About the Journal

Overview

Pertanika Journal of Science & Technology (JST) is the official journal of Universiti Putra Malaysia published by UPM Press. It is an open-access online scientific journal which is free of charge. It publishes the scientific outputs. It neither accepts nor commissions third party content.

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

JST is a **quarterly** (January, April, July and October) periodical that considers for publication original articles as per its scope. The journal publishes in **English** and it is open to authors around the world regardless of the nationality.

The Journal is available world-wide.

Aims and scope

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

History

Pertanika was founded in 1978. A decision was made in 1992 to streamline Pertanika into three journals as Journal of Tropical Agricultural Science, Journal of Science & Technology, and Journal of Social Sciences & Humanities to meet the need for specialised journals in areas of study aligned with the interdisciplinary strengths of the university.

After almost 25 years, as an interdisciplinary Journal of Science & Technology, the revamped journal now focuses on research in science and engineering and its related fields.

Goal of *Pertanika*

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

Quality

We aim for excellence, sustained by a responsible and professional approach to journal publishing. Submissions are guaranteed to receive a decision within 14 weeks. The elapsed time from submission to publication for the articles averages 5-6 months.

Abstracting and indexing of *Pertanika*

Pertanika is almost 40 years old; this accumulated knowledge has resulted in Pertanika JST being abstracted and indexed in SCOPUS (Elsevier), Thomson (ISI) Web of Knowledge [BIOSIS & CAB Abstracts], EBSCO & EBSCOhost, DOAJ, ERA, Cabell's Directories, Google Scholar, MyAIS, ISC & Rubriq (Journal Guide).

Future vision

We are continuously improving access to our journal archives, content, and research services. We have the drive to realise exciting new horizons that will benefit not only the academic community, but society itself.

Citing journal articles

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

Publication policy

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

Code of Ethics

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

International Standard Serial Number (ISSN)

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

Journal of Science & Technology: ISSN 0128-7680 (*Print*); ISSN 2231-8526 (*Online*).

Lag time

A decision on acceptance or rejection of a manuscript is reached in 3 to 4 months (average 14 weeks). The elapsed time from submission to publication for the articles averages 5-6 months.

Authorship

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

Manuscript preparation

Refer to Pertanika's **INSTRUCTIONS TO AUTHORS** at the back of this journal.

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words **I**ntroduction, **M**aterials and Methods, **R**esults, **A**nd, **D**iscussion. IMRAD is simply a more 'defined' version of the "IBC" [Introduction, Body, Conclusion] format used for all academic writing. IMRAD indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: *Title, Authors, Keywords, Abstract, Conclusions*, and *References*. Additionally, some papers include Acknowledgments and Appendices.

The *Introduction* explains the scope and objective of the study in the light of current knowledge on the subject; the *Materials and Methods* describes how the study was conducted; the *Results* section reports what was found in the study; and the *Discussion* section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the Journal's **INSTRUCTIONS TO AUTHORS**.

Editorial process

Authors are notified with an acknowledgement containing a *Manuscript ID* on receipt of a manuscript, and upon the editorial decision regarding publication.

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

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

As articles are double-blind reviewed, material that might identify authorship of the paper should be placed only on page 2 as described in the first-4 page format in Pertanika's **Instructions to Authors** given at the back of this journal.

The Journal's peer-review

In the peer-review process, three referees independently evaluate the scientific quality of the submitted manuscripts.

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

Operating and review process

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

1. The Journal's chief executive editor and the editorial board examine the paper to determine whether it is appropriate for the journal and should be reviewed. If not appropriate, the manuscript is rejected outright and the author is informed.

2. The chief executive editor sends the article-identifying information having been removed, to three reviewers. Typically, one of these is from the Journal's editorial board. Others are specialists in the subject matter represented by the article. The chief executive editor asks them to complete the review in three weeks.

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

3. The chief executive editor, in consultation with the editor-in-chief, examines the reviews and decides whether to reject the manuscript, invite the author(s) to revise and resubmit the manuscript, or seek additional reviews. Final acceptance or rejection rests with the Edito-in-Chief, who reserves the right to refuse any material for publication. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the author) are forwarded to the author. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.

4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the chief executive editor along with specific information describing how they have answered' the concerns of the reviewers and the editor, usually in a tabular form. The author(s) may also submit a rebuttal if there is a need especially when the author disagrees with certain comments provided by reviewer(s).

5. The chief executive editor sends the revised paper out for re-review. Typically, at least one of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the chief executive editor in consultation with the editorial board and the editor-in-chief examine their comments and decide whether the paper is ready to be published, needs another round of revisions, or should be rejected.

7. If the decision is to accept, an acceptance letter is sent to all the author(s), the paper is sent to the Press. The article should appear in print in approximately three months.

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

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

Pertanika Journal of Science & Technology
Vol. 26 (3) Jul. 2018

Contents

Short Communication

Foreword

Welcome to the **Third Issue 2018** of the Journal of Science and Technology (JST)!

Pertanika bids farewell to Dr. Nayandeep Singh Kanwal whose service ended this month. We thank him for his contribiutions and wish him every success in his future endeavour.

JST is an open-access journal for studies in science and technology published by Universiti Putra Malaysia Press. It is independently owned and managed by the university and is run on a non-profit basis for the benefit of the world-wide science community.

This issue contains **45 articles,** of which **four** are review articles, **one** is a short communication and **40** are regular research articles. The authors of these articles hail from several countries namely, **Malaysia**, **Indonesia**, **Germany**, **Denmark**, **Afghanistan**, **Saudi Arabia**, **Canada**, **Italy**, **India**, **Iraq** and **Iran**.

The first review article in this issue reports briefly on remotely sensed imagery data application in a mangrove forest (*Zulfa, A. W.* and *Norizah, K.*), while the second is on a preliminary study on paper-sheet-based epoxy composites designed for repairing work application and their properties (*Muhamad Hellmy Hussin*). The next review article looks at platelet transcriptome-based approaches in the fight against dengue and other diseases (*Suppiah, J., Sakinah, S., Chan, S. Y., Wong, Y. P., Bala, J. A., Lawal, N., Benelli, G., Subbiah, S. K.* and *Chee, H. Y.*), while the final review article examines forensic body height estimation by measuring unsegmented fingers of Javanese subjects in Indonesia (*Athfiyatul Fatati* and *Myrtati D. Artaria*).

The short communication discusses the effectiveness of the quick coherence technique using the heart rate variability-biofeedback technology on the recovery of heart coherence among university students (*Abdul Qahar Sarwari* and *Mohammad Nubli Wahab*).

The 40 regular articles cover a wide range of topics. The first article is on the association of FDG-PET (SUVmax) and an inflammatory marker in predicting tumour aggressiveness (*Ahmad Saad, F. F., Abdullah, N. S., Shaharuddin, S.* and *Nordin, A. J.*). The following articles look at: a comparative analysis of contrast enhancement techniques for medical images (*Randeep Kaur, Meenu Chawla, Navdeep Kaur Khiva* and *Mohd Dilshad Ansari*); the preparation and thermal properties of cellulose acetate/polystyrene blend nanofibres via the electrospinning technique (*Rosdi, N. H., Mohd Kanafi, N.* and *Abdul Rahman, N.*); an experimental analysis of condensation in a helical coil tube (*Rashed Ali* and *Nitin P Gulhane*); a modified wiener filter for restoring landsat images in remote sensing applications (*Kalaivani K* and *Asnath Victy Phamila Y*); GPU-based optimisation of pilgrim simulation for *hajj* and *umrah* rituals (*Abdur Rahman Muhammad Abdul Majid, Nor Asilah Wati Abdul Hamid, Amir Rizaan Rahiman* and *Basim Zafar*); ultrasound-assisted

extraction of natural colourants from the husk of *cocos nucifera* in comparison with agitated-bed extraction (*Rodiah, M. H., Nur Asma Fhadhila, Z., Noor Asiah, H. , Aziah, M. Y.* and *Kawasaki, N.*); distinct pH-dependent aggregation of citrate-capped colloidal gold in presence of citrate competitors (*Fatemeh Javadi-Zarnaghi, Fahimeh Hosseini* and *Dorsa Mohammadrezaei*); experimental evaluation of jatropha oil methyl ester (JOME) and fish oil methyl ester (FOME) in a compression ignition engine with exhaust gas recirculation (*K. Bhaskar* and *S. Sendilvelan*); decomposition of benders for distribution networks with a cross-docking centre (*Manpreet Singh, Divya Aggarwal* and *Vijay Kumar*); distance correlation between plaintext and hash data by genetic algorithm (*Farjami, Y., Rahbari, D.* and *Hosseini, E.*); simulation of fermentation compounds for bioethanol production using different separating agents (*S. M. Anisuzzaman, D. Krishnaiah, A. Bono, F. A. Lahin* and *Syazryn R. I.*); the study of time lag on plant growth under the effect of toxic metal as a mathematical model (*Kalra, P.* and *Kumar, P.*); optimisation of the multireservoir operation policy using a teaching-learning-based optimisation algorithm (*Jayantilal N. Patel* and *Pranita N. Balve*); determination of the optimal pre-processing technique for spectral data of oil palm leaves with respect to nutrients (*Helena Anusia James Jayaselan, Wan Ishak Wan Ismail, Nazmi Mat Nawi* and *Abdul Rashid Mohamed Shariff*); a novel entropy algorithm for state sequence of the Bakis Hidden Markov Model (*Jason Chin-Tiong Chan* and *Hong Choon Ong*); a portable and low-cost multi-sensor for real-time remote sensing of water quality in agriculture (*Sandeep Bansal* and *G. Geetha*); an adaptive mechanism to optimise routing performance in mobile ad hoc networks (*B. Nithya, C. Mala* and *Abhishek Agrawal*); design of the side-sensitive group runs chart with estimated parameters based on expected average run length (*You Huay Woon*); DSSBD, an intelligent decision-support system for residual life estimation of the PN junction diode (*Shivani* and *Cherry Bhargva*); theoretical development of biaxial fabric prestressed composites under tension-tension fatigue loading (*Nawras H. Mostafa, Z. N. Ismarrubie, S. M. Sapuan* and *M. T .H. Sultan*); drying characteristics of *curcuma longa* using solar dryer (*Fhelix August Soebiantoro, Elieser Tarigan, Lie Hwa, Violita Putri Halim* and *Lanny Sapei*); exploration of tritrophic interaction for enhancing conservation biological control of insect pests in the role of analytical chemistry (*Surjani Wonorahardjo, Nurindah, Dwi Adi Sunarto, Sujak* and *Setya Ayu Aprilia*); OPH-LB, an optimal physical host for load balancing in a cloud environment (*Sakshi Chhabra* and *Ashutosh Kumar Singh*); an evaluation of network intrusion detection systems through a statistical analysis of the CIDDS-001 dataset using machine-learning techniques (*Abhishek Verma* and *Virender Ranga*); protocols-performance investigation using ad hoc WLAN for healthcare applications (*Piyush Yadav, Rajeev Agrawal* and *Komal Kashish*); detection of spam using particle swarm optimisation in feature selection (*Surender Singh* and *Ashutosh Kumar Singh*); evaluation of the ball-milling process for the production of carbon particles from rice straw waste (*Asep Bayu Dani Nandiyanto, Rosi Oktiani, Rena Zaen, Ari Arifin Danuwijaya, Ade Gafar Abdullah* and *Nuria Haristiani*); implementation of a markerless augmented reality method to visualise the philosophy of batik based

