call-center interactions, encompassing 363 audio recordings of service interactions between customers and customer service representatives from two pension service providers in the Netherlands [3]. Initially intended for discrete emotion categorization, the dataset covered a range of emotions, including anger/disgust, neutral/sadness, surprise/fear, and happiness. Subsequently, the data was repurposed for stress binary classification, building upon the earlier discrete emotion categorization.

Various spectral, cepstral, and prosodic features were employed in this study, comprising zero-crossing rate, energy, entropy of energy, spectral centroid, spectral spread, spectral entropy, spectral flux, spectral rolloff, mel-cepstral coefficients, chroma vector, and chroma deviations. The authors adopted Long Short-Term Memory (LSTM) with a combination of attention layers to facilitate the classification process. However, it is noteworthy that the study encountered challenges due to imbalanced class distribution, resulting in unsatisfactory outcomes, with an accuracy of only 80.00%.

2.4. MFCCs with CNN (MFCCs-CNN)

Ten participants from the International Islamic University Malaysia (IIUM) engaged in individual virtual interviews conducted through the Discord application [8]. During these interviews, participants were directed to verbally answer the Perceived Stress Scale 10 (PSS-10) questionnaire, comprising ten questions with responses rated on a scale from 0 to 4 (0 for Never, 1 for Almost Never, 2 for Sometimes, 3 for Often, and 4 for Very Often).

The collected data was labeled based on the PSS-10 final score. Notably, the authors opted for binary classification instead of the originally introduced three stress level classifications (low-stress, medium-stress, and high-stress). However, the study reported a suboptimal accuracy of only 61.00%. The utilization of CNN as a classifier was deemed unsuitable due to the limited amount of data points, totaling 150.

Considering the low volume of data points, the authors suggested employing a machine learning (ML) classifier instead of CNN to potentially enhance performance. This adjustment was recommended with the anticipation that an ML classifier could yield improved results under these specific data constraints.

2.5. Pitch, Jitter, Short-term energy variance with 10-fold cross-validation (PJS-10fcv)

Thirty participants volunteered to partake in a stress experiment, spanning ages from 18 to 24 years old [15]. The experimental protocol commenced with participants being tasked to respond to three arithmetic questions interspersed with four passage readings, each requiring two minutes for completion.

The speech features incorporated in this study encompassed all four speech types: prosodic, spectral, cepstral, and Teager Energy Operator (TEO)-based features. Notably, the authors identified pitch, jitter, and short-term energy variance as the most effective speech features in highlighting stress characteristics within the produced speech. Furthermore, the stress classification task was executed utilizing the 10-fold cross-validation (10fcv) method.

https://doi.org/10.31436/iiumej.v25i2.3207

Despite the promising findings, the study limited its focus to binary classification. Regrettably, the classifier encountered an overfitting issue in processing female data, achieving a perfect accuracy of 100.00%, indicative of potential classifier limitation.

2.6. TRIpLet Loss network-vectors with CNN (TRILL-CNN)

The StressDat dataset was employed in a study comprising 30 participants, where various potentially stressful scenarios were presented, spanning low, medium, and high stress levels [11]. Each scenario comprised multiple interconnected sentences, and participants were tasked with reading all scenario sentences across the three stress levels.

The study compared various speech features, including traditional ones such as Low-Level Descriptors (LLD), MFCCs, and Perceptual Linear Prediction (PLP). Additionally, modern non-semantic speech features like TRIpLet Loss network (TRILL) vectors and x-vectors were evaluated. Multiple ML and DL architectures were compared, including Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), CNN, and LSTM. Notably, our study adopted a multi-class stress classification approach, distinguishing between low-stress, medium-stress, and high-stress categories.

The study identified the TRILL-vectors with CNN as the most effective combination, achieving an accuracy of 81.86%. In our study, the goal is to enhance performance by exploring the synergy between MFCCs and CNN within the same multi-class classification framework.

3. METHODOLOGY

In this section, the methodology encompasses the various activities conducted throughout the study, including data acquisition, pre-processing, feature extraction, and, ultimately, stress classification. The proposed system architecture of our study is illustrated in Fig. 1.

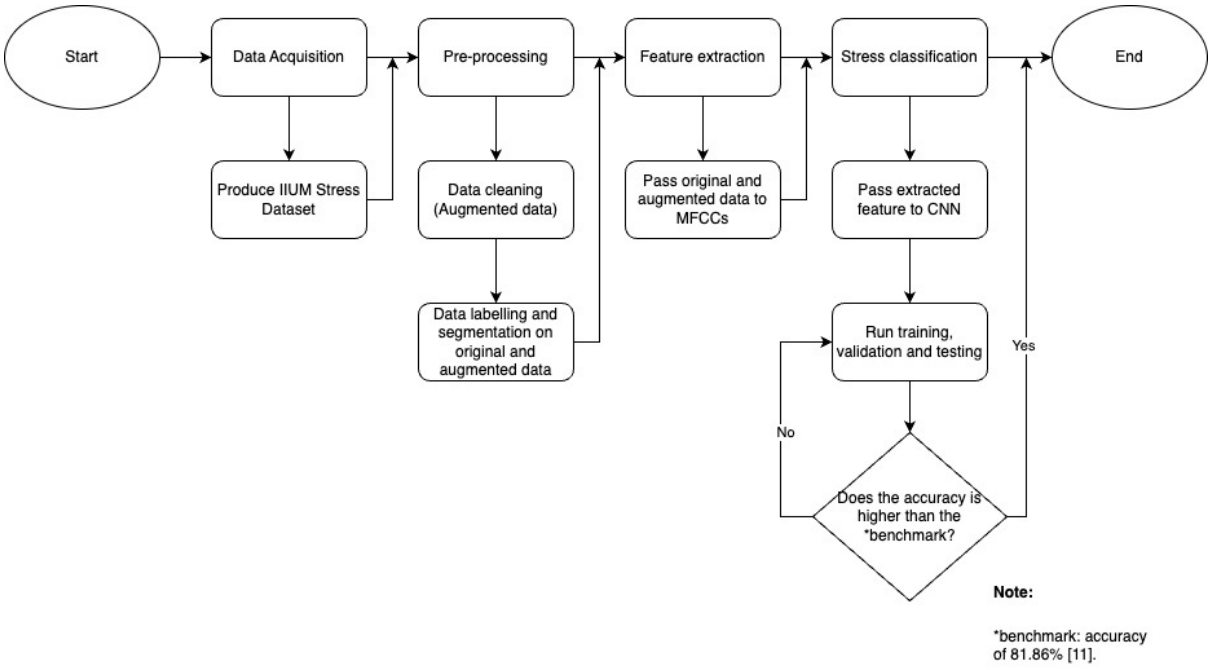


Figure 1. Proposed system architecture of our study

In addition, the study utilized Python programming with several other libraries, such as Pandas, Numpy, OS, and others. Also, the feature extraction and classifier procedures

https://doi.org/10.31436/iiumej.v25i2.3207

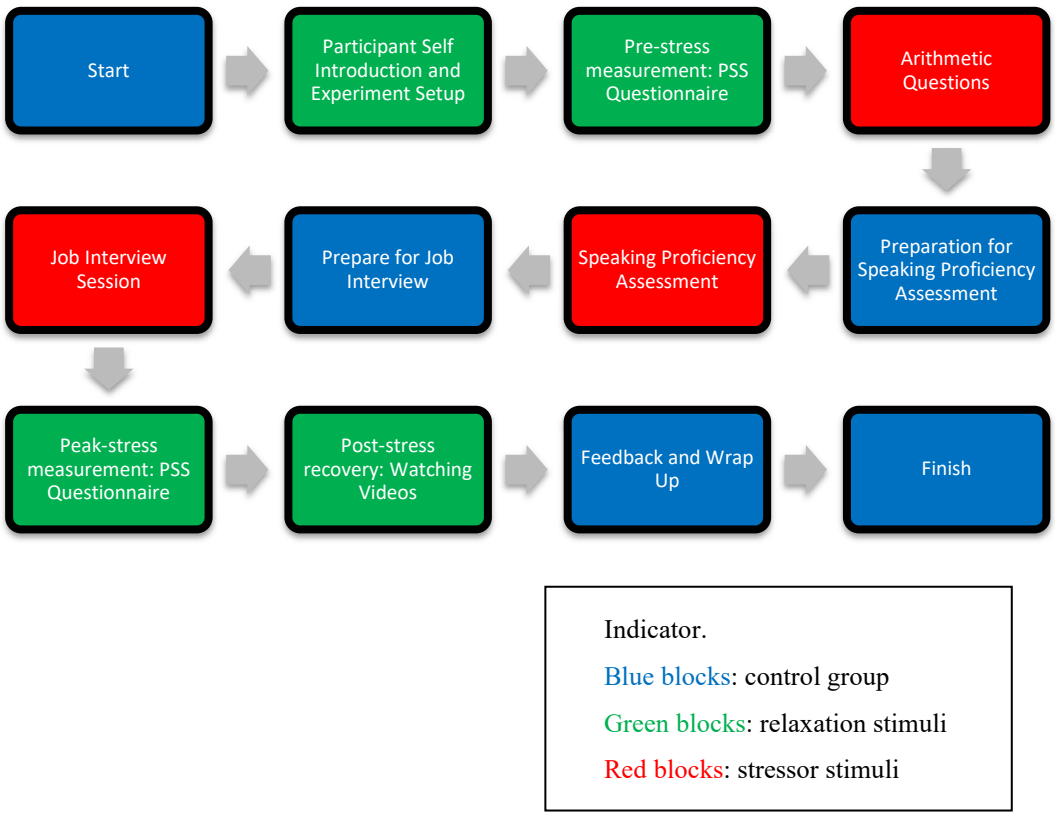
utilized the Pytorch framework. The utilization of this library was motivated by the preference to harness the Graphical Processing Unit (GPU) for accelerated learning.

3.1. Data Acquisition

The experiment aimed to elicit stress responses from participants during stress-inducing activities, and the resulting dataset was named the IIUM Stress Dataset. The activities within the experiment were categorized into three parts: pre-stress activities, stress-induced activities, and post-stress activities, as detailed in Table 1.

Table 1. Flow of experiment of IIUM Stress Dataset

Activity	Tasks during activity
Pre-stress	1. Participant self-introduction and experiment setup
	2. Pre-stress measurement: PSS Questionnaire
Stress-induced	1. Arithmetic Questions
	2. Speaking Proficiency Assessment (SPA)
	3. Job Interview Session
Post-stress	1. Peak-stress measurement: PSS Questionnaire
	2. Post-stress recovery: watching breathing a video
	3. Feedback and wrap-up



© Ani Liza, 2024. All rights reserved.

Figure 2. Experiment flow for the IIUM Stress Dataset (pilot and actual studies)

The experimental sequence commenced with pre-stress activities, followed by stress-induced activities, and concluded with post-stress activities. After an in-depth review of the literature, our study selectively incorporated two stress experimental studies that were

https://doi.org/10.31436/iiumej.v25i2.3207

particularly relevant to our case. Firstly, our study employed the same stimuli pattern created by [14], involving relaxation-stressor-relaxation phases. Secondly, the experimental flow was adapted from the Trier Social Stress Test (TSST) conducted by [16], incorporating time-induced stress tasks proven to elicit stress responses in participants. Additionally, the use of the PSS-5 Questionnaire was adapted from the same study by [16].

Fig. 2 illustrates the complete experiment procedure for the IIUM Stress Dataset. The indicator box showcases blue, green, and red blocks representing the control group, relaxation stimuli, and stressor stimuli, respectively. Notably, a pilot study was conducted before the actual study, providing valuable insights into potential flaws in our methodology, which were subsequently refined in the main study. No supplementary experimental figure is deemed necessary, as no modifications were introduced to the experimental flow for both the pilot and actual studies.

3.2. Pre-processing

Several steps were undertaken during the pre-processing stage, encompassing data cleaning, labeling, and segmentation. Our study employed two types of speech data: original (unaltered) and augmented (edited) data. The augmentation process utilized the Adobe Podcast application, which not only underwent the cleaning process but also underwent several changes, including cleaning up speech recordings, removing background noise, enhancing the overall sound quality by adjusting the pitch, and others.

Both sets of data were labeled based on their respective classes. Class identification for each participant's speech production relied on the PSS-5 questionnaire results, serving as the ground truth for the participants' current stress states. For instance, if the pre-stress PSS-5 result indicated low stress, the speech audio recorded during the pre-stress activities was categorized as low stress. Similarly, if the peak-stress PSS-5 result indicated high stress, the speech audio recorded during stress-induced and post-stress activities was categorized as high-stress.

Given the requirement for the same audio clip length as input in DL, the labeled data was further segmented into five-second snippets for use as input to the classifier. These audio snippets were saved according to their respective stress classes: low-stress, medium-stress, and high-stress. The total number of data points and hours of the snippets are detailed in Table 2.

Table 2. The summary of the data points encompasses both the original and augmented stress speech data

Stress Class	Total Data Points	Total Hours
Low-stress	759	1 hr 6 min 10 sec
Medium-stress	766	1 hr 3 min 50 sec
High-stress	794	1 hr 3 min 15 sec
Total	**2319**	**3hr 13 min 15 sec**

3.3. Feature Extraction

The extraction and representation of features play a pivotal role in influencing the performance of algorithms, particularly in the context of deep learning with speech signals. Feature extraction involves the identification and processing of concealed information within a raw data signal. By discerning and eliminating ineffective features while retaining crucial data, this process streamlines data management.

https://doi.org/10.31436/iiumej.v25i2.3207

In our investigation, we employed Mel-Frequency Cepstral Coefficients (MFCCs) as the speech feature to elucidate the characteristics of stress speeches. This feature was derived from the Torchaudio library developed by PyTorch. The framework of MFCCs is illustrated in Fig. 3.

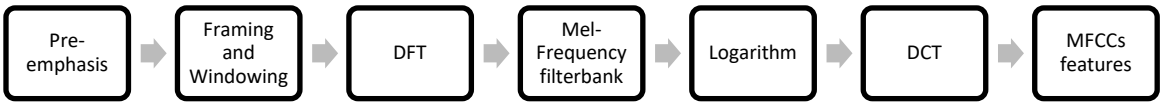


Fig. 3. MFCCs framework [17]

The elaboration of each procedure was as follows [17]:

- **Pre-emphasis**: Pre-emphasis, the first stage of MFCCs is easily achieved by utilizing a high-pass filter with an ordinary filter coefficient (a) of 0.97. The energy distribution across frequencies and the overall energy level were altered by the filtering procedure. Additionally helpful in removing numerical problems during the Fourier Transform operation, this pre-emphasis filter can improve the Signal-to-Noise (SNR) ratio.
  The pre-emphasis filter can be applied to a signal using the first-order filter in Eq. (1) [17].

$$y(t) = x(t) - a(t-1) \quad (1)$$

- **Framing and Windowing**: To make the signal more stationary, the raw signals were divided into frames. Then, speech needs to be assessed briefly enough to have consistent acoustic characteristics. In the speech signal, since there was roughly 20 ms between two glottal closures, 20–30 ms was identified as a Quasi-Stationary Segment (QSS). Conversely, vowel voices were reported to be captured between 40 and 80 ms. Due to this, short-term spectrum measurements were frequently carried out at intervals of 20 ms, with a 10-ms overlap between each frame. Then, tracking the temporal characteristics of the voice signal was possible with frame overlaps of 10 ms.
  A window was employed by each frame to restrict the signal close to the frame's edge. Eq. (2) gave the equation for the Hamming windows, which were chosen for this investigation [17].

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right), 0 \le n \le N. \text{ Window length: } L = N + 1. \quad (2)$$

- **Power Spectrum**: A power spectrum, in general, is the signal's frequency components' power distribution. The power spectrum was computed using the DFT. Each frame's power spectrum was determined using Eq. (3). The signal length was N, and the discrete signal was $x(n)$ [17]:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-\frac{2\pi jnk}{N}}, k = 1, 2, 3 \ldots N-1 \quad (3)$$

- **Mel Filter Bank**: The Mel band-pass filter was a collection of filters based on pitch perception. The Mel filter function was designed to extract a non-linear representation of the speech stream, much like the human ear does. On a Mel scale, the standard Mel filter bank consisted of 40 triangular filters. Each triangular filter in the filter bank had an output of 1 at the centre frequency and was linearly lowered towards 0 until it reached the centre frequencies of the two adjoining filters, where the result was 0. Eq. (4) computes the transfer function (TF) of each m-th filter. Where $\sum_{m}^{M-1} H_m(k) = 1$, and the triangle filter's centre frequency is $f(m)$ [17].

https://doi.org/10.31436/iiumej.v25i2.3207

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \le k < f(m) \\ 1, & k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) < k \le f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (4)$$

Eq. (5) and Eq. (6) were used to compute the conversion of the Mel scale (m) to Hertz (f) and vice versa [17].

$$m = 2595\, log_{10}\left(1 + \frac{f}{700}\right) \quad (5)$$

$$f = 700(10^{\frac{m}{2595}} - 1) \quad (6)$$

- **Discrete Cosine Transform (DCT)**: A limited sequence of data points called a Discrete Cosine Transform (DCT) is used to express a sum of cosine functions that oscillate at different frequencies. To choose the most accelerative coefficients or eliminate the relationship in the log spectral magnitudes from the filter bank, the DCT was applied to the Mel filter bank during the MFCCs process. Eq. (7) was used to compute the DCT. The discrete signal, denoted by x_n in this equation, was of length N [17].

$$X(k) = \sum_{n=0}^{N-1} x_n * \cos\left(\frac{2\pi jnk}{N}\right), k = 1, 2, 3 \ldots N-1 \quad (7)$$

3.4. Stress Classification

Our study utilized Convolutional Neural Network (CNN), a deep learning algorithm, for the multi-class stress classification task. The feature vectors extracted in the previous stage were fed into CNN. The dataset consisted of original and augmented data, and the training set included both. The reason to include the augmented data in the training set was to increase the data points and improve the performance and generalization of the CNN model by having high diversity data on the training set as referred to a finding from a previous study [18]. Meanwhile, the validation and test sets included the original data only. The reason for not using augmented data on the test and validation data was to give an accurate assessment of the model [19].

The dataset was split into three sets: train, validation, and test, with the percentages of each set being 80%, 10%, and 10%, respectively. Both sets of data were combined and used in the train set, while the validation and test sets only comprised the original data. All the data was randomly shuffled before going into CNN. Fig. 4 provides the data point arrangement in each set.

The depicted CNN algorithm, as illustrated in Fig. 5, is composed of six 2-dimensional convolutional layers (Conv2d) interconnected with a Rectified Linear Unit (ReLU) layer and a 2-dimensional Max-pooling (MaxPoold2d) layer placed on each Conv2d layer. Subsequently, after the Conv2d layers, flatten layers were connected to three additional linear layers. A ReLU layer is incorporated at the initiation of the first linear layer, which is then linked to the output layer containing three stress classes. The proposed hyperparameter settings for MFCCs are outlined in Table 3.

https://doi.org/10.31436/iiumej.v25i2.3207

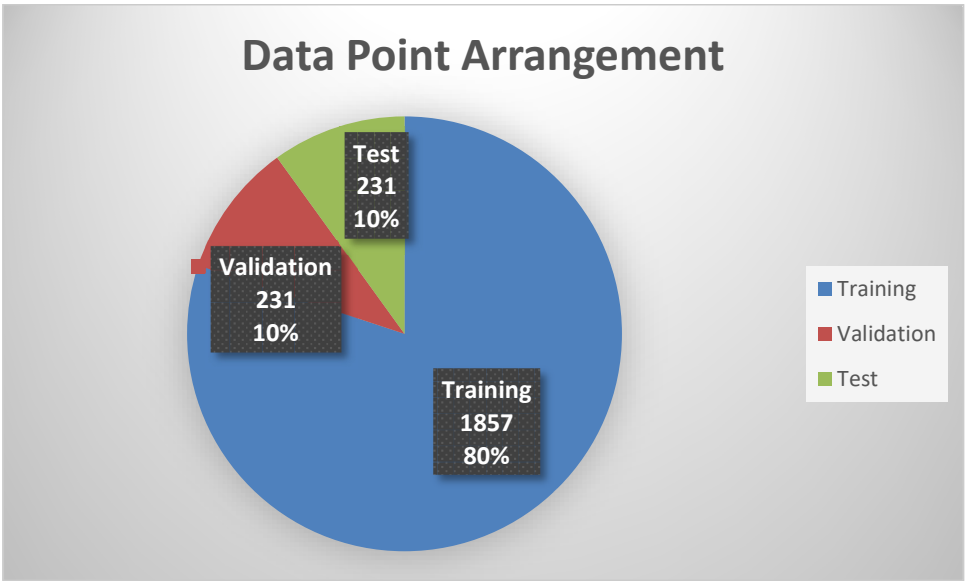


Fig. 4. Arrangement of the data points in each set

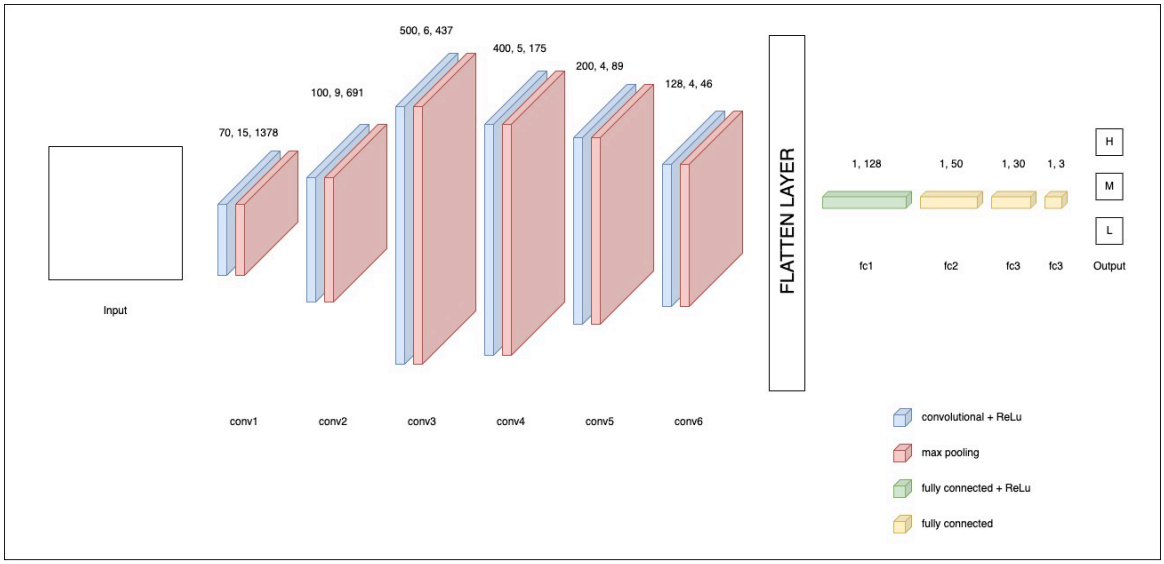


Fig. 5. Proposed CNN algorithm

Table 3: Proposed hyperparameter settings

Epoch	Loss Function	Optimizer	Learning Rate	Momentum
200	Cross-Entropy Loss	Stochastic Gradient Descent	0.0001	0.7

4. RESULT AND DISCUSSION

This section presents a comprehensive examination of the findings, covering the results of the proposed system and a comparative analysis against benchmarks. Firstly, the outcomes from the proposed system include various metrics, such as accuracy, F1-score, individual class accuracies, false positive rate, precision, and recall. Additionally, visual representations of the confusion matrix, accuracy, and loss graphs are provided. Secondly, the comparative analysis against benchmarks is conducted in terms of accuracy only, as the papers available in the literature predominantly reported accuracy metrics in their findings.

https://doi.org/10.31436/iiumej.v25i2.3207

Our study produced better accuracy across all sets: 99.14% for the training set, 92.21% for the validation set, and 95.67% for the test set. As depicted in Fig. 6, the proposed system demonstrated efficient classification of features, exhibiting minimal misclassifications—specifically, fewer than four errors.

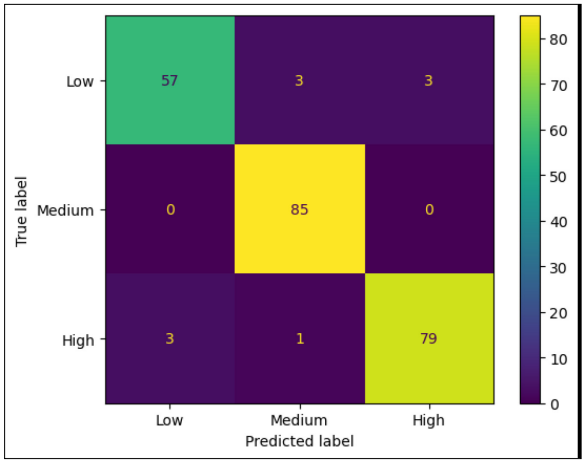


Fig. 6. Confusion matrix of the proposed system

Meanwhile, Table 4 furnishes the performance metrics corresponding to each class for the proposed system, derived from the generated confusion matrix. All scores are normalized to range of 0 to 1. The F1 score for multi-class classification was computed as mean of the F1 scores across individual classes. The proposed system yielded a score approaching 1, indicative of elevated precision and recall, signifying better overall performance.