on android (*Isma Widiaty, Ivan Yustiawan, Yudi Wibisono, Ade Gafar Abdullah, Cep Ubad Abdullah* and *Lala Septem Riza*); metaheuristicopt, an R package for optimisation based on meta-heuristics algorithms (*Lala Septem Riza, Iip, Eddy Prasetyo Nugroho* and *Munir*); a natural circulation system for advanced fast reactors with lead-bismut as a coolant (*Ade Gafar Abdullah, Zaki Su'ud* and *Asep Bayu Dani Nandiyanto*); using the jolly balance spring method to determine the pure water surface tension coefficient (*Duden Saepuzaman, Muhamad Gina Nugraha, Regiana Dewi, Fitri Kafiyani* and *Fanny Herliyana Dewi*); an analysis of attacks on mail disposition systems secured by digital signatures equipped with AES and RSA algorithms (*Herbert Siregar, Enjun Junaeti* and *Try Hayatno*); the technical efficiency chemical industry in Indonesia using the stochastic frontier analysis (SFA) approach (*Amir Machmud, Asep Bayu Dani Nandiyanto* and *Puspo Dewi Dirgantari*); a validation of UML model and OCL expressions using the USE tool (*Arifa Bhutto* and *Dil Muhammad Akbar Hussain*); a photonic crystal-based micro mechanical sensor in an SOI platform (*Indira Bahaddur, Preetha Sharan* and *P. C. Srikanth*); an adaptive MOEMS-based micro pressure sensor using photonic crystal (*Johnson, O. V.* and *Preeta Sharan*); photonic crystal-based micro interferometer biochip (PC-IMRR) for early stage detection of melanoma (*Nandhini, V. L., K. Suresh Babu, Sandip Kumar Roy* and *Ketan Pandit*); cheat-proof communication through a cluster head (C3H) in a mobile ad hoc network (*Abu Sufian, Anuradha Banerjee* and *Paramartha Dutta*); and wavelength selectivity using an adaptive shortest path algorithm for optical network (*Piruthiviraj P, Preeta Sharan* and *Nagaraj Ramrao*).

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

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

I would also like to express my gratitude to all the contributors namely, the authors, reviewers and editors for their professional contribution towards making this issue feasible.

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

Chief Executive Editor
Prof. Dato' Dr. Abu Bakar Salleh
executive_editor.pertanika@upm.my

SCIENCE & TECHNOLOGY

Review Article

Remotely Sensed Imagery Data Application in Mangrove Forest: A Review

Zulfa, A. W.[1] and Norizah, K.[1,2]*

[1]*Department of Forest Management, Faculty of Forestry, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

The mangrove forest ecosystem acts as a shield against the destructive tidal waves, preventing the coastal areas and other properties nearby from severe damages; this protective function certainly deserves attention from researchers to undertake further investigation and exploration. Mangrove forest provides different goods and services. The unique environmental factors affecting the growth of mangrove forest are as follows: distance from the sea or the estuary bank, frequency and duration of tidal inundation, salinity, and composition of the soil. These crucial factors may under certain circumstances turn into obstacles in accessing and managing the mangrove forest. One effective method to circumvent this shortcoming is by using remotely sensed imagery data, which offers a more accurate way of measuring the ecosystem and a more efficient tool of managing the mangrove forest. This paper attempts to review and discuss the usage of remotely sensed imagery data in mangrove forest management, and how they will improve the accuracy and precision in measuring the mangrove forest ecosystem. All types of measurements related to the mangrove forest ecosystem, such as detection of land cover changes, species distribution mapping and disaster observation should take advantage of the advanced technology; for example, adopting the digital image processing algorithm coupled with high-resolution image available nowadays. Thus, remote sensing is a highly efficient, low-cost and time-saving technique for mangrove forest measurement. The application of this technique will further add value to the mangrove forest and enhance its in-situ conservation and protection programmes in combating the effects of the rising sea level due to climate change.

Keywords: Geoinformation technology, mangrove, remote sensing, sustainable management, tropical forest

MANGROVE FOREST

The mangrove forest forms an intertidal ecosystem represented by a variety of tree species that grows mainly along the tropical and subtropical coasts worldwide (Kovacs et al., 2011). Mangrove forest is characterised by dicotyledonous woody shrubs or trees that are virtually confined to the tropics (Hogarth, 2007). Mangrove forest is a highly productive ecosystem usually scattered along the intertidal zone of the low-energy tropical coastlines (Kathiresan et al., 2001 and Lugo et al., 1974). These unique forests grow abundantly in the saline soil and brackish water, subject to periodic fresh and salt water inundation; they are generally found along the sheltered coast where all vegetation adapt to a highly saline environment that would normally be uninhabitable for other kinds of trees (Ibharim et al., 2015).

According to Giri et al. (2011), approximately 13.7 million km^2 of mangrove forest exist worldwide in 118 countries and territories within the tropical and subtropical regions of the world in 2000. During the time this review was written, accurate, up-to date and reliable information about mangrove forest coverage worldwide was not available. Thus, this study reports the statistics of the mangrove forest 11 years ago (Spalding et al, 2010; Giri et al., 2011). According to Giri et al. (2016) and Ghosh et al. (2016), due to the climate change issues and anthropogenic factors, the mangrove forest ecosystem was under pressure. Anthropogenic activities and climate change has led to degradation, pollution, sea-level rise, coastal erosion, increased salinity, increased number of cyclones and higher levels of the storm to the mangrove forest. Over the past 3 decades, the world has lost almost 50% of the mangrove forest areas (Osti et al., 2009). Spalding et al. (2010) reported that based on the 2006 record from Food and Agriculture Organization, (FAO), Asia had the biggest mangrove forest cover, about 40.4%, followed by America 30.4%, Africa 18.4%, Africa 18.4%, Australasia 6.7%, Pacific Ocean 3.8% and Middle East 0.4% (Figure 1). Asia has 25 countries with mangrove forest under a wide range of climatic conditions such as Arid (Arabian Peninsula), Subtropical (China and Japan) and Humid tropical (Southeast Asia). Table 1 shows the 12 countries with the largest mangrove forest areas in the world; Indonesia has the biggest mangrove forest area in the world with 31,894 km^2 covering about 20.9% of 68% of the world's mangrove forest area in 2006 (Spalding et al., 2010).
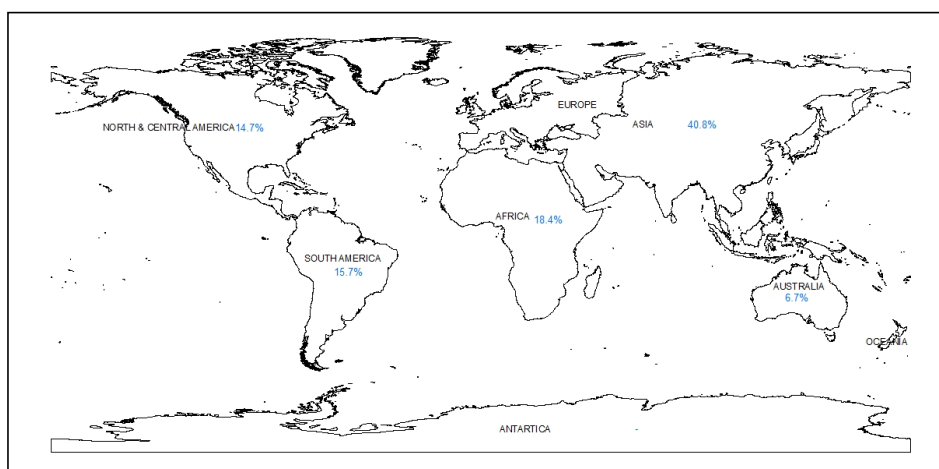
Figure 1. World mangrove forest distribution (Spalding et al., 2010)

Table 1
Mangrove forest coverage in the World (Spalding et al., 2010)

Countries	Region	Area (km^2)
Indonesia	Asia	31,894
Brazil	South America	13,000
Australia	Australia	9910
Mexico	North and Central America	7710
Nigeria	Africa	7356
Malaysia	Asia	7097
Myanmar	Asia	5029
Bangladesh	Asia	4951
Cuba	North and Central America	4944
India	Asia	4326
Papua New Guinea	Oceania	4265
Colombia	South America	4079
Total		104,561

Humans benefit in many ways from the mangrove forest ecosystem, directly or indirectly such as forest products (Kridiborworn et al., 2012; Walters, 2005; Kairo et al., 2002; Adegbehin, 1993), fisheries (Rhyma & Norizah, 2016; Thu & Populus, 2007; Manson et al., 2005; Rönnbäck, 1999; Feller & Sitnik, 1996), and protection against natural disaster (Pearce, 2014; Danielsen et al., 2005; Feller & Sitnik, 1996). From all the benefits gained and impacts reported from the mangrove forest function, the role of mangrove forest in ecological significance is very important; there is an urgent need to examine the role and function of the mangrove forest periodically. Monitoring the spatial and temporal changes of the mangrove forest area is the first step in this process. However, it is a costly and laborious task to obtain reliable, accurate and

timely information, especially when dealing with a tidal event in relation to unique mangrove forest characteristics.

By employing the manual surveying method, it is impossible to identify changes in a large mangrove forest area with a tidal event. However, with the modern technologies of monitoring such as satellite remote sensing and digital image processing algorithm, monitoring work can be done in a couple of hours or days and details can be mapped with improved accuracy and precision (Giri, 2016; Ibrahim et al., 2015; Heenkeda et al., 2014; Rao et al., 2014; Ibrahim et al., 2013; Chun et al., 2011; Wang & Sousa, 2009). These articles review and discuss the use of satellite images in the mangrove forest and the details that can be mapped from the satellite image of this era; i.e., detection of land cover changes, species distribution mapping and disaster observation.

REMOTE SENSING

Remote sensing is the science of obtaining information about objects or areas from a distance, typically from the aircraft or satellite known as a sensor. There are two categories of sensors: passive and active sensors. A passive sensor will detect the sunlight radiation reflected from the earth and thermal radiation in the visible and infrared of the electromagnetic spectrum (Barrett, 2013; Tyo et al., 2006; Turner et al., 2003; Fingas & Brown, 1997; Jackson, 1995). It does not emit its own radiation, but it receives the light and thermal radiation from the Earth's surface. Landsat, SPOT, Quickbird, IKONOS and Worldview are some examples of the passive sensors, and they have been widely used to monitor land cover and land use changes (Goetz & Dubayah, 2011; Muttitanon & Tripathi, 2005; Turner et al., 2003). An active satellite sensor functions as it emits artificial radiation to monitor the earth surface or atmospheric feature. Examples of active sensors are radar and laser scanner; Light Detection and Ranging (LIDAR), which uses a short pulse of electromagnetic radiation in the microwave spectral range. These sensors do not rely on daylight, and they are slightly affected by clouds, dust, fog, wind and weather conditions (Starek, 2016; Zink et al., 2004; Turner et al., 2003; Clothiaux et al., 2000; Fingas & Brown, 1997). Active sensors are used to identify the vegetation structure and ground surface elevation (Faridhouseini et al., 2011; Rosette et al., 2008; Holmgren & Persson, 2004; Patenaude et al., 2004; Dowling & Accad, 2003; Lefsky et al., 2002).