Table 4: The performance metrics based on each class

Class	Accuracy	False Positive Rate	Precision	Recall	F1-score (for all classes)
Low-stress	0.961	0.018	0.950	0.905	
Medium-stress	0.983	0.027	0.955	1.000	0.960
High-stress	0.970	0.020	0.963	0.952	

Furthermore, as illustrated in Figures 7 and 8, there was a clear positive correlation between the number of epochs and accuracy. The proposed CNN algorithm showed a gradual increase in accuracy over time, unlike the baseline, which showed no improvement. This trend contrasted with the loss curve behavior, where the proposed system experienced a gradual decline in loss as the number of epochs increased, while the baseline system again showed no improvement. This observation supported the claim that the proposed system demonstrated robust generalization capabilities and continuous improvement over time. Notably, the decision to set the final epoch number at 200 was deliberate, as further increases indicated signs of overfitting.

https://doi.org/10.31436/iiumej.v25i2.3207

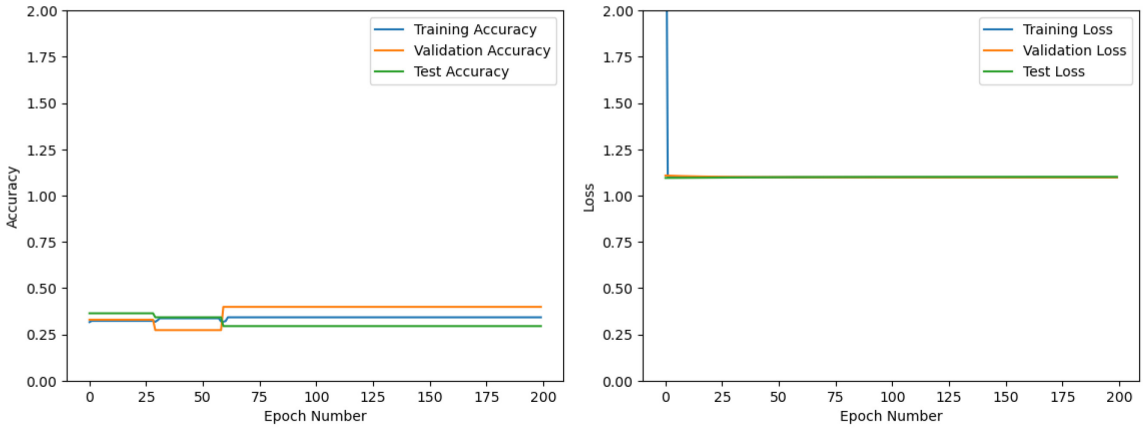


Fig. 7. Accuracy and loss curves for the baseline CNN algorithm, which consists of one layer each of Conv2d, ReLU, MaxPool2d, and Linear

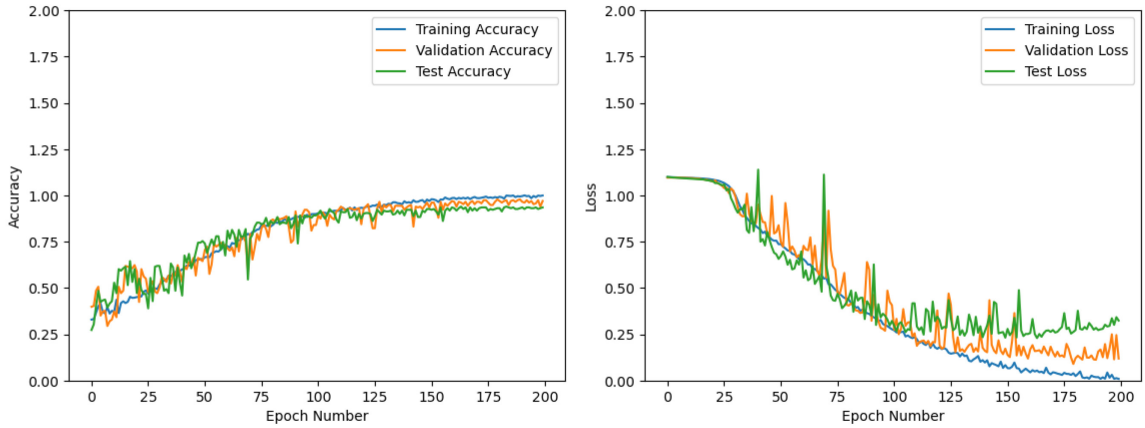


Fig. 8. Accuracy and lost curves for the proposed CNN algorithm

The proposed system in our paper demonstrated the best performance metrics for an unscripted dataset, as evidenced by the results. The study revealed that Mel-Frequency Cepstral Coefficients (MFCCs) exhibit the capability to highlight essential speech characteristics that differentiate between low-stress, medium-stress, and high-stress categories. The implementation of CNN proved instrumental in accurately categorizing the input into the respective stress levels.

Moreover, our work conducted a comparison of performance metrics with several benchmark studies. The papers selected were using one unscripted/freestyle dataset, which was similar to our study. Thus, the comparison was made due to the similar nature of the chosen database. Given that the available metrics from other studies were solely accuracy-related, our study specifically focuses on accuracy, as illustrated in Table 5. Notably, our study achieved an improved accuracy of 95.67%, surpassing previous studies [3, 8-11, 12] with lower accuracy that predominantly engaged in binary classification. While the parameter used in the previous study [8] was similar to our study, variations in the data distribution, including the incorporation of augmented data that increased the number of data points, as well as differences in the selection of CNN parameters' algorithm, collectively contributed to improving the overall performance of our study. In addition, the inclusion of multiple stress classes in our study enhances the accurate categorization of stress speech into respective stress levels. The

proposed system had better performance compared to previous studies, and this could serve as the framework for multi-class speech stress classification.

Table 5. Comparison between our study and previous research utilizing an unscripted dataset

No.	Speech Features-Classifier	Distribution of Data	Accuracy (%)	Class
1	(MFCCs-CNN) [8]	Training set: 112 datapoints Test set: 38 datapoints	61.00	Two
2	(40M-LSTM-SVM) [12]	Training set: 9.52 hours Test set: 2.36 hours	66.40	Two
3	(Spec-CNN) [9]	Training set: 2627 datapoints Test set: 642 datapoints	77.00	Two
4	(STFT-CRNN) [10]	Training set: 9.52 hours Test set: 2.36 hours	77.30	Two
5	(SCP-LSTM) [3]	Training set: 3738 datapoints Test set: 934 datapoints	80.00	Two
6	(TRILL-CNN) [11]	Training set: 9208 datapoints Test set: 1023 datapoints	81.86	Three
7	**Our study (MFCCs-CNN)**	**Training set: 1857 datapoints** **Validation set: 231 datapoints** **Test set: 231 datapoints**	**95.67**	**Three**

As highlighted in Section 2, MFCCs were widely employed in stress speech recognition due to their outstanding performance. The positive outcomes suggested that MFCCs possessed the capability to accentuate stress-related characteristics in the produced speech, leveraging their features that emulated the response of the human auditory system to sound. Subsequently, the extracted features were categorized into three stress classes with the assistance of CNN. Based on the literature [6, 9-11], CNN performed better in classification tasks due to its ability to effectively capture spatial hierarchies and patterns in images. Their ability to acquire features in a hierarchical and translation-invariant way was made possible by these architectural decisions, which were essential for comprehending complicated visual data.

5. CONCLUSION

The research aimed to classify speech stress into three classes—low-stress, medium-stress, and high-stress—using a deep learning approach to improve upon a prior three-class speech classification task. MFCCs were chosen as the speech feature and integrated with CNN for stress classification. The IIUM Stress dataset, derived from an experimental investigation involving Malay students with tertiary education backgrounds, provided unscripted data representing stressed speech characteristics. The dataset parameters ensured the generation of authentic stress-related speech data, subsequently transformed into MFCC features for stress classification using CNN.

The findings demonstrate that the utilization of MFCCs yielded a significantly higher level of accuracy, achieving 95.67% when distinguishing the three stress classes. Importantly, our analysis has been proven to outperform prior studies that used an unscripted dataset in terms of accuracy, highlighting the effectiveness of the proposed system in generating the best stress classification model. In addition, the utilization of an augmented data strategy, which increased the total number of data points, combined with the integration of MFCCs and CNN, proved effective in improving the overall performance.

The significance of conducting our study lies in its impactful contributions. Firstly, our study produced new evidence for the research field, demonstrating the effectiveness of augmented data strategies and the integration of MFCCs with CNN. Secondly, our study produced an unscripted dataset that comprised authentic recordings of stressed speakers' original speeches. Thirdly, our study proved able to categorize the dataset into three distinct stress classes (low-stress, medium-stress, and high-stress), effectively delineating speakers' perceived stress levels. Lastly, the research yielded improved results, achieving a better accuracy rate, surpassing the benchmarks set by the previous studies in the related field.

In future research, a thorough examination and comparison of different speech features can be conducted to discern the most effective feature that results in superior performance metrics for the identification and classification of individual stress levels. This exploration could involve an in-depth analysis of the strengths and weaknesses of various features, shedding light on their respective contributions to accurate stress level categorization. Such a comprehensive investigation would contribute valuable insights to the field, guiding the development of more refined and effective stress classification systems.

ACKNOWLEDGEMENT

This research was made possible by the Fundamental Research Grant Scheme (FRGS), which is administered by the Ministry of Higher Education (MoHE) and was granted funding with the reference code FRGS/1/2021/TK0/UIAM/02/29.

REFERENCES

[1] S. A. Kriakous, K. A. Elliott, C. Lamers, and R. Owen, "The Effectiveness of Mindfulness-Based Stress Reduction on the Psychological Functioning of Healthcare Professionals: a Systematic Review," *Mindfulness (N Y)*, vol. 12, no. 1, pp. 1–28, Jan. 2021, doi: 10.1007/s12671-020-01500-9.

[2] S. Liu, A. Lithopoulos, C.-Q. Zhang, M. A. Garcia-Barrera, and R. E. Rhodes, "Personality and perceived stress during COVID-19 pandemic: Testing the mediating role of perceived threat and efficacy," *Pers Individ Dif*, vol. 168, p. 110351, Jan. 2021, doi: 10.1016/j.paid.2020.110351.

[3] S. Bromuri, A. P. Henkel, D. Iren, and V. Urovi, "Using AI to predict service agent stress from emotion patterns in service interactions," *Journal of Service Management*, vol. 32, no. 4, pp. 581–611, 2020, doi: 10.1108/JOSM-06-2019-0163.

[4] H. Aouani and Y. Ben Ayed, "Speech Emotion Recognition with deep learning," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 251–260. doi: 10.1016/j.procs.2020.08.027.

[5] S. Mihalache, D. Burileanu, and C. Burileanu, "Detecting Psychological Stress from Speech using Deep Neural Networks and Ensemble Classifiers," Institute of Electrical and Electronics Engineers (IEEE), Nov. 2021, pp. 74–79. doi: 10.1109/sped53181.2021.9587430.

[6] P. Chyan, A. Achmad, I. Nurtanio, and I. S. Areni, "A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis," Institute of Electrical and Electronics Engineers (IEEE), Mar. 2023, pp. 1–5. doi: 10.1109/icitisee57756.2022.10057845.

[7] N. A. Zainal, A. L. Asnawi, A. Z. Jusoh, S. N. Ibrahim, H. A. M. Ramli, and N. F. M. Azmin, "MFCCs and TEO-MFCCs for Stress Detection on Women Gender through Deep Learning Analysis," in *2023 9th International Conference on Computer and Communication Engineering (ICCCE)*, IEEE, Aug. 2023, pp. 283–288. doi: 10.1109/ICCCE58854.2023.10246098.

[8] M. S. Hafiy Hilmy *et al.*, "Stress Classification based on Speech Analysis of MFCC Feature via Machine Learning," in *2021 8th International Conference on Computer and Communication Engineering (ICCCE)*, IEEE, Jun. 2021, pp. 339–343. doi: 10.1109/ICCCE50029.2021.9467176.

[9] K. Chlasta, K. Wołk, and I. Krejtz, "Automated speech-based screening of depression using deep convolutional neural networks," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 618–628. doi: 10.1016/j.procs.2019.12.228.

[10] H.-K. Shin, H. Han, K. Byun, and H.-G. Kang, "Speaker-invariant Psychological Stress Detection Using Attention-based Network," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2020, pp. 308–313.

[11] J. Kejriwal, S. Benus, and M. Trnka, "Stress detection using non-semantic speech representation," in *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*, IEEE, Apr. 2022, pp. 1–5. doi: 10.1109/RADIOELEKTRONIKA54537.2022.9764916.

[12] H. Han, K. Byun, and H.-G. Kang, "A Deep Learning-based Stress Detection Algorithm with Speech Signal," in *Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia*, New York, NY, USA: ACM, Oct. 2018, pp. 11–15. doi: 10.1145/3264869.3264875.

[13] M. E. Damayanti and L. Listyani, "AN ANALYSIS OF STUDENTS' SPEAKING ANXIETY IN ACADEMIC SPEAKING CLASS," *ELTR Journal*, vol. 4, no. 2, pp. 152–170, Aug. 2020, doi: 10.37147/eltr.v4i2.70.