In principle, an object captured by a sensor can be identified from the spectral reflectance signature (or electromagnetic radiation-EMR) of the remote sensing, if the sensing system has sufficient spectral resolution to distinguish its spectrum from those of other materials (Zulfa & Norizah, 2016; Ranchin &Wald, 2000; Martin & Aber, 1997; Goetz et al., 1985). A different sensor has a different wavelength. The finer the spectral resolution and the higher the resolution, the narrower is the wavelength range for a particular channel or band; thus, objects on the Earth can be easily identifiable and differentiated as spectral resolution increases (Zhang & Zhu, 2011; Ehlers et al., 2003; Turner et al., 2003; Clark et al., 1990). To date, the spectral resolutions of some satellite sensors have up to hundreds and thousands spectral wavelength, which can be used to identify objects easily. Here are some examples of sensors that have high spectral and spatial resolutions: Worldview sensor (Kamal et al., 2014; Heenkenda et al., 2014; Heumann, 2011b), airborne sensor (Herweg et al., 2012; Tarabalka et al., 2010; Dalponte et al., 2008; Van

Aardt et al., 2007; Kamaruzaman & Kasawani, 2007; Kruse et al., 2003; Cocks et al., 1998; Martin & Aber, 1997), LiDAR sensor (Greaves et al., 2016; Wei et al., 2012; Hakala et al., 2012; Dalponte et al., 2008) and some multispectral cameras used with unmanned automated vehicles (UAV); these sensors are more flexible in terms of time of image capturing (Bareth et al., 2015; Saari et al., 2011; Mäkynen et al., 2011; Berni et al.,2009; Held et al., 2003).

The spatial resolution also plays an important role in viewing, identifying and evaluating captured images (Lillesand et al., 2014; Verma et al., 2014; Xie et al., 2008). The spatial resolution is presented by the square picture elements or pixels. The pixel size is determined by the sampling distance (Ayoub et al., 2009; Leprince et al., 2007; Schowengerdt, 2006; Kaufman et al., 1997). For example, 30 m data spatial resolution from Landsat sensor refers to data in a matrix of 30 m x 30 m pixels; this is an example of low spatial resolution image where only coarse features can be observed in the image (Lillesand et al., 2014). Meanwhile, a high spatial resolution image refers to one with a small spatial resolution size (Lillesand et al., 2014; Sawaya et al., 2003). Fine details can be seen in a high spatial resolution image, and these are some matrix examples: 0.41 m x 0.41 m pixels and 1.65 m x 1.65 m pixels respectively from panchromatic and multispectral sensor of GeoEye-1; 0.5 m x 0.5 m pixels and 2 m x 2 m pixels respectively from panchromatic and multispectral sensor of Worldview2; 0.6 m x 0.6 m pixels and 2.4 m x 2.4 m pixels respectively from panchromatic and multispectral sensor of Quickbird; and 1 m x 1 m pixels and 4 m x 4 m pixels respectively from panchromatic and multispectral sensor of IKONOS (Rhyma et al., 2015). The matrix of pixels is often called a scene. A different type of sensor has a different scene size. Sensors widely used in mangrove forest are listed in Table 2.

Table 2
Sensor used in mangrove forest mapping

Sensor	No. of band(s)	Spectral range	Spatial resolution	Sources
Spot XS	3	Red, green, blue	20 m X 20 m	Pasqualini et al. (1999); Gao, (1999); Jensen et al. (1991); Green et al. (1998);
Landsat TM	7	Blue, green, red, NIR, shortwave infrared and thermal	30 m	Rhyma et al. (2016); Che Ku Akmar et al. (2009); Alatorre et al. (2016); Kovacs et al. (2001)
Quickbird	5	Panchromatic, Red, green, blue and NIR	Panchromatic:0.65 m and Multispectral 2.62 m	Lee and Yeh. (2009) Neukermans et al. (2008); Wang et al. (2004)
IKONOS	4 bands and panchromatic	Panchromatic, Red, Green, Blue and NIR	Panchromatic:0.82 m and Multispectral 3.2 m	Huang et al. (2009); Proisy et al. (2007); Kovacs et al. (2005); Rodriguez and Feller. (2004); Wang et al. (2004); Wang et al. (2004)

Table 2 *(continue)*

Sensor	No. of band(s)	Spectral range	Spatial resolution	Sources
RapidEye	5	Blue, red, green, red edge and NIR	5 m	Son et al. (2017); Giardino et al. (2015); Roslani et al. (2014); Roslani et al. (2013); Ibrahim et al. (2013)
Spot 4 and 5	5	Monospectral, green, red, NIR and short-wave infrared	Panchromatic: 10 m and Multispectral: 20 m	Santos et al. (2015); Vo et al. (2013); Conchendda et al. (2008); Saito et al. (2003)
LiDAR		Blue, red and green	Laser scanner system: x,y,z coordinate	Wannasiri et al. (2013); Chadwick, (2011); Knight et al. (2009); Proisy et al. (2009); Zhang, (2008)
WorldView	9	Blue, red and green Panchromatic, multispectral (red, green, blue, near-infrared, coastal, yellow, red edge and near-infrared-2)	0.5 m	Wang et al. (2015); Zhu et al.,(2015); Hassan et al. (2014); Heenkenda et al. (2014); Kamal et al. (2013); Kux and Souza. (2012); Heumann. (2012)

With the advances and precision of remotely sensed imagery data and the modern technology of digital image processing algorithm (Giri, 2016), the application in management, monitoring and mapping of mangrove forest ecosystem has been proven very valuable. The use of remotely sensed imagery offers many advantages because obtaining information and performing observations by satellite sensors are beyond human ability. With the advances in the spatial and spectral resolution of remotely sensed imagery data, the application of satellite image in the mangrove forest area has been intensified. The following section reviews the use of remotely sensed imagery data with image/data analysis, and it discusses details mapped from the mangrove forest.

Early Application in Mangrove Forest

Traditional approaches to mangrove forest remote sensing has been described by the uses of pure visual imagery; aerial photography (AP) as primary source particularly for surveys conducted before 1990s (Kuenzer et al., 2011; Martinussi et al., 2009; Newton et al., 2009). Limited and lack of utility in data sources make mapping of mangrove forest difficult or even impossible to be surveyed. As technology has grown rapidly with the importance of the uses with mangrove forest, digital imagery; thematic mapper has been evolved to describe and monitor a variety of systems on a local or global scale (Newton et al., 2009; Farid, 2002). There are still limitations related to the early application of digital imagery in mangrove forest such as pointed by Adam et al. (2010) and Ozesmi and Bauer (2002). Their paper stresses that these limitations were related to spatial and spectral resolution where differentiating the spectral reflectance is difficult

to perform especially when dealing with low resolution images and the fact that vegetation in mangrove forest has the same basic components that contribute to its spectral reflectance (Kokaly et al., 2003; Price, 1992). Spectral reflections are measurements of the spectral response of different features (in case of mangrove forest, the feature refers to vegetation species) in the bands of the satellite image. Although digital imagery has evolved with the high resolution of data - discussed in the following section - AP sometimes excel in monitoring small area of mangrove forest and a number of classification are easily distinguishable (Heumann 2011b; Ozesmi & Bauer, 2002; Sulong et al., 2002).

Recently, the decline of mangrove forest has become major environmental issues worldwide. Mangrove forest become severe due to human activity or natural disaster. The needs to study mangrove forest highly important and the application of remote sensing has been widely used over a few decades ago with the development from pure visual imagery to multi-spectral imagery to the advances of narrow spectral imagery. Table 3 shows some of early remote sensing application in mangrove forest.

Table 3
Early remote sensing system and mangrove studies (Heumann, 2011b)

Sensor(s)	Studies
Aerial photography	Hossain et al. (2009); Eslami-Andargoli et al. (2009); Everitt et al. (2007); Thampanya et al. (2007); Dahdouh-Guebas et al. (2006); Benfield et al. (2005); Rodriguez and Feller (2004; Fromard et al. (2004); Jones et al. (2004); Krause et al. (2004); Manson et al. (2001); Murray et al. (2003); Chauvaud et al. (1998); Sulong (1999); Sulong and Ismail (1990)
Landsat MSS, TM, or ETMþ	Long et al. (2011); Che Ku Akmar et al., (2009); Alatorre et al., (2016; Beland et al. (2006); Cornejo et al. (2005); Giri et al. (2008); Green et al. (1998); James et al. (2007); Krause et al. (2004); Lee and Yeh (2009); Liu et al. (2008); Long and Skewes (1996); Kovacs et al., (2001); Manson et al. (2001); Mumby et al. (1999); Paling et al. (2008), Ruiz-Luna and Berlanga-Robles (2003); Vasconcelos et al. (2002)
SPOT HVR, HRVIR, or HRG	Chauvaud et al. (2001); Gao (1998, 1999); Green et al. (1998); Lee and Yeh (2009); Mumby et al. (1999); Rasolofoharinoro et al. (1998); Saito et al. (2003)
ASTER	Al Habshi et al. (2007); Vaiphasa et al. (2006)
IRS C or D	Mantri and Mishra (2006); Pattanaik et al. (2008); Ramachandran et al. (1998); Reddy and Pattanaik (2007)

Mangrove Forest Mapping

Monitoring and/or mapping mangrove forest using remotely sensed imagery data has been described as moderate and sometimes poor by Kamal et al. (2015), Wang et al. (2004), Liu et al. (2008) and Heumann (2011b), due to the presence of homogenous species in the mangrove forest and due to the limited spectral signature and spatial resolution of conventional imagery (Wang et al., 2004; Chun et al., 2011). The complexity of separating the spectral reflectance between species is the reason for their report. According to Ajithkumar et al. (2008) and Blasco and Aizpuru (2002), the spectral reflectance of the mangrove leaf is affected by the chlorophyll

content; a high chlorophyll concentration will give lower reflectance value, and thus it is difficult to discriminate the mangrove species. On the other hand, others reported an opposite view, stating that the use of remotely sensed imagery data is easy in mangrove forest mapping (Giri, 2016) since the mangrove forest possesses a very distinct spectral signature. The general consensus seems to be that mangrove forest mapping is not straightforward with the remote sensing application. It may be based on the precision of the image, resolution, processing algorithm, or expertise in observing the data; in addition, it might be affected by the different location, as a different location has different vegetation composition and structure (Hossain & Nuruddin, 2016; Ghosh et al., 2016; Matsui et al., 2015; Heumann, 2011b; Adam et al., 2010).

A recent trend in processing the satellite image for mangrove forest is to perform it using science knowledge and engineering technology. Over the past few decades, innovations in remote sensing sensors and systems such as very High Resolution System (VHR) and Synthetic Aperture Radar (SAR) (i.e., Quickbird, IKONOS, GeoEye-1, Worldview-3, PRISM-ALOS PALSAR, ASAR ENVISAT) and airborne sensors (i.e., hyperspectral remote sensing) are a breakthrough due to their high resolution sensor and continuous spectral data that are helpful in discriminating features having similar spectra in the multispectral domain (Rhyma et al., 2016; Prasad et al., 2014). In parallel with the advances of sensors and systems that are extensively applied in mangrove forest, analysis techniques in order to improve the accuracy of mangrove forest classification have also been developed such as object-based classifications integrated with one of these methods: pixel-based classification (Walter, 2004), decision tree learning analysis of pixel-based classification (Liu et al., 2008a), receiver operating characteristics (ROC) curve analysis of spectral analysis (Alatorre et al., 2011), threshold and fuzzy rule classification approaches with that of the pixel-based (Hussain et al. 2013), and support vector machine (SVM) approach of object-based classification (Liu et al., 2008b, Heumann (2011a), Vidhya et al., 2014). The following section reviews details mapped from the mangrove forest with a number of examples of image/data analysis techniques.

Land Cover Changes. Mangrove forests have been altered by direct or indirect uses of the environment, socio-economic and natural resources. The mangrove forest change dynamics worldwide due to natural and anthropogenic forces such as reported by Misra et al. (2015) and Giri et al. (2007) in India, Ibrahim et al. (2015) and Abdullah and Nakagosi (2007) in Malaysia, Muttitanon and Tripathi (2005) in Thailand, Kirui et al. (2013) in Kenya, Nguyen (2014), Nguyen et al. (2013), and Thu and Populus (2007) in Vietnam, Souza-Filho and Paradella (2003) in Brazil. Due to the dynamic changes, area extent and distribution of mangrove forest need to be monitored as frequently as possible for management and conservation. Image classification technique is the main obstacle to have an accurate land use/cover detection in mangrove forest. Nowadays, with advance algorithm and/or procedure in image processing makes mangrove forest distinguishable from other land cover in one scene.

Alatorre et al. (2016) studied the temporal evolution of vegetation activity of mangroves in the South-eastern coastal area of the Gulf of California, Mexico by using multi-temporal Landsat TM images for 20 years (1990-2010). They used NDVI analysis to detect the changes within 20 years and used multivariate regression analysis to show the coverage of the mangrove forest. From the pixel-by-pixel spatial analysis complemented through image interpretation

they conducted, they found that shrimp farms in the study area showed a spatial relationship with the zones of the great loss of vegetation activity. Meanwhile, Kanniah et al. (2015) used a maximum likelihood classification (MLC) and support vector machine (SVM) to classify the mangrove forest areas in Southern Johore, Malaysia to analyse the changes over a period of 25 years. Between these two techniques, MLC was reported to provide the significantly higher user, producer and overall accuracy compared with SVM. In different mangrove forest area in North of Malaysia, Perak, Ibrahim et al. (2015) also used MLC to classify land use and land cover for 18 years from Landsat TM and RapidEye imageries. To ensure the accuracy of their classification, normalised different vegetation index (NDVI) technique have been used. When dealing with area classification using remotely sensed imageries data, the accuracy of land cover changes for a certain period of time series with a number of periods will be affected with cloud cover. Kirui et al. (2013) in their study have paired (overlay and differentiate) all images for each time series to remove the areas with cloud cover, leaving images as cloud free over the same locations for area changes assessment. This study used MLC for classification as well.

Mathematical approach in image classification to enhance the accuracy to detect land use/cover changes in mangrove forest has been widely used by several researchers. For example, Misra et al. (2015) used Principal Component Analysis (PCA) as a pre-classification and continue the common classification; supervised and unsupervised classification. Their accuracy reported as 95.56% (Kappa=0.0556), 92.93% (Kappa=0.92), 84.64% (Kappa=0.85) and 86.36% (Kappa =0.85) for four images in 2011, 2001, 1989 and 1973. According to Kanellopulos and Wilkinson (1997) and Civco (1993), advance algorithm coupled with artificial intelligence using software can be used reliably for routing operational requirement in remote sensing and will provide more accurate and useful data.

Species Distribution Mapping. From the previous research, a number of researchers have established a method that is able to distinguish the mangrove species from other species with the laboratory measurement of hyperspectral leaf reflectance (Wang & Sousa, 2016; Zhang, Kovacs, Liu, Flores-Verdugo, & Flores-de-Santiago, 2014; Chun et al., 2011; Chun et al., 2015; Neukermans et al., 2008). This is important as it provides reliable and accurate information about mangrove forest. According to Asner et al. (2009), having maps of individual tree locations is fundamental to understanding forest responses to global change, providing a basis for monitoring species distribution patterns, responding to stress, disease, and exotic species spread and deforestation.

Wang and Sousa (2009) find that the accuracy of classification of mangrove species increased by using a narrow band of hyperspectral data. They conducted a laboratory study of mangrove leaves using a high-resolution spectrometer in the Carribean coast. For bands that have significant difference (P value <0.01) in the mean reflectance across tree species measured, linear discriminant analysis (LDA) is performed to detail the classified the mangrove species. Eventually, the most useful bands for mangrove species classification are found to be at 780, 790, 800, 1480, 1530 and 1550 nm. In addition, their study used four narrow band ratios (R_{695}/R_{420}, R_{605}/R_{760}, R_{695}/R_{760} and R_{710}/R_{760}) to diagnose stress condition across mangrove species, and results revealed that at least one ratio index was proven useful from ANOVA. In a paper by Zhang et al. (2014), FieldSpec® 3JR spectrometer was used to examine the mangrove species

in Mexico. Seven wavebands (520, 560, 650, 710, 760, 2100 and 2230 nm) are selected based on the principle component analysis and stepwise discriminant analyses to classify mangrove species. The waveband selected is able to identify the mangrove species and mangrove forest conditions with an overall accuracy of higher than 90% and a Khat coefficient higher than 0.9. Their study also examined the stress condition of the mangrove forest (poor and dwarf), and it is found to be satisfactory with accuracy higher than 80%. Roslani et al. (2014) used RapidEye satellite data in their study, and they concluded that using a high-resolution image, mangrove forest classification can be mapped with reliable information. In their study, textural analysis is used to make the classification. Their results show that the textured image produced high overall classification assessment recorded at 84%, and kappa statistics recorded at 0.8016. Meanwhile, the non-textural image produced about 80% of the overall accuracy and kappa statistics of 0.7061.

Advanced processing technique and algorithm can also yield reliable and accurate information of mangrove forest instead of using high-resolution remotely sensed imagery data. The classification approach, a part of the processing analysis, is found to influence discriminating the mangrove species. Kamal et al. (2015) stress that using a number of satellite data, various spatial and spectral resolution, and mapping technique would provide effective multi-scale mangrove forest composition mapping. Their study used the object-based approach to classify the mangrove species in the Moretan Bay, Australia by using various satellite data such as Landsat TM, ALOS AVNIR-2, WorldView-2, and LiDAR. Ghosh et al. (2016) in Sundarbans use the maximum likelihood classifier technique to classify objects from Landsat satellite image and utilise the post-classification comparison techniques to detect changes at the species level. From their study, accuracy rates of about 72%, 83%, 79% and 89% are reported from the images of 1977, 1989, 2000 and 2015 respectively. A total of five major species are detected from the Landsat image. Wang et al. (2016) explore the use of textural and differential spectral features classification technique to discriminate species that are complex to be identified. Their study use the WorldView-3 image in Hong Kong. They find that the differential spectral features could aid in reducing inner-species variability and increasing intra-species separation that might be due to the different arrangement of leaves, the branch density, and the average height and size of plants. Zhu et al. (2015) use a back-propagation artificial-neural-network (BP ANN) to accurately estimate the uneven-aged and dense mangrove forest biomass at the individual species level. Their study show a lower residual mean square error (RMSE), of about 19.17% from the classification conducted to estimate the biomass. Umroh et al. (2016) use the standard false colour composite of Landsat band 564 to map the mangrove forest areas in Pongok Island, South Bangka, Indonesia. They use NDVI value ranging from -1 to 0.33 to represent the area dominated with <1000 trees/ha; NDVI 0.33-0.42 to represent the area dominated with >1000 to <1500 trees/ha; and NDVI 0.42-1 to represent the area dominated with >1500 trees/ha. Their study recognized that Pongok Island are dominated by *Rhizophora* sp., *Avicennia* sp. and *Bruguiera* sp. Study by Heenkenda et al. (2014) used WorldView 2 and high-resolution aerial photograph to discriminate mangrove species in Australia. Initial step in their image analysis use object-based image classification to determine mangrove forest and non-mangrove forest area. Later, support vector machine

algorithm with best-fit parameters are used to classify individual species. Species classified with WorldView 2 image show acceptable accuracy with 89%.

Disaster Management. The mangrove forest serves as a form of protection, especially sheltering the coastal inhabitants from natural disasters. The natural disaster (i.e., sea level rise) are mostly as a result of climate change. Advancement in remote sensing; spectral and spatial resolution including techniques in image processing have provided an opportunity to observe and monitor mangrove forest from local to global scales with unprecedented spatial and thematic detail. Practically, remote sensing application to map disaster observation are based on preparedness to support disaster and management of the risk. For example, Giri (2016) points out that the level of disaster protection will depend on the size of the tsunami; vegetation structure may influence the role of mangrove forest as a form of disaster protection. Thus, it is necessary to know the extent of the area and changes of mangrove forest to manage the probability of risk from disaster-tsunami and early warning for resource planning. While the remote sensing of disaster observation and early warning are major areas of research for terrestrial areas (Imen et al., 2015; Agatsiva & Oroda, 2000), relatively little research has been done on mangrove forest.

REMOTE SENSING AND MANGROVE FOREST; WAY FORWARD

Mapping extent and changes, species composition and disaster observation of mangrove forest by using remote sensing need to be significantly improved worldwide. Practical method in improving the accuracy of classification at early stage of remote sensing analysis show an extensive study done by innumerable researchers. Innovation and emerging of sensors and systems such as VHR, SAR, unmanned aerial vehicles (UAV) and LiDAR provide users with a high resolution of spatial and spectral imagery allowing them to map mangrove forest in detail and accurately. The UAV and LiDAR are a novel remote sensing platform to have the capability to store a large volume of data using cloud computing. Thus, management, monitoring work, resource planning and conservation become easier than before for a forest that has a different level of stress due to its inaccessibility. Recognising the features or objects that exist within a set of remotely satellite imagery data nowadays has become more advanced. Integration of mathematical algorithm- artificial intelligence with image processing have been developed to exploit the uses of VHR and SAR data. For example, individual species of mangrove can be discriminated by using PCA to perform linear discriminate analysis and/or stepwise discriminate analysis for narrow band imagery and use multiple sensors with object-based analysis. When dealing with species variability- due to similar chemical component for mangrove vegetation, textural and differential spectral classification techniques has been introduced. While, biomass estimation at species level has been assessed by artificial intelligence of BP ANN. Such advance image processing shows that the science of remote sensing and mangrove forest has advanced in the last decade. Yet, there are still gaps and limitations to overcome mangrove forest as well as in terrestrial remote sensing (Wang et al., 2009). Opportunities always exist in technology. For mangrove forest, this is not exempted. Heumann (2011b) suggested five opportunities to be improved in order to enhance remote sensing in mangrove forest; i) apply the existing

sensors in mangrove forest that has never been used before, ii) use the existing methods of image processing from terrestrial forests into mangrove forest, iii) investigate new sensors that able to provide high quality data of mangrove forest, iv) integrate multiple types of remotely sensed imagery data to improve the accuracy of data processing, and v) global monitoring of mangrove forest to understand the extent and changes of mangrove forest worldwide including the structure, function and ecosystem services.