[14] X. T. Tee, T. A. T. Joanna, and W. Kamarulzaman, "Self-regulatory Strategies Used by Malaysian University Students in Reducing Public Speaking Anxiety: A Case Study," *Proceedings of the 2nd International Conference on Social Science, Humanities, Education and Society Development (ICONS 2021)*, vol. 629, no. Icons 2021, pp. 146–152, 2022, doi: 10.2991/assehr.k.220101.023.

[15] N. Li, N. Li, M. Guo, and J. Feng, "Research of Speech Biomarkers for Stress Recognition Using Linear and Nonlinear Features," in *2021 7th International Conference on Computer and Communications, ICCC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 509–513. doi: 10.1109/ICCC54389.2021.9674330.

[16] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, "How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life," *Sensors (Switzerland)*, vol. 20, no. 3, Feb. 2020, doi: 10.3390/s20030838.

[17] Z. K. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," *IEEE Access*, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 122136–122158, 2022. doi: 10.1109/ACCESS.2022.3223444.

[18] D. C. Marcu and C. Grava, "The Importance of Data Quality in Training a Deep Convolutional Neural Network," in *2023 17th International Conference on Engineering of Modern Electric Systems (EMES)*, IEEE, Jun. 2023, pp. 1–4. doi: 10.1109/EMES58375.2023.10171785.

[19] Saulo Barreto, "Data Augmentation | Baeldung on Computer Science." Accessed: Jul. 24, 2023. [Online]. Available: https://www.baeldung.com/cs/ml-data-augmentation

https://doi.org/10.31436/iiumej.v25i2.3250

DESIGN AND DEVELOPMENT OF A SLIDER-CRANK ACTUATED KNEE EXOSKELETON WITH OPTIMIZED MOTION CONTROLLER

MARIAM MD GHAZALY[1,2]*, NAI JUN AN[1,2], LAW HIN KWEE[1,2], ZULKEFLEE ABDULLAH[3], NORHASLINDA HASIM[1,2], ISA HALIM[3], NASHARUDDIN ZAINAL[4]

[1]Motion Control Research Laboratory (MCon Lab), Fakulti Teknologi dan Kejuruteraan Elektrik (FTKE), Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Center for Robotics and Industrial Automation (CeRIA), Universiti Teknikal Malaysia Melaka, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[3]Fakulti Teknologi dan Kejuruteraan Industri dan Pembuatan (FTKIP), Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[4]Department of Electrical, Electronic, and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

**Corresponding author: mariam@utem.edu.my*

(Received: 4 April 2024; Accepted: 27 June 2024; Published online: 15 July 2024)

ABSTRACT: The rising incidence of injuries and neurological disorders has highlighted the critical need for accessible and affordable rehabilitation solutions. In response to this demand, robotic exoskeletons have become a popular option for rehabilitation. However, current rehabilitation exoskeletons are generally expensive due to the high force of the actuators used, i.e., electric motors. Therefore, the availability is limited to patients who can afford to pay for physiotherapy using these robotic exoskeletons. Because of the demand for high force, the exoskeleton is heavy, impacting patient safety. In response to these challenges, the main contribution of this study is to develop a lightweight lower-body rehabilitation exoskeleton with sufficient force while maintaining a fast response time and precise motion control for rehabilitation purposes. In this research, a lower body knee joint rehabilitation exoskeleton prototype implementing a slider-crank mechanism was meticulously designed and optimized using Finite Element Analysis (FEA) via SolidWorks software. After optimising the design, the lower body exoskeleton (LBE) was fabricated and assembled. Next, the LBE system was characterized to understand its non-linear behaviour, as the LBE uses a double-acting pneumatic cylinder that is known to exhibit non-linear behaviour. To further analyse the effectiveness of LBE for rehabilitation, a Proportional-Integral-Derivative (PID) controller was adopted for its simplicity in controlling the exoskeleton's angular motions. Excellent results were obtained using a PID controller at the angular displacement of 75^{o}, with a 96.5% reduction in overshoot (OS%), a 92.9% decrease in steady-state error (E_{ss}), a 3.2% reduction of rise time (T_r), and a minimal 0.006% reduction in settling time (T_s). These findings indicate that the LBE with the slider-crank mechanism is a promising device, particularly for knee joint rehabilitation, and that it can be applied to other rehabilitation applications that require a lightweight design and high force application.

ABSTRAK: Peningkatan kecederaan dan gangguan neurologi menyebabkan keperluan kritikal terhadap pemulihan yang senang diakses dan berpatutan. Sebagai solusi kepada keperluan ini, robot eksoskleton telah menjadi pilihan popular bagi sesi pemulihan. Namun, eksoskleton pemulihan sedia ada adalah secara amnya mahal kerana memerlukan daya

https://doi.org/10.31436/iiumej.v25i2.3250

penggerak yang tinggi, contohnya motor elektrik. Maka, ketersediaan menggunakan eksoskleton pemulihan ini terhad kepada pesakit yang mampu membayar fisioterapi mahal menggunakan robot eksoskleton. Selain itu, disebabkan permintaan pada daya penggerak tinggi, robot eksoskleton secara tidak langsung adalah berat dan ini akan memberi kesan kepada keselamatan pesakit. Sebagai solusi kepada permasalahan ini, sumbangan utama kajian ini adalah bagi membangunkan eksoskleton pemulihan bahagian bawah badan yang ringan dan mempunyai daya penggerak yang mencukupi, di samping mengekalkan masa tindak balas yang cepat dan kawalan pergerakan yang tepat bagi tujuan pemulihan. Penyelidikan ini membangunkan prototaip eksoskleton pemulihan sendi lutut bawah badan (LBE) yang menggunakan mekanisme engkol gelangsar dan dioptimumkan dengan teliti menggunakan Analisis Unsur Terhingga (FEA), menggunakan perisian SolidWorks. Selepas reka bentuk dioptimumkan, eksoskleton LBE telah difabrikasi dan dipasang. Seterusnya sistem LBE telah direka bagi memahami ciri-ciri tidak linear, kerana sistem LBE ini menggunakan silinder pneumatik dwitindakan, dimana pneumatik terkenal sebagai sistem tidak linear. Bagi menganalisa lebih lanjut keberkesanan LBE sebagai sistem pemulihan, kawalan Berkadaran-Kamiran-Pembeza (PID) telah digunakan bagi memudahkan kawalan sudut gerakan eksoskleton. Dapatan kajian menunjukkan, kawalan PID adalah sangat baik pada gerakan sudut maksimum, anjakan sudut 75^{o}, di mana pengurangan 96.5% yang ketara dalam lajakan (OS%), penurunan 92.9% dalam ralat keadaan mantap (E_{ss}), 3.2% pengurangan masa naik (T_r), dan pengurangan minimum 0.006% dalam masa penetapan (T_s). Penemuan ini menunjukkan bahawa sistem LBE dengan menggunakan mekanisme engkol gelangsar adalah peralatan yang berkesan, terutama bagi pemulihan sendi lutut, dan ia juga boleh digunakan bagi aplikasi pemulihan lain yang memerlukan reka bentuk ringan dan aplikasi daya yang tinggi.

KEYWORDS: *Lower Body Exoskeleton (LBE), Knee Joint Rehabilitation, Slider-Crank, Finite Element Analysis (FEA), PID controller.*

1. INTRODUCTION

Exoskeletons are versatile external wearable devices with a wide range of applications. They are useful tools for specific tasks, daily activities, and medical rehabilitation, particularly for people who have mobility impairments as a result of spinal cord injuries, stroke, or traumatic events [1-6]. Figure 1 depicts examples of a lower body exoskeleton (LBE), BLEEX (Berkeley Lower Extremity Exoskeleton), designed by the University of California, Berkeley, in 2000. BLEEX was the first autonomous robotic exoskeleton demonstrated to allow the user to carry significant loads over various terrains [7,8]. In contrast, Hybrid Assistive Leg (HAL) is used for walking gait training, whilst Lokomat and ReWalk are used for walking rehabilitation therapy [10-12]. These advanced exoskeletons not only greatly assist in the rehabilitation of stroke patients and the well-being of the elderly, but they can also reduce physical strain in industries such as construction, manufacturing, and logistics [13-17]. In Figure 1, HAL-ML05 refers to the Hybrid Assistive Leg, model ML05, which is used as a medical device for patients suffering from musculoskeletal ambulation disability symptom complex, which includes spinal cord injuries, traumatic brain injuries, cerebrovascular diseases, and other brain and neuromuscular disorders.

https://doi.org/10.31436/iiumej.v25i2.3250

(a) BLEEX [7] HAL-ML05 [9]

Figure 1. Examples of a lower body exoskeleton (LBE)

Nevertheless, the adoption of lower body exoskeletons (LBE) is hampered by limitations such as high costs and complex controllers, making them unfeasible in some regions despite their improved functionality [18,19]. Simplified control systems are required for effective user mobility. Moreover, problems such as the weight of the exoskeletons, low torque, and risk of falls to patients with existing exoskeletons are some of the challenges that need to be considered [20-22]. In response, researchers are working on designing LBE with higher torque and power for lower-body applications using advanced materials and control algorithms [23]. They are also trying to enhance the precision and the rate of the intention-detecting sensors by integrating advanced sensor technologies with machine learning techniques like EMG and motion capture [5,20,21,24,25]. Large-size machine-type supported exoskeletons offer mechanical aid for patients with full disability; however, they are often too bulky, expensive, and too sophisticated for use in hospitals [26-28]. Mobile exoskeletons are lighter and more portable but have drawbacks like shorter battery life and no back support, which restricts the duration of use [11, 29-31]. Moreover, fixed exoskeletons, which are intended for the treadmill, are rather bulky, costly, and possess low mobility [32,33]. Therefore, the aim of this research is to design a lower body rehabilitation exoskeleton that is lightweight and has enough force to support the patient's body weight during rehabilitation while at the same time having a fast response time and accurate control of the motion.

Exoskeletons employ different types of actuators, such as pneumatic, hydraulic, and electrical systems, with their own advantages and disadvantages. For instance, pneumatic actuators can offer linear motion and high-speed benefits with some designs, but they may not be as accurate and may be affected by issues like non-linear airflow and air leakage [34-36]. Hydraulic actuators, for instance, in BLEEX, offer accurate control because the hydraulic fluid is incompressible, but they are known to leak and cause-related injuries [7,8,37]. Electrical actuators employ brushed DC motors for high accuracy; however, they are costly compared to other types [38,39]. In addition, exoskeletons also employ several mechanisms for motion conversion. The slider-crank mechanism is widely used and is a type of kinematic pair that translates rotational motion into linear motion to help in joint

movement, especially in the knee and elbow joints; it offers simplicity and high rigidity within a small range of motion with an added bonus of being lightweight [40-41]. The rack and pinion system that translates rotary motion to linear motion is employed in exoskeletons due to its compactness and high torque density, but it is relatively large and thus may not be portable [42]. Ball screw mechanisms are also used, which are efficient and precise, but they produce high heat and are quite bulky [43].

For proper movement of the exoskeleton, appropriate controller techniques are needed. Traditional approaches like the PID controller use feedback control algorithms to achieve the desired motion while being easy to design [44-47]. Tuning of a PID controller is crucial in order to get the desired performance of a system and this can be done in several ways, each of which has its own benefits. Heuristic methods such as the Trial-and-error method, Ziegler-Nichols, and Cohen-Coon are some of the systematic tuning methods that use system responses to give initial settings for further fine-tuning. The trial-and-error method is a simple method of adjusting the system based on the response of the system, which is easy to implement and gives an immediate response to the system, hence making it effective despite its simplicity. Root locus, bode plot, and Nyquist plot are other methods that provide graphical solutions for stability and performance analysis, which are more complex and need more control theory knowledge. Genetic algorithms, particle swarm optimization, and simulated annealing are some of the computational methods that employ sophisticated techniques to search for the best parameter space and optimize PID gains for complicated systems, but they are time-consuming. Model-based optimization techniques such as Linear Quadratic Regulator (LQR) and Model Predictive Control (MPC) employ state-space models and predictive algorithms to determine the best control actions, offering high accuracy in exchange for accurate system models and high computational demands. Each method is useful, and the choice depends on the application needs and the system's characteristics, which enables the best approach to achieve the best performance of the controller and the system's stability. On the other hand, Artificial Intelligence (AI) based controller strategies, such as machine learning and model-based control, can enhance flexibility and accuracy at the cost of high computational complexity. In addition, a mixed approach integrates machine learning and model-based control, which brings complexity but also flexibility and accuracy [48,49]. Position-based trajectory tracking control involves following joint paths and while it has its drawbacks, it is not easy to model dynamics [50]. Some exoskeletons rely on electromyographic (EMG) signals to control the motion intent and patient recovery status, using methods like stiffness regulation and joint position with EMG information [51-56].