CONCLUSION

The purpose of this review paper is to provide a comprehensive overview of the usage and application of remote sensing, especially the mangrove forest. Remote sensing is a tool for accomplishing many objectives related to the mangrove forest ecosystem. The mangrove forest ecosystem is unique and difficult to be measured at the terrestrial landscape; aerial measurement is a very good alternative. Details can be measured by using remotely sensed imagery data with multiple available up-to-date sensors. Key requirements for a sustainable mangrove forest management include high-resolution images, advanced image analysis algorithm and integration with precision tools to improve spectral properties identification and characterisation. Since aerial measurement deals with the upper surface of an area, canopy reflectance, along with wet soil and water characteristics, was expected to distort the image analysis processing. Notwithstanding the above shortcoming, the use of remote sensing in managing the mangrove forest remains advantageous. It relies on the technical support drawn from the continuous research in the techniques and methods for image analyses, particularly in differentiating the biochemical and biophysical canopy attributes, which can improve the accuracy and precision of mangrove forest mapping and monitoring. It is important to carry out research in the mangrove forest ecosystem so that the quality of the mangrove forest can be maintained, and it continues to play its protective function against the effects of sea-level rise due to the increasing emission of atmospheric heat from the climate change.

REFERENCES

Abdullah, S. A., & Nakagoshi, N. (2007). Forest fragmentation and its correlation to human land use change in the state of Selangor, peninsular Malaysia. *Forest Ecology and Management, 241*(1), 39-48.

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management, 18*(3), 281-296.

Adegbehin, J. O. (1993). Mangroves in Nigeria. *Conservation and sustainable utilization of mangrove forests in Latin America and African Regions (Part 2: Africa). Mangrove Ecosystem Technical Reports, 3*, 135-153.

Adi, W., & Sari, S. P. & Sari, S. P. (2016). Detection of mangrove distribution in Pongok Island. *Procedia Environmental Sciences, 33*, 253-257.

Agatsiva, J., & Oroda, A. (2000). Remote sensing and GIS in the development of a decision support system for sustainable management of a drylands of eastern Africa: a case of the Kenyan drylands. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34*(6), 42-49.

Ajithkumar, T. T., Thangaradjou, T., & Kannan, L. (2008). Spectral reflectance properties of mangrove species of the Muthupettai mangrove environment, Tamil Nadu. *Journal of Environmental Biology, 29*(5), 785-788.

Alatorre, L. C., Sánchez-Carrillo, S., Miramontes-Beltrán, S., Medina, R. J., Torres-Olave, M. E., Bravo, L. C., ... & Uc, M. (2016). Temporal changes of NDVI for qualitative environmental assessment of mangroves: Shrimp farming impact on the health decline of the arid mangroves in the Gulf of California (1990–2010). *Journal of Arid Environments, 125*, 98-109.

Al Habshi, A., Youssef, T., Aizpuru, M., & Blasco, F. (2007). New mangrove ecosystem data along the UAE coast using remote sensing. *Aquatic Ecosystem Health and Management, 10*(3), 309-319.

Aschbacher, J., Ofren, R., Delsol, J. P., Suselo, T. B., Vibulsresth, S., & Charrupat, T. (1995). An integrated comparative approach to mangrove vegetation mapping using advanced remote sensing and GIS technologies: preliminary results. *Hydrobiologia, 295*(1-3), 285-294.

Asner, G. P., Knapp, D. E., Balaji, A., & Páez-Acosta, G. (2009). Automated mapping of tropical deforestation and forest degradation: CLASlite. *Journal of Applied Remote Sensing, 3*(1), 033543.

Ayoub, F., Leprince, S., & Avouac, J. P. (2009). Co-registration and correlation of aerial photographs for ground deformation measurements. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(6), 551-560.

Bareth, G., Aasen, H., Bendig, J., Gnyp, M. L., Bolten, A., Jung, A., … & Soukkamäki, J. (2015). Low-weight and UAV-based hyperspectral full-frame cameras for monitoring crops: Spectral comparison with portable spectroradimeter measurements. *Photogrammetrie Fernerkundung- Geoinformation, 2015*(1), 69-79.

Barrett, E. C. (2013). *Introduction to environmental remote sensing.* New York, NY: Routledge.

Beland, M., Goita, K., Bonn, F., & Pham, T. T. H. (2006). Assessment of land-cover changes related to shrimp aquaculture using remote sensing data: a case study in the Giao Thuy District, Vietnam. *International Journal of Remote Sensing, 27*(8), 1491-1510.

Benfield, S. L., Guzman, H. M., & Mair, J. M. (2005). Temporal mangrove dynamics in relation to coastal development in Pacific Panama. *Journal of Environmental Management, 76*(3), 263-276.

Berni, J. A. J., Zarco-Tejada, P. J., Suárez, L., González-Dugo, V., & Fereres, E. (2009). Remote sensing of vegetation from UAV platforms using lightweight multispectral and thermal imaging sensors. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Hannover, Germany.

Blasco, F., & Aispuru, M. (2002). Mangroves along the coastal stretch of the Bay of Bengal: Present status. *Indian Journal of Marine Sciences, 31*(1), 9-20.

Chadwick, J. (2011). Integrated LiDAR and IKONOS multispectral imagery for mapping mangrove distribution and physical properties. *International jJournal of Remote Sensing, 32*(21), 6765-6781.

Chauvaud, S., Bouchon, C., & Maniere, R. (1998). Remote sensing techniques adapted to high resolution mapping of tropical coastal marine ecosystems (coral reefs, seagrass beds and mangrove). *International Journal of Remote Sensing, 19*(18), 3625-3639.

Chauvaud, S., Bouchon, C., & Maniere, R. (2001). Thematic mapping of tropical marine communities (coral reefs, seagrass beds and mangroves) using SPOT data in Guadeloupe Island. *Oceanologica Acta, 24*, S3-S16.

Che Ku Akmar, C. K. O., Mohd Hasmadi, I., Kasawani, I., Norsaliza, U., & Kamaruzaman, J. (2009). Comparison of several vegetation indices for mangrove mapping using temotely sensed data. In *Environmental Science and Technology Conference (ESTEC2009)* (pp. 620–628). Kuala Terengganu Malaysia.

Chun, B. B., Jafri, M. Z. M., & San, L. H. (2011, July 12-13). Reflectance characteristic of certain mangrove species at Matang Mangrove Forest Reserve, Malaysia. In *Space Science and IEEE. Communication (IconSpace), 2011 IEEE International Conference on* (pp. 147-151). IEEE, Penang, Malaysia.

Chun, B. B., Keat, S. C., Syahreza, S., Jafri, M. Z. M., & San, L. H. (2015, April 24). Discrimination of mangrove species in Matang Mangrove Forest Reserve, Perak using in-situ measurement of hyperspectral leaf reflectance. In *AIP Conference Proceedings* (Vol. 1657, No. 1, p. 110004). AIP Publishing, United States of America.

Civco, D. L. (1993). Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Science, 7*(2), 173-186.

Clark, R. N., Gallagher, A. J., & Swayze, G. A. (1990, June). Material absorption band depth mapping of imaging spectrometer data using a complete band shape least-squares fit with library reference spectra. In *Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop* (pp. 176-186). JPL Publication, California, United States of America.

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., & Martner, B. E. (2000). Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *Journal of Applied Meteorology, 39*(5), 645-665.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., & Shields, T. (1998, October). The HyMapTM airborne hyperspectral sensor: the system, calibration and performance. In *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy* (pp. 37-42). EARSeL, University of Zurich, Switzerland.

Conchedda, G., Durieux, L., & Mayaux, P. (2008). An object-based method for mapping and change analysis in mangrove ecosystems. *ISPRS Journal of Photogrammetry and Remote Sensing, 63*(5), 578-589.

Dahdouh-Guebas, F. (2002). The use of remote sensing and GIS in the sustainable management of tropical coastal ecosystems. *Environment, Development and Sustainability, 4*(2), 93-112.

Dahdouh-Guebas, F., Verheyden, A., Kairo, J. G., Jayatissa, L. P., & Koedam, N. (2006). Capacity building in tropical coastal resource monitoring in developing countries: a re-appreciation of the oldest remote sensing method. *The International Journal of Sustainable Development and World Ecology, 13*(1), 62-76.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing, 46*(5), 1416-1427.

Danielsen, F., Sørensen, M. K., Olwig, M. F., Selvam, V., Parish, F., Burgess, N. D., ... & Quarto, A. (2005). The Asian tsunami: A protective role for coastal vegetation. *Science (Washington), 310*(5748), 643.

Dowling, R., & Accad, A. (2003). Vegetation classification of the riparian zone along the Brisbane River, Queensland, Australia, using light detection and ranging (lidar) data and forward looking digital video. *Canadian Journal of Remote Sensing, 29*(5), 556-563.

Ehlers, M., Gähler, M., & Janowsky, R. (2003). Automated analysis of ultra high resolution remote sensing data for biotope type mapping: New possibilities and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing, 57*(5), 315-326.

Eslami-Andargoli, L., Dale, P. E. R., Sipe, N., & Chaseling, J. (2009). Mangrove expansion and rainfall patterns in Moreton Bay, southeast Queensland, Australia. *Estuarine, Coastal and Shelf Science, 85*(2), 292-298.

Everitt, J. H., Yang, C., Summy, K. R., Judd, F. W., & Davis, M. R. (2007). Evaluation of color-infrared photography and digital imagery to map black mangrove on the Texas Gulf Coast. *Journal of Coastal Research, 23*(1), 230-235.

Faridhouseini, A., Mianabadi, A., Bannayan, M., & Alizadeh, A. (2011). Lidar remote sensing for forestry and terrestrial applications. *International Journal of Applied Environmental Sciences, 6*(1), 99-114.

Feller, I. C., & Sitnik, M. (1996). Mangrove ecology: A manual for a field course. *Smithsonian Institution, Washington, DC*, 1-135.

Fingas, M. F., & Brown, C. E. (1997). Review of oil spill remote sensing. *Spill Science and Technology Bulletin, 4*(4), 199-208.

Fromard, F., Vega, C., & Proisy, C. (2004). Half a century of dynamic coastal change affecting mangrove shorelines of French Guiana: A case study based on remote sensing data analyses and field surveys. *Marine Geology, 208*(2), 265-280.

Gao, J. (1998). A hybrid method toward accurate mapping of mangroves in a marginal habitat from SPOT multispectral data. *International Journal of Remote Sensing, 19*(10), 1887-1899.

Gao, J. (1999). A comparative study on spatial and spectral resolutions of satellite data in mapping mangrove forests. *International Journal of Remote Sensing, 20*(14), 2823-2833.

Ghosh, M. K., Kumar, L., & Roy, C. (2016). Mapping long-term changes in mangrove species composition and distribution in the Sundarbans. *Forests, 7*(12), 305-321.

Giardino, C., Bresciani, M., Fava, F., Matta, E., Brando, V. E., & Colombo, R. (2015). Mapping submerged habitats and mangroves of Lampi Island Marine National Park (Myanmar) from in situ and satellite observations. *Remote Sensing, 8*(1), 2-14.

Giri, C., & Long, J. (2016). Is the Geographic Range of Mangrove Forests in the Conterminous United States Really Expanding? *Sensors, 16*(12), 2010-2026.

Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., ... & Duke, N. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography, 20*(1), 154-159.

Giri, C., Pengra, B., Zhu, Z., Singh, A., & Tieszen, L. L. (2007). Monitoring mangrove forest dynamics of the Sundarbans in Bangladesh and India using multi-temporal satellite data from 1973 to 2000. *Estuarine, Coastal and Shelf Science, 73*(1), 91-100.

Giri, C., Zhu, Z., Tieszen, L. L., Singh, A., Gillette, S., & Kelmelis, J. A. (2008). Mangrove forest distributions and dynamics (1975–2005) of the tsunami-affected region of Asia. *Journal of Biogeography, 35*(3), 519-528.

Goetz, A. F., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for earth remote sensing. *Science, 228*(4704), 1147-1152.

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management, 2*(3), 231-244.

Greaves, H. E., Vierling, L. A., Eitel, J. U., Boelman, N. T., Magney, T. S., Prager, C. M., & Griffin, K. L. (2016). High-resolution mapping of aboveground shrub biomass in Arctic tundra using airborne lidar and imagery. *Remote Sensing of Environment, 184*, 361-373.

Green, E. P., Clark, C. D., Mumby, P. J., Edwards, A. J., & Ellis, A. C. (1998). Remote sensing techniques for mangrove mapping. *International Journal of Remote Sensing, 19*(5), 935-956.

Hakala, T., Suomalainen, J., Kaasalainen, S., & Chen, Y. (2012). Full waveform hyperspectral LiDAR for terrestrial laser scanning. *Optics Express, 20*(7), 7119-7127.

Hassan, N., Hamid, J. R. A., Adnan, N. A., & Jaafar, M. (2014, August 26-29). Delineation of wetland areas from high resolution WorldView-2 data by object-based method. In *IOP Conference Series: Earth and Environmental Science* (Vol. 18, No. 1, p. 012017). IOP Publishing, Kuching, Sarawak, Malaysia.

Heenkenda, M. K., Joyce, K. E., Maier, S. W., & Bartolo, R. (2014). Mangrove species identification: Comparing WorldView-2 with aerial photographs. *Remote Sensing, 6*(7), 6064-6088

Held, A., Ticehurst, C., Lymburner, L., & Williams, N. (2003). High resolution mapping of tropical mangrove ecosystems using hyperspectral and radar remote sensing. *International Journal of Remote Sensing, 24*(13), 2739-2759.

Hernández Cornejo, R., Koedam, N., Ruiz Luna, A., Troell, M., & Dahdouh-Guebas, F. (2005). Remote sensing and ethnobotanical assessment of the mangrove forest changes in the Navachiste-San Ignacio-Macapule lagoon complex, Sinaloa, Mexico. *Ecology and Society, 10*(1), 16.

Herweg, J. A., Kerekes, J. P., Weatherbee, O., Messinger, D., van Aardt, J., Ientilucci, E., ... & Meola, J. (2012, May 24). Spectir hyperspectral airborne rochester experiment data collection campaign. In *SPIE Defense, Security, and Sensing* (pp. 839028-839028). International Society for Optics and Photonics, Maryland, United States of America.

Heumann, B. W. (2011a). An object-based classification of mangroves using a hybrid decision tree— Support vector machine approach. *Remote Sensing, 3*(11), 2440-2460.

Heumann, B. W. (2011b). Satellite remote sensing of mangrove forests: Recent advances and future opportunities. *Progress in Physical Geography, 35*(1), 87-108.

Hogarth, P. J. (2007). *The biology of mangroves and seagrasses (Biology of Habitats)*. New York, NY: Oxford University Press.

Holmgren, J., & Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment, 90*(4), 415-423.

Hossain, M. D., & Nuruddin, A. A. (2016). Soil and mangrove: A review. *Journal of Environmental Science and Technology, 9*(2), 198.

Hossain, M. Z., Tripathi, N. K., & Gallardo, W. G. (2009). Land use dynamics in a marine protected area system in lower Andaman coast of Thailand, 1990–2005. *Journal of Coastal Research, 25*(5), 1082-1095.

Huang, X., Zhang, L., & Wang, L. (2009). Evaluation of morphological texture features for mangrove forest mapping and species discrimination using multispectral IKONOS imagery. *IEEE Geoscience and Remote Sensing Letters, 6*(3), 393-397.

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing, 80*, 91-106.

Ibharim, N. A., Mustapha, M. A., Lihan, T., & Mazlan, A. G. (2015). Mapping mangrove changes in the Matang Mangrove Forest using multi temporal satellite imageries. *Ocean and Coastal Management, 114*, 64-76.

Ibrahim, N. A., Mustapha, M. A., Lihan, T., & Ghaffar, M. A. (2013, November 31). Determination of mangrove change in Matang Mangrove Forest using multi temporal satellite imageries. In *AIP Conference Proceedings* (Vol. 1571, No. 1, pp. 487-492). AIP Publishing, United States of America.

Ibrahim, S., & Hashim, I. (1990). Classification of mangrove forest by using 1: 40 000-scale aerial photograph. *Forest Ecology and Management, 33*, 583-592.

Imen, S., Chang, N. B., & Yang, Y. J. (2015). Developing the remote sensing-based early warning system for monitoring TSS concentrations in Lake Mead. *Journal of Environmental Management, 160*, 73-89.

Jackson, T. J., Le Vine, D. M., Swift, C. T., Schmugge, T. J., & Schiebe, F. R. (1995). Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92. *Remote Sensing of Environment, 54*(1), 27-37.

James, G. K., Adegoke, J. O., Saba, E., Nwilo, P., & Akinyede, J. (2007). Satellite-based assessment of the extent and changes in the mangrove ecosystem of the Niger Delta. *Marine Geodesy, 30*(3), 249-267.

Jensen, J. R., Lin, H., Yang, X., Ramsey III, E., Davis, B. A., & Thoemke, C. W. (1991). The measurement of mangrove characteristics in southwest Florida using SPOT multispectral data. *Geocarto International, 6*(2), 13-21.

Jones, J., Dale, P. E. R., Chandica, A. L., & Breitfuss, M. J. (2004). Changes in the distribution of the grey mangrove Avicennia marina (Forsk.) using large scale aerial color infrared photographs: are the changes related to habitat modification for mosquito control? *Estuarine, Coastal and Shelf Science, 61*(1), 45-54.

Kairo, J. G., Dahdouh-Guebas, F., Gwada, P. O., Ochieng, C., & Koedam, N. (2002). Regeneration status of mangrove forests in Mida Creek, Kenya: a compromised or secured future? *AMBIO: A Journal of the Human Environment, 31*(7), 562-568.

Kamal, M., Phinn, S., & Johansen, K. (2014). Characterizing the spatial structure of mangrove features for optimizing image-based mangrove mapping. *Remote Sensing, 6*(2), 984-1006.

Kamal, M., Phinn, S., & Johansen, K. (2015). Object-based approach for multi-scale mangrove composition mapping using multi-resolution image datasets. *Remote Sensing, 7*(4), 4753-4783.

Kamaruzaman, J., & Kasawani, I. (2007). Imaging spectrometry on mangrove species identification and mapping in Malaysia. *WSEAS Trans Biol Biomed, 8*, 118-126.

Kanellopoulos, I., & Wilkinson, G. G. (1997). Strategies and best practice for neural network image classification. *International Journal of Remote Sensing, 18*(4), 711-725.

Kanniah, K. D., Sheikhi, A., Cracknell, A. P., Goh, H. C., Tan, K. P., Ho, C. S., & Rasli, F. N. (2015). Satellite images for monitoring mangrove cover changes in a fast-growing economic region in southern Peninsular Malaysia. *Remote Sensing, 7*(11), 14360-14385.

Kathiresan, K., & Bingham, B. L. (2001). Biology of mangroves and mangrove ecosystems. *Advances in Marine Biology, 40*, 81-251.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research: Atmospheres, 102*(D14), 17051-17067.

Kirui, K. B., Kairo, J. G., Bosire, J., Viergever, K. M., Rudra, S., Huxham, M., & Briers, R. A. (2013). Mapping of mangrove forest land cover change along the Kenya coastline using Landsat imagery. *Ocean and Coastal Management, 83*, 19–24. https://doi.org/10.1016/j.ocecoaman.2011.12.004

Knight, J. M., Dale, P. E., Spencer, J., & Griffin, L. (2009). Exploring LiDAR data for mapping the micro-topography and tidal hydro-dynamics of mangrove systems: An example from southeast Queensland, Australia. *Estuarine, Coastal and Shelf Science, 85*(4), 593-600.

Kokaly, R. F., Despain, D. G., Clark, R. N., & Livo, K. E. (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment, 84*(3), 437-456.

Kovacs, J. M., Liu, Y., Zhang, C., Flores-Verdugo, F., & de Santiago, F. F. (2011). A field based statistical approach for validating a remotely sensed mangrove forest classification scheme. *Wetlands Ecology and Management, 19*(5), 409-421.

Kovacs, J. M., Wang, J., & Flores-Verdugo, F. (2005). Mapping mangrove leaf area index at the species level using IKONOS and LAI-2000 sensors for the Agua Brava Lagoon, Mexican Pacific. *Estuarine, Coastal and Shelf Science, 62*(1), 377-384.

Krause, G., Bock, M., Weiers, S., & Braun, G. (2004). Mapping land-cover and mangrove structures with remote sensing techniques: A contribution to a synoptic GIS in support of coastal management in North Brazil. *Environmental Management, 34*(3), 429-440.

Kridiborworn, P., Chidthaisong, A., Yuttitham, M., & Tripetchkul, S. (2012). Carbon sequestration by mangrove forest planted specifically for charcoal production in Yeesarn, Samut Songkram. *Journal of Sustainable Energy and Environment, 3*(2), 87-92.

Kruse, F. A., Boardman, J. W., & Huntington, J. F. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing, 41*(6), 1388-1400.

Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T. V., & Dech, S. (2011). Remote sensing of mangrove ecosystems: A review. *Remote Sensing, 3*(5), 878-928.

Kux, H. J., & Souza, U. D. (2012). Object-based image analysis of WORLDVIEW-2 satellite data for the classification of mangrove areas in the city of São Luís, Maranhão State, Brazil. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 95-100). Melbourne, Australia.

Lee, T. M., & Yeh, H. C. (2009). Applying remote sensing techniques to monitor shifting wetland vegetation: A case study of Danshui River estuary mangrove communities, Taiwan. *Ecological Engineering, 35*(4), 487-496.

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography, 11*(5), 393-399.

Leprince, S., Barbot, S., Ayoub, F., & Avouac, J. P. (2007). Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing, 45*(6), 1529-1558.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). *Remote sensing and image interpretation.* New York, NY: John Wiley & Sons.

Liu, K., Li, X., Shi, X., & Wang, S. (2008). Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning. *Wetlands, 28*(2), 336-346.

Long, J. B., & Giri, C. (2011). Mapping the Philippines' mangrove forests using Landsat imagery. *Sensors, 11*(3), 2972-2981.

Lugo, A. E., & Snedaker, S. C. (1974). The ecology of mangroves. *Annual Review of Ecology and Systematics, 5*(1), 39-64.

Manson, F. J., Loneragan, N. R., Harch, B. D., Skilleter, G. A., & Williams, L. (2005). A broad-scale analysis of links between coastal fisheries production and mangrove extent: A case-study for northeastern Australia. *Fisheries Research, 74*(1), 69-85.