In conclusion, this research aims to develop a lightweight lower body exoskeleton (LBE) movement using a slider-crank mechanism coupled with a double-acting pneumatic actuator to overcome the limitations of current exoskeletons for knee joint rehabilitation in a seated position, using conventional control strategies, that is, a PID controller due to its simplicity, reliability, and flexibility. The slider-crank mechanism and pneumatic actuators are more preferred than other mechanisms and actuation systems, which makes the design lightweight, user-friendly, and safe for patients who are undergoing rehabilitation. The slider-crank mechanism translates rotational motion into linear movement, thus making joint motion simpler, stronger, and more effective, especially in the knee joints. Pneumatic actuators are particularly suitable for linear motion, speed, and force, which makes them suitable for applications that require fast and accurate control of motion.

https://doi.org/10.31436/iiumej.v25i2.3250

2. METHODOLOGY

This section presents the design of the lower body exoskeleton (LBE) prototype using SolidWorks software, the optimization of the LBE design using Finite Element Analysis (FEA), and the discussion on the LBE full experimental setup. To validate the LBE prototype, the LBE was characterized in order to gain insight into the system's non-linear response. After the LBE was characterized, the PID controller was designed using a heuristic approach because of its ability to handle non-linear systems such as the LBE system.

2.1 Lower Body Exoskeleton (LBE) Experiment Setup

This research focuses on the design of a lower-body exoskeleton (LBE) prototype with the aim of knee joint rehabilitation using a slider-crank mechanism in a seated position. To ensure that the exoskeleton device is practical for use, the LBE was designed with the human anatomy specification in mind. It is vital that the device does not restrict the user's movement and allows for safe movement. The biomechanical properties of the body need to be considered, including the location of joints, limb length, limb weight, and the maximum range of motion at the knee joint. This information is necessary to design a device compatible with the patient's body [57,58]. Therefore, the working range for the LBE is set to between 0° to 75°. The knee possesses two rotational degrees of freedom and is classified as a condyloid joint; however, it is frequently reduced to a one DOF joint since the twisting rotation of the joint is restricted and not always necessary for movement. Figure 2 shows the overview of the LBE prototype, which was meticulously designed using SolidWorks. The individual components were designed separately, optimized, and assembled. The LBE consists of two (2) double-acting pneumatic cylinders, pusher holders, leg holders, piston end pins, knee rods, cranks, shank holders, and sliders, with plywood as the frame material. The Finite Element Analysis (FEA) method was employed to comprehensively assess the LBE structural integrity and performance under varying conditions and loads, as illustrated in Figure 3. The FEA analysis was conducted on four crucial LBE components, i.e., Shank 1, Shank 2 (plywood with yield strength 13.8×10^6 N/m^{2}Pa), Pusher Holder (Aluminum 7075 with yield strength 480×10^6 N/m^2), and Slider.

Based on Figure 3, the FEA analysis aims to assess the stress distribution and the safety factor, contributing to a comprehensive understanding of the prototype's performance and durability. The red-colored area in the Finite Element Analysis (FEA) result indicates the areas of high-stress concentration in Shank 2 of the LBE design. The FEA result for Shank 2 suggests that this component experienced significant stress when the load was applied to it, which is crucial information for assessing the structural integrity and performance of the LBE prototype under varying conditions and loads. Table 1 summarizes the FEA results for each component. Based on the analysis, the FEA confirms that the component can withstand applied forces of up to 367 N, which represents the total force exerted by two (2) double-acting pneumatic cylinders, which is still within its yield strength limits, and ensuring resilience against deformation or failure. Furthermore, the factor of safety analysis affirms that all materials used have a factor of safety exceeding one, signifying their capacity to withstand applied forces without damage.

Based on the FEA results, the LBE prototype was fabricated. Figure 4 shows the LBE overall experiment setup comprised of the LBE prototype, which is connected with the OMRON Rotary Encoder (E6B2CWZ6C). The encoder acts as the feedback to the LBE system, which is used for angular motion control. Two (2) double-acting cylinders (Chelic SDA 20/075) convert the LBE linear motion to angular motion using the slider-crank mechanism shown in Figure 2. To control the double-acting cylinders, the Enfield

https://doi.org/10.31436/iiumej.v25i2.3250

Technologies 5/3 Proportional Directional Valve (LS-V05S) was implemented to control the linear bi-directional motion, hence converting the linear motion to rotary motion via the slider-crank mechanism. For the purpose of controlling the pressure input to the LBE system, two (2) pressure sensors, FESTO SDE3-D10S-B-HQ4-2P-M8, were mounted on the LBE system. To evaluate the control performances of the LBE system, MATLAB/ SIMULINK was implemented and was connected in real-time via a DSP board, i.e., Micro-Box 2000/2000C for the real-time control system.

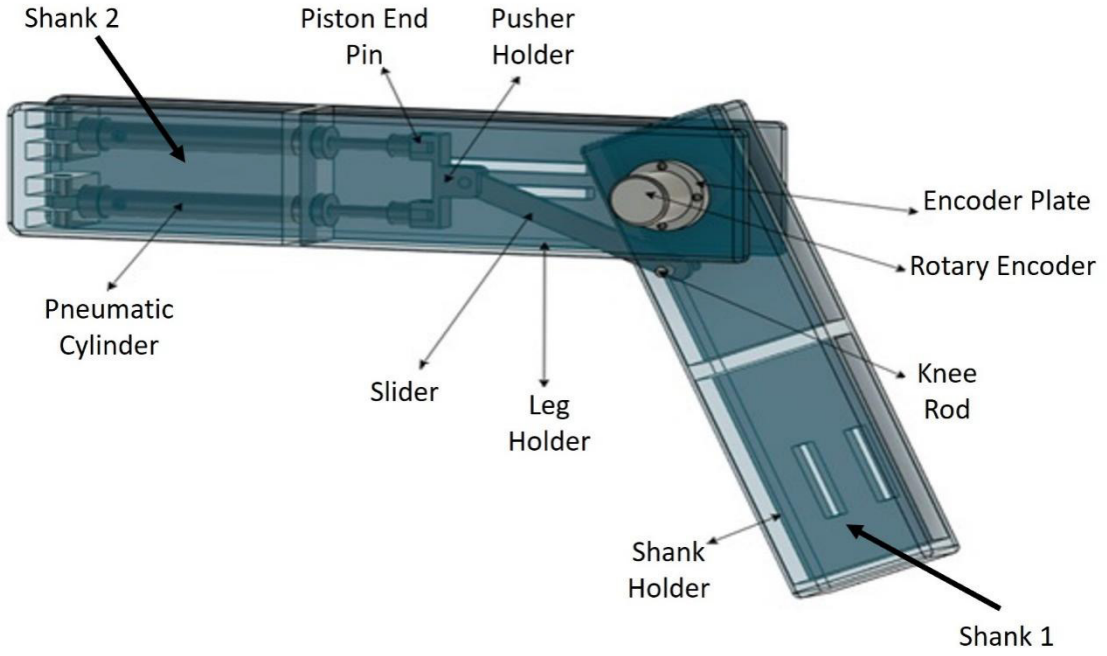
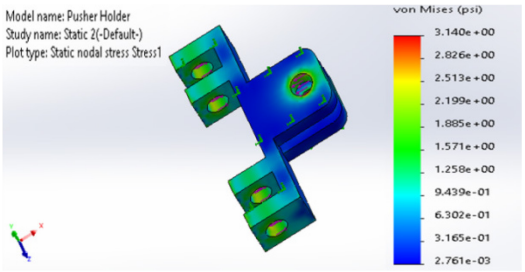
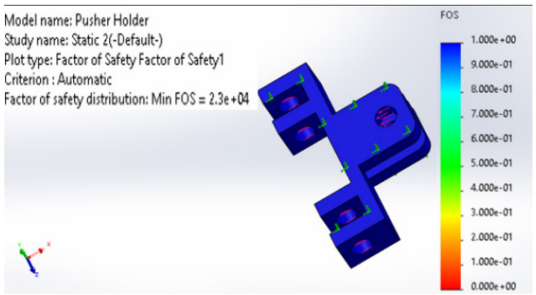


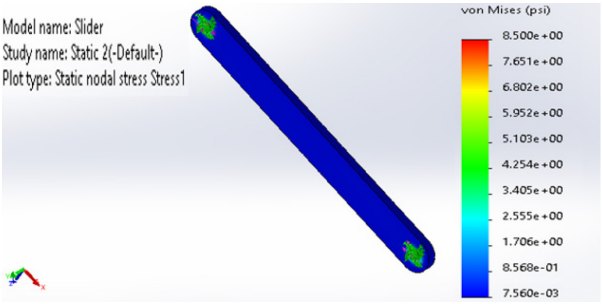
Figure 2. Design overview of the LBE system



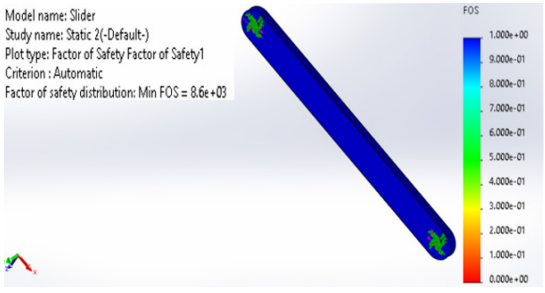
(a) Push Holder Stress Analysis



(b) Push Holder Factor of Safety Analysis

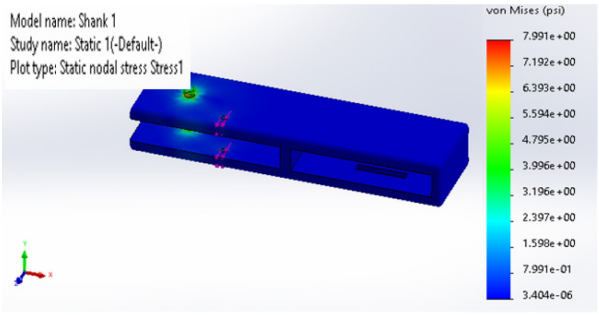


(c) Slider Stress Analysis

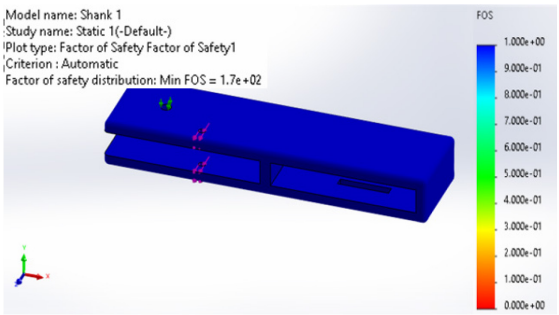


(d) Slider Factor of Safety Analysis

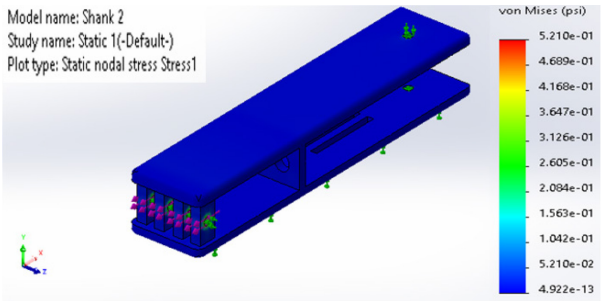
https://doi.org/10.31436/iiumej.v25i2.3250



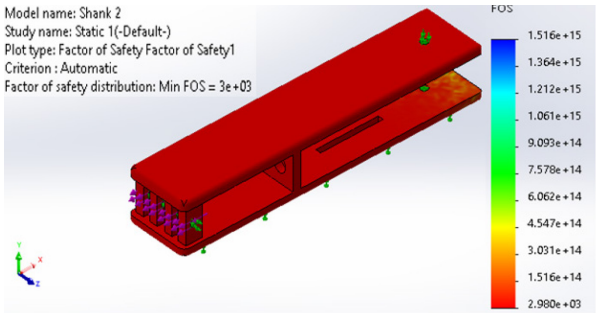
(e) Shank 1 Stress Analysis



(f) Shank 1 Factor of Safety Analysis



(g) Shank 2 Stress Analysis



(h) Shank 2 Factor of Safety Analysis

Figure 3. Finite Element Analysis (FEA) results for the LBE components.