Manson, F. J., Loneragan, N. R., McLeod, I. M., & Kenyon, R. A. (2001). Assessing techniques for estimating the extent of mangroves: topographic maps, aerial photographs and Landsat TM images. *Marine and Freshwater Research, 52*(5), 787-792.

Manson, F. J., Loneragan, N. R., Skilleter, G. A., & Phinn, S. R. (2005). An evaluation of the evidence for linkages between mangroves and fisheries: a synthesis of the literature and identification of research directions. In *Oceanography and marine biology* (pp. 493-524). USA: CRC Press.

Mantri, V. A., & Mishra, A. K. (2006). On monitoring mangrove vegetation of Sagar Island by remote sensing. *National Academy Science Letters, 29*(1/2), 45-48.

Martin, M. E., & Aber, J. D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications, 7*(2), 431-443.

Martinuzzi, S., Gould, W. A., Lugo, A. E., & Medina, E. (2009). Conversion and recovery of Puerto Rican mangroves: 200 years of change. *Forest Ecology and Management, 257*(1), 75-84.

Matsui, N., Meepol, W., & Chukwamdee, J. (2015). Soil Organic Carbon in Mangrove Ecosystems with Different Vegetation and Sedimentological Conditions. *Journal of Marine Science and Engineering, 3*(4), 1404-1424.

Mesta, P. N., Setturu, B., Subash Chandran, M. D., Rajan, K. S., & Ramachandra, T. V. (2014). Inventorying, mapping and monitoring of mangroves towards sustainable management of West Coast, India. *J Geophysics Remote Sensing, 3*(3), 130-138.

Misra, A., Murali, M. R. & Vethamony, P. (2015). Assessment of the land use/land cover (LU/LC) and mangrove changes along the Mandovi–Zuari estuarine complex of Goa, India. *Arabian Journal of Geosciences, 8*(1), 267-279.

Mumby, P. J., Green, E. P., Edwards, A. J., & Clark, C. D. (1999). The cost-effectiveness of remote sensing for tropical coastal resources assessment and management. *Journal of Environmental Management, 55*(3), 157-166.

Murray, M. R., Zisman, S. A., Furley, P. A., Munro, D. M., Gibson, J., Ratter, J., ... & Place, C. J. (2003). The mangroves of Belize: Part 1. distribution, composition and classification. *Forest Ecology and Management, 174*(1), 265-279.

Muttitanon, W., & Tripathi, N. K. (2005). Land use/land cover changes in the coastal zone of Ban Don Bay, Thailand using Landsat 5 TM data. *International Journal of Remote Sensing, 26*(11), 2311-2323.

Mäkynen, J., Holmlund, C., Saari, H., Ojala, K., & Antila, T. (2011). Unmanned aerial vehicle (UAV) operated megapixel spectral camera. In *Proceedings of SPIE, Electro-Optical Remote Sensing, Photonic Technologies, and Applications* (Vol. 8186, p. 81860Y). International Society for Optics and Photonics.

Nayak, S., & Bahuguna, A. (2001). Application of remote sensing data to monitor mangroves and other coastal vegetation of India. *Indian Journal of Marine Sciences, 30*(4), 195-213.

Neukermans, G., Dahdouh-Guebas, F. J. G. K., Kairo, J. G., & Koedam, N. (2008). Mangrove species and stand mapping in Gazi Bay (Kenya) using Quickbird satellite imagery. *Journal of Spatial Science, 53*(1), 75-86.

Newton, A. C., Hill, R. A., Echeverría, C., Golicher, D., Rey Benayas, J. M., Cayuela, L., & Hinsley, S. A. (2009). Remote sensing and the future of landscape ecology. *Progress in Physical Geography, 33*(4), 528-546.

Nguyen, H., Mcalpine, C., Pullar, D., Johansen, K., & Duke, N. C. (2013). Ocean and coastal management the relationship of spatial e temporal changes in fringe mangrove extent and adjacent land-use: Case study of Kien Giang coast, Vietnam. *Ocean and Coastal Management, 76*, 12–22. https://doi.org/10.1016/j.ocecoaman.2013.01.003

Nguyen, H. H. (2014). The relation of coastal mangrove changes and adjacent land-use: A review in Southeast Asia and Kien Giang, Vietnam. *Ocean and Coastal Management, 90*, 1-10.

Osti, R., Tanaka, S., & Tokioka, T. (2009). The importance of mangrove forest in tsunami disaster mitigation. *Disasters, 3*3(2), 203-213.

Ozesmi, S. L., & Bauer, M. E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management, 10*(5), 381-402.

Paling, E. I., Kobryn, H. T., & Humphreys, G. (2008). Assessing the extent of mangrove change caused by Cyclone Vance in the eastern Exmouth Gulf, northwestern Australia. *Estuarine, Coastal and Shelf Science, 77*(4), 603-613.

Pasqualini, V., Iltis, J., Dessay, N., Lointier, M., Guelorget, O., & Polidori, L. (1999). Mangrove mapping in North-Western Madagascar using SPOT-XS and SIR-C radar data. *Hydrobiologia, 413*(0), 127-133.

Patenaude, G., Hill, R. A., Milne, R., Gaveau, D. L., Briggs, B. B. J., & Dawson, T. P. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment, 93*(3), 368-380.

Pattanaik, C., Reddy, C. S., & Prasad, S. N. (2008). Mapping, monitoring and conservation of Mahanandi wetland ecosystem, Orissa, India using remote sensing and GIS. In *Proceedings of the National Academy of Sciences India Section B-Biological Science*s (vol. 78, pp. 81-89). India.

Pearce, F. (2014). Power to the people. *New Scientist, 223*(2980), 26-27.

Pohl, C., & Van Genderen, J. L. (1998). Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing, 19*(5), 823-854.

Prasad, K. A., & Gnanappazham, L. (2014). Species discrimination of mangroves using Derivative Spectral Analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2*(8), 45-52.

Price, J. C. (1992). Variability of high-resolution crop reflectance spectra. *International Journal of Remote Sensing, 13*(14), 2593-2610.

Proisy, C., Couteron, P., & Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote Sensing of Environment, 109*(3), 379-392.

Proisy, C., Gratiot, N., Anthony, E. J., Gardel, A., Fromard, F., & Heuret, P. (2009). Mud bank colonization by opportunistic mangroves: A case study from French Guiana using lidar data. *Continental Shelf Research, 29*(3), 632-641.

Ramachandran, S., Sundaramoorthy, S., Krishnamoorthy, R., Devasenapathy, J., & Thanikachalam, M. (1998). Application of remote sensing and GIS to coastal wetland ecology of Tamil Nadu and Andaman and Nicobar group of islands with special reference to mangroves. *Current Science, 75*(3), 236-244.

Ranchin, T., & Wald, L. (2000). Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing, 66*(1), 49-61.

Rao, K. R., & Yip, P. (2014). *Discrete cosine transform: algorithms, advantages, applications.* United States of America, USA: Academic press.

Rasolofoharinoro, M., Blasco, F., Bellan, M. F., Aizpuru, M., Gauquelin, T., & Denis, J. (1998). A remote sensing based methodology for mangrove studies in Madagascar. *International Journal of Remote Sensing, 19*(10), 1873-1886.

Reddy, C. S., & Pattanaik, C. (2007). Mangrove vegetation assessment and monitoring in Balasore district, Orissa using remote sensing and GIS. *National Academy Science Letters, 30*(11/12), 377-382.

Rhyma, P. P., & Norizah, K. (2016). Kriging analysis- optimizing values in unknown areas using known data point. In *Proceedings of International Conference on Sustainable Forest Development in view of Climate Change (SFDCC2016),* (pp. 96-99). Hotel Bangi-Putrajaya, Malaysia.

Rhyma, P. P., Norizah, K., Ismail Adnan, A. M., Faridah-hanum, I., & Ibrahim, S. (2015). Canopy density classification of Matang mangrove forest reserve using machine learning approach in remote sensing. *The Malaysian Forester, 78*(1&2), 75-86.

Rodriguez, W., & Feller, I. C. (2004). *Mangrove landscape characterization and change in Twin Cays, Belize using aerial photography and IKONOS satellite data.* National Museum of Natural History.

Rosette, J. A. B., North, P. R. J., & Suarez, J. C. (2008). Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing, 29*(5), 1475-1493.

Roslani, M. A., Mustapha, M. A., Lihan, T., & Juliana, W. W. (2013, November). Classification of mangroves vegetation species using texture analysis on Rapideye satellite imagery. In *AIP Conference Proceedings* (Vol. 1571, No. 1, pp. 480-486). AIP Publishing, United States of America.

Roslani, M. A., Mustapha, M. A., Lihan, T., & Juliana, W. W. (2014). Applicability of Rapideye satellite imagery in mapping mangrove vegetation species at Matang mangrove forest reserve, Perak, Malaysia. *Journal of Environmental Science and Technology, 7*(2), 123-136.

Ruiz-Luna, A., & Berlanga-Robles, C. A. (2003). Land use, land cover changes and coastal lagoon surface reduction associated with urban growth in northwest Mexico. *Landscape Ecology, 18*(2), 159-171.

Rönnbäck, P. (1999). The ecological basis for economic value of seafood production supported by mangrove ecosystems. *Ecological Economics, 29*(2), 235-252.

Saari, H., Pellikka, I., Pesonen, L., Tuominen, S., Heikkilä, J., Holmlund, C., ... & Antila, T. (2011). Unmanned Aerial Vehicle (UAV) operated spectral camera system for forest and agriculture applications. In *Proceedings of the SPIE, Remote Sensing for Agriculture, Ecosystems, and Hydrology XIII.* (Vol. 8174, p. 81740H). International Society for Optics and Photonics.

Saito, H., Bellan, M., Al-Habshi, A., Aizpuru, M., & Blasco, F. (2003). Mangrove research and coastal ecosystem studies with SPOT-4 HRVIR and TERRA ASTER in the Arabian Gulf. *International Journal of Remote Sensing, 24*(21), 4073-4092.

Santos, L. C. M., Matos, H. R., Schaeffer-Novelli, Y., Cunha-Lignon, M., Bitencourt, M. D., Koedam, N., & Dahdouh-Guebas, F. (2014). Anthropogenic activities on mangrove areas (São Francisco River Estuary, Brazil Northeast): A GIS-based analysis of CBERS and SPOT images to aid in local management. *Ocean and Coastal Management, 89*, 39-50.

Sawaya, K. E., Olmanson, L. G., Heinert, N. J., Brezonik, P. L., & Bauer, M. E. (2003). Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery. *Remote Sensing of Environment, 88*(1), 144-156.

Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing.* USA: Academic press.

Setlur, V., Gleicher, M., Gooch, B., Takagi, S., & Raskar, R. (2009). *U.S. Patent No. 7,574,069.* Washington, DC: U.S. Patent and Trademark Office.

Son, N. T., Chen, C. F., & Chen, C. R. (2017). Mapping mangrove density from RapidEye Data in Central America. *Open Geosciences, 9*(1), 211-220.

Souza-Filho, P. W., & Paradella, W. R. (2003). Use of synthetic aperture radar for recognition of Coastal Geomorphological Features, land-use assessment and shoreline changes in Bragança coast, Pará, Northern Brazil. *Anais da Academia Brasileira de Ciências, 75*(3), 341-356.

Spalding, M. (2010). *World atlas of mangroves.* London: Routledge.

Spalding, M., Blasco, F., & Field, C. (1997). *World mangrove atlas.* Okinawa (Japan): International Society for Mangrove Ecosystems.