Table 1. FEA test result data

Component part Test	Stress Limit	Strain Limit	Static Displacement Limit	Factor of Safety	Component part Test
Pusher Holder	21649.54 (N/m2)	6.649e-8	6.284e-6 mm	2.3 ×104, >1	Pusher Holder
Slider	58605.4 (N/m2)	10.991e-8	3.549e-7 mm	8.6 ×103, >1	Slider
Shank 1	55096.01 (N/m2)	6.634e-6	3.142e-1 mm	1.7 ×10, >1	Shank 1
Shank 2	3592.17 (N/m2)	4.612e-7	3.965e-6 mm	3 ×103, >1	Shank 2

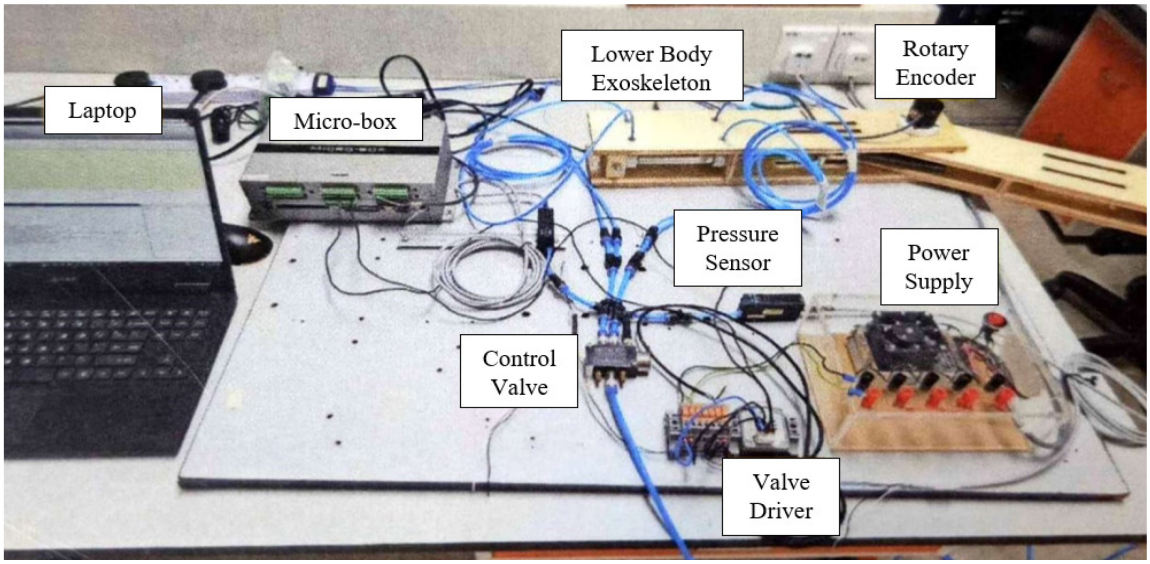


Figure 4. Overall experiment setup for the LBE system

2.2 Characterization Of The Lower Body Exoskeleton (LBE)

The studies in this research were conducted within the operational working range of the Lower Body Exoskeleton (LBE) that is between 0° to 75° angular movement. To characterize the LBE system, the open-loop control system experiment was executed to evaluate the behavior of the LBE system. Figure 5 shows the angular motion of the LBE system to several step input voltages from 0V to 10V, under both without load and with load conditions. The displacement with load refers to the angular motion of the LBE system when a user wears the LBE device, which is equivalent to the estimated applied force of a human leg, 367 N. The LBE design had previously been verified to withstand 367N load using FEA analysis in the previous subsection. The displacement without load evaluation means that the LBE system is evaluated independently. From Figure 5, the retraction motion by the double-acting cylinder was observed below the 5V input signal, while extension motion was observed between 5-10V for both loads and without load. The findings support the hypothesis from the FEA result that the system was able to cause leg movement in the presence of an external load, though to a reduced extent. The relationship between the input voltages and the respective pressures is shown in Figure 6. The results depicted that the LBE system exhibits non-linear behaviors due to the implementation of the pneumatic cylinders. However, it does not significantly affect LBE performance. The results of this experiment demonstrate that the LBE effectively aids in leg movement. Additionally, the Finite Element Analysis (FEA) analysis confirms the functionality of the LBE prototype.

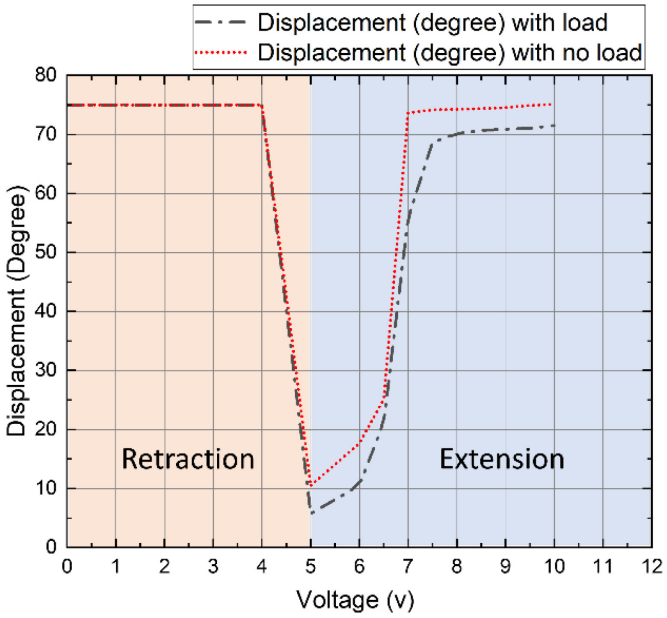


Figure 5. Angular motion of LBE with respect to several input voltages

2.3 Uncompensated Control System For The Lower Body Exoskeleton (LBE)

In order to validate the effectiveness of the Lower Body Exoskeleton (LBE) for knee rehabilitation, initially, a close-loop uncompensated control was designed, whereby a negative feedback loop was introduced, as shown in Figure 7. Figure 8 shows the testing & validation of the LBE prototype implemented, which was conducted at 45° and 75° reference, respectively. Figure 9, Table 2, and Table 3 show the results for the close-loop uncompensated, whereby the experiments were conducted with three (3) times repeatability, the results were averaged, and the standard deviation was calculated to ensure the data was reliable with low standard deviation.

https://doi.org/10.31436/iiumej.v25i2.3250

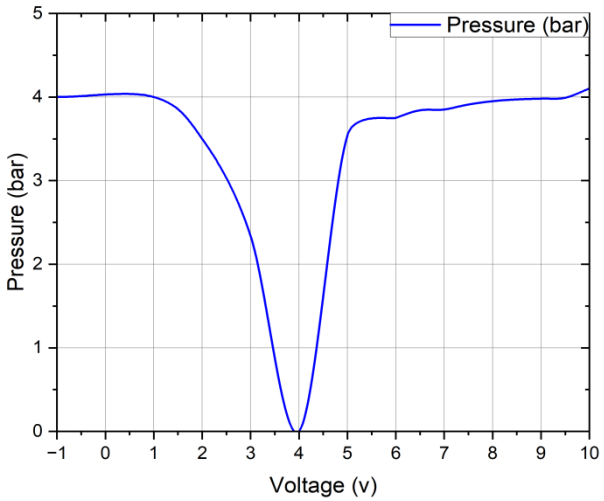


Figure 6. Non-linear characteristic of the double-acting pneumatic cylinder

It can be depicted that the LBE system shows the inability to reach the reference set point at 45° and 75°, with a high error rate, E_{ss} of 5.26° and 4.87°, respectively. It can also be observed in Tables 2 and 3 that the uncompensated control system exhibits high rise time, Tr and settling times, and T_s for the two (2) references, with relatively small overshoot values that do not significantly impact stability.

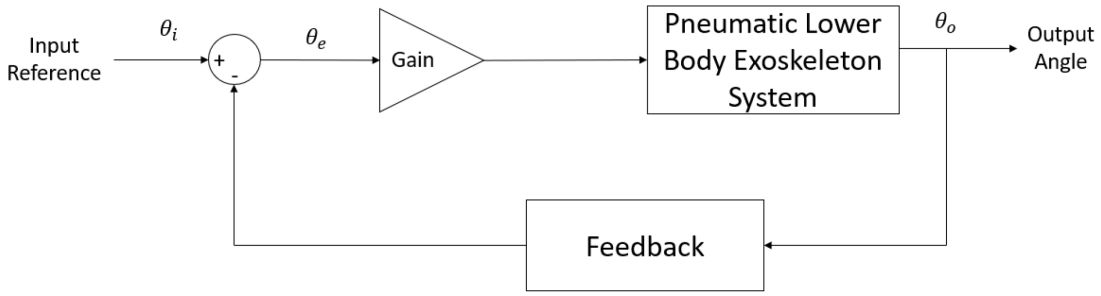
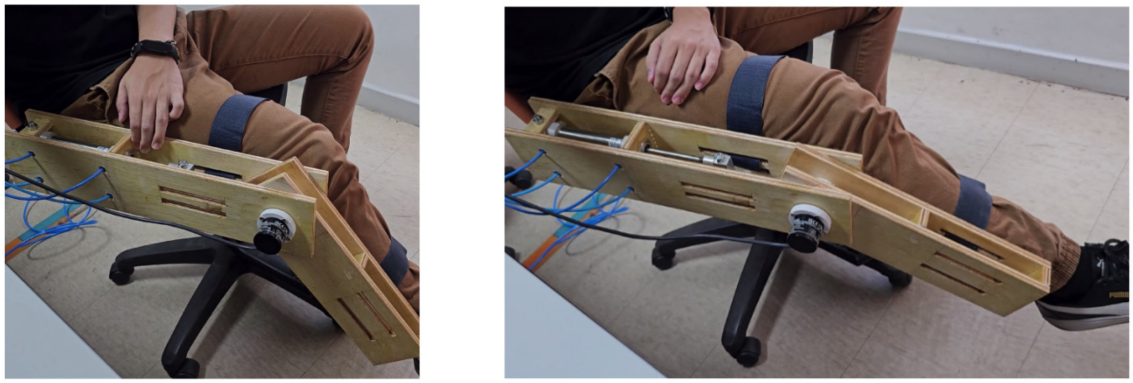


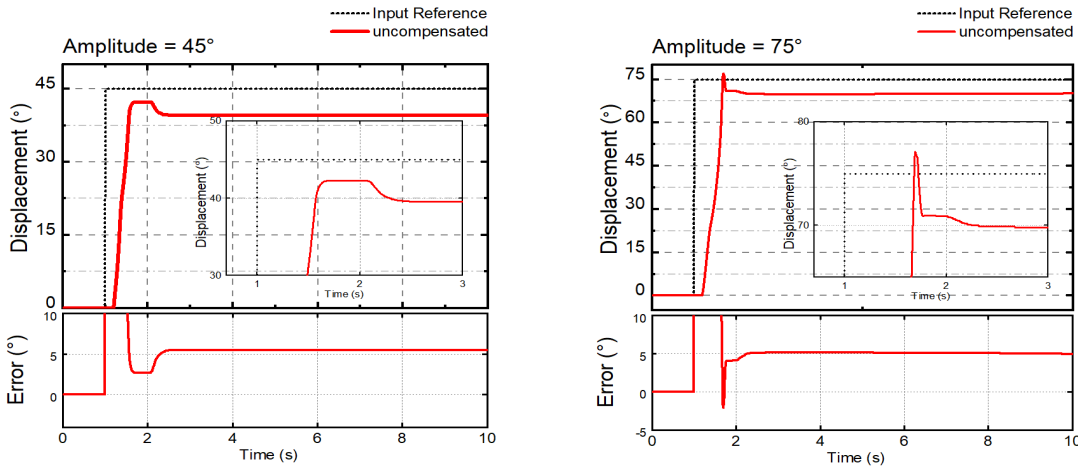
Figure 7. Close-loop uncompensated block diagram for LBE



(a) Reference set point at 45° (b) Reference set point at 75°

Figure 8. Testing and validation of the LBE prototype

https://doi.org/10.31436/iiumej.v25i2.3250



(a) Reference 45° (b) Reference 75°

Figure 9. Comparative transient parameters result for the uncompensated controller.

Table 2. The control performances of the uncompensated closed-loop controller at 45°.

System Response	Experiment			Average	Standard Deviation
	Rep 1	Rep 2	Rep 3		
Rise Time, T_r (ms)	293.0	292.1	295.3	293.5	0.00769
Settling Time, T_s (ms)	2238.1	2321.7	2227.3	2262.4	0.04103
Overshoot, OS (%)	6.9663	7.012	6.9742	6.9848	0.00817
Steady State Error, E_{ss} (°)	5.4677	5.3452	4.9789	5.2639	0.29315

Table 3. The control performances of the uncompensated closed-loop controller at 75°.

System Response	Experiment			Average	Standard Deviation
	Rep 1	Rep 2	Rep 3		
Rise Time, T_r (ms)	395.3	399.8	388.4	394.5	0.00832
Settling Time, T_s (ms)	1752.2	178.33	1728.9	1754.5	0.02693
Overshoot, OS (%)	10.0127	9.991	10.342	10.1159	0.09101
Steady State Error, E_{ss} (°)	5.09	4.52	5.01	4.8733	0.16635

2.4 Proportional Integral Derivative (PID) Controller for the LBE system

To improve the control performance, the Proportional Integral Derivative (PID) controller scheme was proposed and implemented in the LBE system, as shown in Figure 10, to reduce the rise time, settling time, percentage overshoot, and error based on the previous result. In this work, the heuristic method, i.e., trial and error method, was chosen for its practicality due to the highly non-linear behavior of the system, which was verified in the open-loop characterization. Despite being a simpler form of optimization, it allows for effective tuning of the PID controller to achieve the desired performance. The iterative nature of this method means that it inherently seeks to optimize the gain values by continuously refining them based on observed system behavior. The tuning of the PID gain was executed by initially setting the proportional parameter, K_p, and gradually increasing it from zero value until the system exhibited steady-state output. This phase determined the desired rise time, and this K_p value was then selected as the optimized value. In order to reduce the error, while maintaining K_p and K_d, which were set at zero value, the integral parameter, Ki, was then

tuned. Fine-tuning K_i effectively eliminated system errors and enhanced the control accuracy. Finally, the derivative parameter, K_d, was tuned to enhance transient response and reduce high overshoot. The optimized PID gain for the LBE system is shown in Table 4. Tables 5 and 6 show the control performances for the PID controller scheme at 45° and 75° reference, respectively. Table 5 represents the control performances at reference 45°; the standard deviation values for parameters, i.e., rise time (T_r), settling time (T_s), overshoot, and steady-state error (E_{ss}), are low, indicating consistent and reliable results across the repeated experiments. This consistency is crucial in experimental settings as it ensures that the data obtained is reliable and the system's behavior is accurately captured. Similarly, Table 6, which shows the control performances at reference 75°, also exhibits small standard deviation values for the measured parameters. This consistency in the results at different reference angles reinforces the repeatability and reliability of the experiments with the PID gain parameters, demonstrating the controller's robustness in relation to various setpoints.

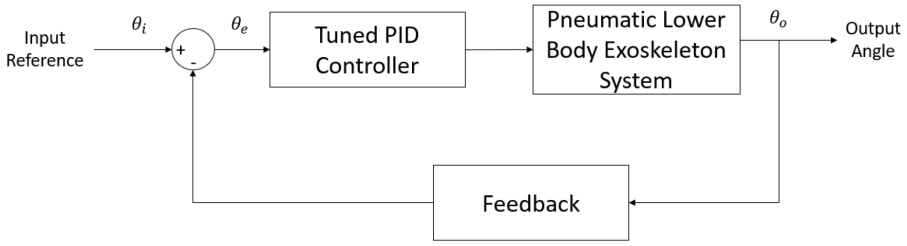


Figure 10. Close-loop block diagram with PID controller

Table 4. PID gain parameters.

PID Gain	Value
K_p	4
K_i	0.35
K_d	0.3

Table 5. The control performances of the Closed-Loop PID Controller at 45°

System Response	Experiment			Average	Standard Deviation
	Rep 1	Rep 2	Rep 3		
Rise Time, T_r (ms)	329.6	339.6	333.6	0.3343	0.00422
Settling Time, T_s (ms)	3485.9	306.79	2968.9	3.1742	0.2302
Overshoot, OS (%)	2.3669	2.56	2.56	2.4976	0.0887
Steady State Error, E_{ss} (°)	-0.04	-0.04	-0.04	-0.04	0

Table 6. The control performances of the Closed-Loop PID Controller at 75°

System Response	Experiment			Average	Standard Deviation
	Rep 1	Rep 2	Rep 3		
Rise Time, T_r (ms)	379.4	384.8	381.8	0.3819	0.00222
Settling Time, T_s (ms)	1747.1	1763.1	1753.1	1.7544	0.00702
Overshoot, OS (%)	0.1195	0.5967	0.3584	0.3582	0.2254
Steady State Error, E_{ss} (°)	0.5549	0.1107	0.3772	0.3476	0.2149

https://doi.org/10.31436/iiumej.v25i2.3250

3. RESULTS AND DISCUSSION

This section will discuss and compare the effectiveness of the closed-loop PID controller to the uncompensated system. Tables 7 and 8 discussed the comparative analysis of transient parameters for both the uncompensated closed-loop system and the PID controller at reference angles of 45° and 75°, respectively. Concurrently, the comparative result is visually represented in Figure 11. At reference 45°, the PID controller showed an improvement in rise time, Tr, by 40.8ms, equivalent to 13.9% compared to the uncompensated system, as well as a significant improvement by 911.8ms in the settling time, Ts, i.e., 40.3% improvement. Additionally, there was a distinct reduction in percentage overshoot, OS% by 4.4872%, signifying a substantial 64.26% decrease and a 5.2639° reduction in steady-state error, E_{ss} achieving a complete 100% reduction at the 45° reference angle. The increased rise time, T_r, and settling time, Ts, in this context can be attributed to the unique challenges posed by small reference angles and the non-linear behavior of the LBE system, as mentioned in Section 2. In such scenarios, the control system must meticulously fine-tune its responses to reach the desired setpoint. At the 75° reference, the PID controller demonstrates notable improvements, with a 3.2% reduction in rise time, $T_{r,}$ and a mere 0.006% change in settling time, T_s. Moreover, there is a substantial 96.5% decrease in overshoot, OS%, and a remarkable 92.9% reduction in steady-state error, E_{ss}. In summary, the proposed PID controller significantly enhanced the performance of the LBE system, as the analysis of the system's response to reference angles of 45° and 75° revealed substantial improvements in various key performance parameters.

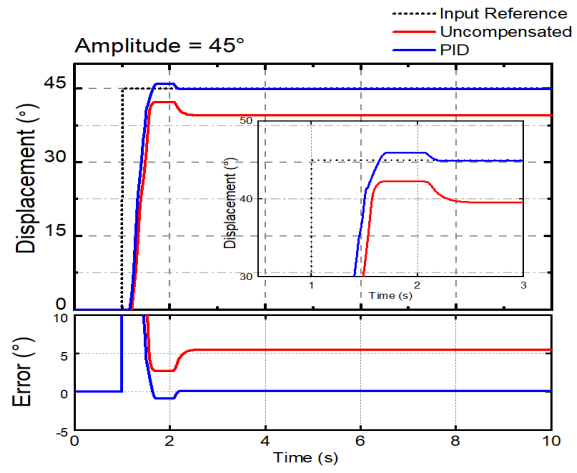
Table 7. Controller performance comparison at reference 45°

Parameters		**Reference 45^0**		
		Uncompensated	**PID**	**Improvement (%)**
Overshoot, OS (%)	**Average**	6.9848	2.4976	64.26%
	Standard Deviation	0.00817	0.00887	
Steady State Error, E_{ss} (^{o})	**Average**	5.2639	0	100%
	Standard Deviation	0.29315	0	
Rise Time, T_r (ms)	**Average**	293.5	334.3	13.9%
	Standard Deviation	7.69	4.22	
Settling Time, T_s (ms)	**Average**	2262.4	3174.2	40.3%
	Standard Deviation	41.03	230.2	

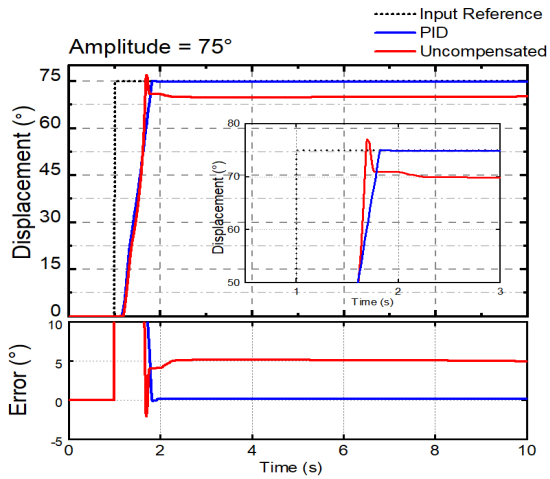
https://doi.org/10.31436/iiumej.v25i2.3250

Table 8. Controller performance comparison at reference 75°

Parameters		Reference 75^0		
		Uncompensated	PID	Improvement (%)
Overshoot, OS (%)	**Average**	10.1159	0.3582	96.5%
	Standard Deviation	0.09101	0.2254	
Steady State Error, E_{ss} (°)	**Average**	4.8733	0.3476	92.9%
	Standard Deviation	0.16635	0.2149	
Rise Time, T_r (ms)	**Average**	394.5	381.9	3.2%
	Standard Deviation	8.32	2.22	
Settling Time, T_s (ms)	**Average**	1754.5	1754.4	0.006%
	Standard Deviation	26.93	7.02	



(a) Reference 45°



(a) Reference 75°

Figure 11. Comparative transient parameters result in uncompensated and PID controller

4. CONCLUSION

In summary, this research successfully achieved its main objective, which was to design a motion controller for a lightweight lower body exoskeleton (LBE) using a slider-crank mechanism dedicated to knee joint rehabilitation. The lower body exoskeleton (LBE) prototype was developed by first implementing the Finite Element Analysis (FEA) method to optimize the design parameters, ensuring the reliability of the prototype. The characteristics of the LBE prototype were then evaluated to confirm the non-linear behavior of the LBE system due to the implementation of the pneumatic double-acting cylinders. The system was then further evaluated using a PID controller to analyze the control performances. The PID closed-loop system exhibited excellent control performances, whereby at reference 45°, the controller showed an improvement in terms of reduction of the overshoot, OS% by 64.26%, and an impressive decrease in steady-state error, E_{ss} by 100% even though there was an increased in rise time, T_r and settling time, T_s. At the 75° reference angle, there was an improvement in rise time, T_r, and settling time; Ts were a substantial 96.5% reduction in

overshoot, OS% accompanied by a significant 92.9% decrease in steady-state error, E_{ss}. These findings depicted the PID controller's efficacy in achieving faster response times, enhancing stability, minimizing overshoot, and greatly improving accuracy and precision in maintaining desired reference angles. Potential future research may aim to enhance the extent of movement and modify the exoskeleton to accommodate bilateral leg utilization. By integrating motion and torque analysis, as well as finite element analysis, the optimization of gait training can be enhanced while also guaranteeing the exoskeleton's safety and durability. These areas of research have the potential to advance the field of lower body exoskeletons for rehabilitation and provide additional benefits to people undergoing rehabilitation.

ACKNOWLEDGEMENT

The authors wish to express their gratitude to Motion Control Research Laboratory (MCon Lab), Center for Robotics and Industrial Automation (CeRIA), Centre for Research and Innovation Management (CRIM), Fakulti Teknologi dan Kejuruteraan Elektrik (FTKE), and Universiti Teknikal Malaysia Melaka (UTeM) for supporting this research and publication.

REFERENCES

[1] Pamungkas DS, Caesarendra W, Soebakti H, Analia R, Susanto S. (2019) Overview: Types of Lower Limb Exoskeletons. Electronics, 8(11): 1283. doi: 10.3390/electronics8111283.

[2] Wu Q, Wang X, Du F, Zhang X. (2015) Design and control of a powered hip exoskeleton for walking assistance. International Journal of Advanced Robotic Systems, 12(3): 18. doi: 10.5772/59757.

[3] Yeem S, Heo J, Kim H, Kwon Y. (2019) Technical analysis of exoskeleton robot. World Journal of Engineering and Technology, 7:68–79. http://dx.doi.org/10.4236/wjet.2019.71004.

[4] Tang X, Wang X, Ji X, Zhou Y, Yang J, Wei Y, Zhang W. (2022) A Wearable Lower Limb Exoskeleton: Reducing the Energy Cost of Human Movement. Micromachines, 13:900. doi: 10.3390/mi13060900.

[5] Rupal BS, Rafique S, Singla A, Singla E, Isaksson M, Virk GS. (2017) Lower-limb exoskeletons: Research trends and regulatory guidelines in medical and non-medical applications. International Journal of Advanced Robotic Systems, 14(6):172988141774355. doi:10.1177/1729881417743554.

[6] Karis MS, Kasdirin HA, Abas N, Saad WHM, Aras MSM. (2023) Emg based control of wrist exoskeleton. IIUM Engineering Journal, 24:391–406. doi: 10.31436/iiumej.v24i2.2804.

[7] Sharma CAA, Sai VKASK, Prasad A, Begum R, Sharvani GS, Manjunath AE. (2018) Multifaceted bio-medical applications of exoskeleton: A review. 2018 2nd International Conference on Inventive Systems and Control (ICISC). doi: 10.1109/icisc.2018.8399053.

[8] Song G, Huang R, Qiu J, Cheng H, Fan S. (2020) Model-based control with interaction predicting for human-coupled lower exoskeleton systems. Journal of Intelligent & Robotic Systems, 100:389–400. doi: 10.1007/s10846-020-01200-5.

[9] Tanaka T, Matsumura R, Miura T. (2022) Influence of varied load assistance with exoskeleton-type robotic device on gait rehabilitation in healthy adult men. International Journal of Environmental Research and Public Health, 19:9713. doi: 10.3390/ijerph19159713.

[10] Van DF, Hesse N, Labruyère R. (2023) Markerless motion tracking to quantify behavioural changes during robot-assisted gait training: A validation study. Frontiers in Robotics and AI, 10. doi: 10.3389/frobt.2023.1155542.

[11] Van NIJ, Van DRB, Van HFH, Rijken H, Geurts AC, Keijsers NL. (2022) Improvement of quality of life after 2-month exoskeleton training in patients with chronic spinal cord injury. The Journal of Spinal Cord Medicine, 47(3): 354–360. doi: 10.1080/10790268.2022.2052502.

[12] Gomez VD, Ballen MF, Rodriguez GC, Munera M, Cifuentes CA. (2021) Experimental characterization of the T-FLEX ankle exoskeleton for gait assistance. Mechatronics, 78: 102608. https://doi.org/10.1016/j.mechatronics.2021.102608.

[13] Masengo G, Zhang X, Dong R, Alhassan AB, Hamza K, Mudaheranwa E. (2023) Lower limb exoskeleton robot and its cooperative control: A review, trends, and challenges for future research. Frontiers in Neurorobotics, 16. doi: 10.3389/fnbot.2022.913748.

[14] Shi D, Zhang W, Zhang W, Ding X. (2019) A Review on lower limb rehabilitation exoskeleton robots. Chinese Journal of Mechanical Engineering, 32(1). doi: 10.1186/s10033-019-0389-8.

[15] Ren B, Luo X, Li H, Chen, J, Wang Y. (2021) Gait trajectory-based interactive controller for lower limb exoskeletons for construction workers. Computer-Aided Civil and Infrastructure Engineering, 37(5): 558–572. doi: 10.1111/mice.12756.

[16] Pina D, Gabriel J, Natal R (2019) Motion capture and multibody simulations to determine actuation requirements for an assistive exoskeleton. Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies. doi: 10.5220/0007403601830191.

[17] Gong H, Song Z, Dario P. (2021) Design of a novel wheelchair-exoskeleton robot for human multi-mobility assist. Intelligent Robotics and Applications: 281–292. doi: 10.1007/978-3-030-89134-3_26.

[18] Foroutannia A, Akbarzadeh TMR, Akbarzadeh A, Tahamipour ZSM. (2023) Adaptive fuzzy impedance control of exoskeleton robots with electromyography-based convolutional neural networks for human intended trajectory estimation. Mechatronics, 91: 102952. doi: 10.1016/j.mechatronics.2023.102952.

[19] Ali SA, Annuar KAM, Miskon MF. (2016) Trajectory planning for exoskeleton robot by using cubic and quintic polynomial equation. International Journal of Applied Engineering Research, 11(13): 7943-7946.

[20] Wu Q, Wang X, Du F, Zhang X. (2015) Design and control of a powered hip exoskeleton for walking assistance. International Journal of Advanced Robotic Systems, 12(3): 18. doi: 10.5772/59757.

[21] Yeem S, Heo J, Kim H, Kwon Y. (2019) Technical analysis of exoskeleton robot. World Journal of Engineering and Technology, 07(01): 68–79. doi: 10.4236/wjet.2019.71004.

[22] Chen G, Chan CK, Guo Z, Yu H. (2013) A review of lower extremity assistive robotic exoskeletons in rehabilitation therapy. Critical Reviews in Biomedical Engineering, 41(4–5): 343–363. doi: 10.1615/critrevbiomedeng.2014010453.

[23] Rupal BS, Rafique S, Singla A, Singla E, Isaksson M, Virk GS. (2017) Lower-limb exoskeletons. International Journal of Advanced Robotic Systems, 14(6): 172988141774355. doi: 10.1177/1729881417743554.

[24] Yuden MMA, Ghazaly MM, Amran AC, Jamaludin IW, Khoo HY, Yaacob MR, Abdullah Z, Yeo CK. (2016) Positioning control performances of a robotic hand system. Jurnal Teknologi, 79(1). doi: 10.11113/jt.v79.8726.

[25] Miao Y, Gao F, Pan D. (2013) Mechanical design of a hybrid leg exoskeleton to augment load-carrying for walking. International Journal of Advanced Robotic Systems, 10(11): 395. doi: 10.5772/57238.

[26] Guo Z, Yu H, Yin YH (2014) Developing a mobile lower limb robotic exoskeleton for gait rehabilitation. Journal of Medical Devices, 8(4). doi: 10.1115/1.4026900.

[27] Luu TP, Low KH, Qu X, Lim HB, Hoon KH. (2014) Hardware development and locomotion control strategy for an over-ground gait trainer: NaTUre-Gaits. IEEE Journal of Translational Engineering in Health and Medicine, 2: 1–9. doi: 10.1109/jtehm.2014.2303807.

[28] Yves S, Bouri M, Clavel R, Allemand Y, Brodard R. (2010) A novel verticalized reeducation device for spinal cord injuries: the walktrainer, from design to clinical trials. Robotics 2010 Current and Future Challenges. doi: 10.5772/7328.

[29] Stearns YKA, Brenner LA. (2018) Novel psychological outcomes with ekso bionics technology. Archives of Physical Medicine and Rehabilitation, 99(10). doi: 10.1016/j.apmr.2018.07.249.

[30] Strausser KA, Swift TA, Zoss AB, Kazerooni H, Bennett BC. (2011) Mobile exoskeleton for spinal cord injury: development and testing. ASME 2011 Dynamic Systems and Control Conference and Bath/ASME Symposium on Fluid Power and Motion Control, 2. doi:

10.1115/dscc2011-6042.

[31] Jyräkoski T, Merilampi S, Puustinen J, Kärki A. (2021) Over-ground robotic lower limb exoskeleton in neurological gait rehabilitation: User experiences and effects on walking ability. Technology and Disability, 33(1): 53–63. doi: 10.3233/tad-200284.

[32] Baronchelli F, Zucchella C, Serrao M, Intiso D, Bartolo M. (2021) The effect of robotic assisted gait training with lokomat® on balance control after stroke: systematic review and meta-analysis. Frontiers in Neurology, 12. doi: 10.3389/fneur.2021.661815.

[33] Ruffaldi E, Barsotti M, Leonardis D, Bassani G, Frisoli A, Bergamasco M. (2014) Evaluating virtual embodiment with the alex exoskeleton. Haptics: Neuroscience, Devices, Modeling, and Applications: 133–140. doi: 10.1007/978-3-662-44193-0_18.

[34] Sapiee MR, Marhaban MHM, Miskon MF, Ishak AJ (2020) Walking simulation model of lower limb exoskeleton robot design. Journal of Mechanical Engineering and Sciences, 14(3): 7071–7081. doi: 10.15282/jmes.14.3.2020.09.0554.

[35] Bogue R. (2017). Robots that interact with humans: a review of safety technologies and standards. Industrial Robot: An International Journal, 44(4): 395–400. doi: 10.1108/ir-04-2017-0070.

[36] Ward D, Epstein B, Tiziani L, Hammond FL. (2021) Optimal design of a mechatronic lever arm for pneumatic exoskeleton: design and validation. 2021 Design of Medical Devices Conference. doi: 10.1115/dmd2021-1093.

[37] Zoss A, Kazerooni H, Chu A. (2005) On the mechanical design of the Berkeley Lower Extremity Exoskeleton (BLEEX). 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. doi: 10.1109/iros.2005.1545453.

[38] Li Y, Guan X, Han X, Tang Z, Meng K, Shi Z, Penzlin B, Yang Y, Ren J, Yang Z, Li Z, Leonhardt S, Ji L. (2020) Design and preliminary validation of a lower limb exoskeleton with compact and modular actuation. IEEE Access, 8: 66338–66352. doi: 10.1109/access.2020.2985910.

[39] Lee H, Ferguson PW, Rosen J. (2020) Lower limb exoskeleton systems—overview. wearable robotics: 207–229. doi: 10.1016/b978-0-12-814659-0.00011-4.

[40] Song J, Zhu A, Tu Y, Zhang X, Cao G. (2023) Novel Design and control of a crank-slider series elastic actuated knee exoskeleton for compliant human–robot interaction. IEEE/ASME Transactions on Mechatronics, 28(1): 531–542. doi: 10.1109/tmech.2022.3204921.

[41] Shao Y, Zhang W, Su Y, Ding X. (2021) Design and optimisation of load-adaptive actuator with variable stiffness for compact ankle exoskeleton. Mechanism and Machine Theory, 161: 104323. doi: 10.1016/j.mechmachtheory.2021.104323.

[42] Brisson N, Krämer M, Reichenbach J, Duda G. (2021) Dynamic and guided knee motion under loading during fast magnetic resonance imaging: a novel device. Osteoarthritis and Cartilage, 29: S350–S351. doi: 10.1016/j.joca.2021.02.455.

[43] Miura T, Matsubara A, Kono D, Otaka K, Hoshide K. (2017) Design of high-precision ball screw based on small-ball concept. Precision Engineering, 47: 452–458. doi: 10.1016/j.precisioneng.2016.09.020.

[44] Sankai Y. (2010) HAL: Hybrid assistive limb based on cybernics. Springer Tracts in Advanced Robotics: 25–34. doi: 10.1007/978-3-642-14743-2_3.

[45] Sun Y, Hu J, Huang R. (2023) Negative-stiffness structure vibration-isolation design and impedance control for a lower limb exoskeleton robot. Actuators, 12(4): 147. doi: 10.3390/act12040147.

[46] Krause J. (2019) The Methodology of conventional arms control. Prospects for Conventional Arms Control in Europe: 28–62. doi: 10.4324/9780429303241-4.

[47] Zaway I, Jallouli KR, Maaleja B, Medhaffar H, Derbela N. (2022) Multi-objective fractional order pid controller optimization for kid's rehabilitation exoskeleton. International Journal of Robotics and Control Systems, 3(1): 32–49. doi: 10.31763/ijrcs.v3i1.840.

[48] Mathew M, Thomas MJ, Navaneeth M, Sulaiman S, Amudhan A, Sudheer A. (2022) A systematic review of technological advancements in signal sensing, actuation, control and training methods in robotic exoskeletons for rehabilitation. Industrial Robot: The International

Journal of Robotics Research and Application, 50(3): 432–455. doi: 10.1108/ir-09-2022-0239.

[49] Vantilt J, Tanghe K, Afschrift M, Bruijnes AK, Junius K, Geeroms J, Aertbeliën E, De Groote F, Lefeber D, Jonkers I, De Schutter J. (2019) Model-based control for exoskeletons with series elastic actuators evaluated on sit-to-stand movements. Journal of NeuroEngineering and Rehabilitation, 16(1). doi: 10.1186/s12984-019-0526-8.

[50] Kardan I, Akbarzadeh A. (2017) Robust output feedback assistive control of a compliantly actuated knee exoskeleton. Robotics and Autonomous Systems, 98: 15–29. doi: 10.1016/j.robot.2017.09.006.

[51] Villa PAC, Delisle RD, Botelho T, Mayor JJV, Delis AL, Carelli R, Frizera NA, Bastos TF. (2018) Control of a robotic knee exoskeleton for assistance and rehabilitation based on motion intention from sEMG. Research on Biomedical Engineering, 34(3): 198–210. doi: 10.1590/2446-4740.07417.

[52] Lenzi T, De Rossi SMM, Vitiello N, Carrozza MC. (2012) Intention-based emg control for powered exoskeletons. IEEE Transactions on Biomedical Engineering, 59(8): 2180–2190. doi: 10.1109/tbme.2012.2198821.

[53] Ling L, Wang Y, Ding F, Jin L, Feng B, Li W, Wang C, Li X. (2023) An efficient method for identifying lower limb behaviour intentions based on surface electromyography. computers, Materials & Continua, 77(3): 2771–2790. doi: 10.32604/cmc.2023.043383.

[54] De Miguel FJ, Lobo PJ, Prinsen E, Font LJM, Marchal CL. (2023) Control strategies used in lower limb exoskeletons for gait rehabilitation after brain injury: a systematic review and analysis of clinical effectiveness. Journal of NeuroEngineering and Rehabilitation, 20(1). doi: 10.1186/s12984-023-01144-5.

[55] Longatelli V, Pedrocchi A, Guanziroli E, Molteni F, Gandolla, M. (2021) Robotic exoskeleton gait training in stroke: an electromyography-based evaluation. Frontiers in Neurorobotics, 15. doi: 10.3389/fnbot.2021.733738.

[56] Adeola-Bello ZA, Azlan NZ. (2022) Power assist rehabilitation robot and motion intention estimation. International Journal of Robotics and Control Systems, 2(2): 297–316. doi: 10.31763/ijrcs.v2i2.650.

[57] Wickiewicz TL, Roy RR, Powell PL, Edgerton VR. (1983) Muscle architecture of the human lower limb. Clinical Orthopaedics and Related Research, 179: 275-283. doi: 10.1097/00003086-198310000-00042.

[58] Motl R, Gosney J. (2008) Effect of exercise training on quality of life in multiple sclerosis: a meta-analysis. Multiple Sclerosis Journal, 14(1): 129–135. doi: 10.1177/1352458507080464.