Starek, J. A., Açıkmeşe, B., Nesnas, I. A., & Pavone, M. (2016). Spacecraft autonomy challenges for next-generation space missions. In *Advances in Control System Technology for Aerospace Applications* (pp. 1-48). Springer Berlin Heidelberg.

Sulong, I. (1999). Remote sensing and geographic information system (GIS) application to environmental sensitivity index mapping. In *Environmental Sensitivity Index Mapping of the Chukai to Penor Coastline* (pp. 29-39). Terengganu, Malaysia.

Sulong, I., & Veddin, I. (1999). *Mapping in the coastal area of Sabak Bernam, Kuala Selangor and Klang district with emphasis on mangrove forest using remote sensing and geographical information systems'*. Assessment and Monitoring of Marine Systems, Universiti Putra Malaysia Terengganu (in press). Google Scholar.

Sulong, I., Mohd-Lokman, H., Mohd-Tarmizi, K., & Ismail, A. (2002). Mangrove mapping using Landsat imagery and aerial photographs: Kemaman District, Terengganu, Malaysia. *Environment, Development and Sustainability, 4*(2), 135-152.

Tarabalka, Y., Fauvel, M., Chanussot, J., & Benediktsson, J. A. (2010). SVM-and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters, 7*(4), 736-740.

Terchunian, A., Klemas, V., Segovia, A., Alvarez, A., Vasconez, B., & Guerrero, L. (1986). Mangrove mapping in Ecuador: the impact of shrimp pond construction. *Environmental Management, 10*(3), 345-350.

Thampanya, U., Vermaat, J. E., Sinsakul, S., & Panapitukkul, N. (2006). Coastal erosion and mangrove progradation of Southern Thailand. *Estuarine, Coastal and Shelf Science, 68*(1), 75-85.

Thu, P. M., & Populus, J. (2007). Status and changes of mangrove forest in Mekong Delta: Case study in Tra Vinh, Vietnam. *Estuarine, Coastal and Shelf Science, 71*(1), 98-109.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution, 18*(6), 306-314.

Tyo, J. S., Goldstein, D. L., Chenault, D. B., & Shaw, J. A. (2006). Review of passive imaging polarimetry for remote sensing applications. *Applied Optics, 45*(22), 5453-5469.

Vaiphasa, C., Skidmore, A. K., & de Boer, W. F. (2006). A post-classifier for mangrove mapping using ecological data. *ISPRS Journal of Photogrammetry and Remote Sensing, 61*(1), 1-10.

Van Aardt, J. A. N., & Wynne, R. H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing, 28*(2), 431-436.

Vasconcelos, M. J. P., Biai, J. M., Araujo, A., & Diniz, M. A. (2002). Land cover change in two protected areas of Guinea-Bissau (1956-1998). *Applied Geography, 22*(2), 139-156.

Verma, M., Fisher, J. B., Mallick, K., Ryu, Y., Kobayashi, H., Guillaume, A., ... & Sikka, M. (2016). Global surface net-radiation at 5 km from MODIS Terra. *Remote Sensing, 8*(9), 739-758.

Vo, Q. T., Oppelt, N., Leinenkugel, P., & Kuenzer, C. (2013). Remote sensing in mapping mangrove ecosystems—an object-based approach. *Remote Sensing, 5*(1), 183-201.

Walters, B. B. (2005). Patterns of local wood use and cutting of Philippine mangrove forests. *Economic Botany, 59*(1), 66-76.

Wang, L., Sousa, W. P., & Gong, P. (2004). Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing, 25*(24), 5655-5668.

Wang, L., Sousa, W. P., Gong, P., & Biging, G. S. (2004). Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sensing of Environment, 91*(3), 432-440.

Wang, L. E., & Sousa, W. P. (2009). Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance. *International Journal of Remote Sensing, 30*(5), 1267-1281.

Wang, L. E., & Sousa, W. P. (2009). Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance. *International Journal of Remote Sensing, 3*0(5), 1267-1281. https://doi.org/10.1080/01431160802474014

Wannasiri, W., Nagai, M., Honda, K., Santitamnont, P., & Miphokasap, P. (2013). Extraction of mangrove biophysical parameters using airborne LiDAR. *Remote Sensing, 5*(4), 1787-1808.

Wei, G., Shalei, S., Bo, Z., Shuo, S., Faquan, L., & Xuewu, C. (2012). Multi-wavelength canopy LiDAR for remote sensing of vegetation: Design and system performance. *ISPRS Journal of Photogrammetry and Remote Sensing, 69*, 1-9.

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: A review. *Journal of Plant Ecology, 1*(1), 9-23.

Zhang, C., Kovacs, J. M., Liu, Y., Flores-Verdugo, F., & Flores-de-Santiago, F. (2014). Separating mangrove species and conditions using laboratory hyperspectral data: A case study of a degraded mangrove forest of the Mexican Pacific. *Remote Sensing, 6*(12), 11673–11688. https://doi.org/10.3390/rs61211673

Zhang, K. (2008). Identification of gaps in mangrove forests with airborne LIDAR. *Remote Sensing of Environment, 112*(5), 2309-2325.

Zhang, R., & Zhu, D. (2011). Study of land cover classification based on knowledge rules using high-resolution remote sensing images. *Expert Systems with Applications, 38*(4), 3647-3652.

Zhu, Y., Liu, K., Liu, L., Wang, S., & Liu, H. (2015). Retrieval of mangrove aboveground biomass at the individual species level with WorldView-2 images. *Remote Sensing, 7*(9), 12192-12214.

Zink, F., Vincent, R. A., Murphy, E., & Cote, O. (2004). Comparison of radar and in situ measurements of atmospheric turbulence. *Journal of Geophysical Research: Atmospheres, 109*(D11), 1-7.

Zulfa, A. W., & Norizah, K. (2016). Examining the radiation of electromagnetic reflectance of satellite image in identifying object on earth. In *Proceedings of International Conference on Sustainable Forest Development in view of Climate Change (SFDCC2016)*, (pp. 100-104). Hotel Bangi-Putrajaya, Malaysia. 8-11 August 2016.

Review Article

A Preliminary Study on Paper Sheets Based Epoxy Composites Designed for Repairing Work Application and Its Properties – A Review

Muhamad Hellmy Hussin

Fabrication and Joining Section, Universiti Kuala Lumpur Malaysia France Institute, Section 14, Jalan Teras Jernang, 43650 Bandar Baru Bangi, Selangor, Malaysia

ABSTRACT

This is a review of studies on various types of paper-based epoxy composites currently being designed and developed for technological use. The concept of designing composite materials is very significant for small to large industry and it is important where initiation of repairing work is now being considered for engineering applications. This composite material is of interest due to its advantages compared with others, including low environmental effects and low cost for a wide range of works. This review aims to provide an overview of morphological, physical and mechanical properties of various paper sheets-based epoxy composites and details of achievements made. From this approach, this paper also presents the preliminary study of SEM results of paper sheets-based epoxy composites designed for repairing work applications. It has been found that a well-arranged laminated paper sheet layers could help the bond strength with epoxy matrix. Thus, this paper sheet-based epoxy composite can be considered as an easiest way, cheap and biodegradable that can be used for various small repairing works in structural and automotive applications.

Keywords: Composites, mechanical properties, paper sheets, physical properties, SEM-EDX

INTRODUCTION

Papers are substances made from natural plant fibres called cellulose and they are originally derived from cloth rags and grasses. Nowadays, papers are made from wood (Pal & Joyce, 2017). Cellulose is an organic compound and is an important structural component of the primary cell wall of green

plants, and the cellulose content of wood is approximately 40 – 50% by volume (Bajpai, 2012). Basically, wood pulp is a lignocellulosic fibrous material prepared by chemical or mechanical means by separating cellulose fibres from wood, fibre crops or waste paper. Bleaching is a chemical process where various wood pulp is added so the pulp becomes white which is an important characteristic of paper. Additionally, the demand and consumption of paper in daily life are still very high and increases every year and this leads to waste problem. In this study, paper sheets refer to waste paper sheets which are basically useless papers, and reject paper from printing works, department stores, self-service stores and homes (Moral et al., 2014; Shafiur Rahman et al., 2014).

Basically, a composite material is defined as a combination of two or more materials that have better properties than when individual components are used alone (Madsen & Gamstedt, 2013; Oladele & Okoro, 2015; Radif, Ali, & Abdan, 2011). This material is designed to display a combination of the best properties of each of the component materials and usually strongly depends on the properties of their constituent materials, their distributions, and the interactions among them (Oladele & Okoro, 2015; Qiu, Liu, & Li, 2015). In this work, composite used consisted of at least two materials, which is paper and the polymeric thermoset matrix (Kroling et al., 2014). In addition, the task of fibres in composite materials is to take tensile and flexural loads, while the matrix material is keeping the fibres in their place and act to transfer the loads between the fibres. Wood pulp fibres in composite materials has gained major interest because of their potential in increasing the mechanical properties of some material and also act as reinforcement in hydrophilic and hydrophobic matrices (Chinga-Carrasco et al., 2011). Most of the engineering properties are usually considered in material selection and design of the structure. It depends on the application and obviously, it is not only on the fibre length or its distribution. It also depends on the orientation and quantity of fibre which are of great importance in composite materials structure.

Instead of using wood pulp in paper making process, non-wood fibre represents a substantial raw material which is widely cultivated throughout the world and used as major oilseed in biodiesel production. Replacing wood pulp with agro residues has less adverse impact on the environment, aside being more economical and easily accessible (Mohd Kassim et al., 2016). They found that together with the cost-effectiveness and less abundance of these crop residues, it was important to consider its use in the pulp and papermaking production (Kiaei et al., 2014). The authors reported that instead of non-fibres wood, some agricultural residues such as bagasse, wheat and rice straws, sorghum stalks and hemp were widely used as raw materials for pulp making process.

Physical Properties

According to Han et al. (2011) when adding the composites additives, little change occured to the physical properties such as the weight, the thickness and the opacity of the composites materials. Instead of the important role played by fibres, resin type used have gained wide acceptance in terms of surface quality. Istek et al. (2010) stated that it was possible to impregnate

the paper with appropriate synthetic resins which include urea formaldehyde, melamine formaldehyde, acrylic, phenolic resins, and mixtures thereof to ensure the paper bond providing a resin-rich finish on the surface under heat and pressure. Thus, it was found that the type of resin influences the quality of the paper composite materials.

Kroling et al. (2014) reported the differences in the apparent density which reflected in the fibre volume fractions of the composite. They found that the specialty paper and lab sheets achieve a fibre volume fraction of almost 40%. Meanwhile low density tea-bag papers and spun laces achieve volume fraction of below 20%. It is important to note that a high volume fraction is beneficial because the mechanical properties of the composite are directly linked to the fibre volume fraction. Praharaj et al. (2014) concluded that the composites with less content of filler absorb less moisture as compared to the composites with high filler content. It was found that the reason was because of the phase separation of matrix and the dispersed phase that took place at higher filler content. Habibi et al. (2016) showed the three important parameters for manufacturing the paper layer, namely surface density, flax content and flax fibre length, influence two of its important properties, which are paper structure and permeability. They confirmed that the reinforcement's permeability was influenced by the paper layer structure and the differences in permeability values were obtained by varying paper layer surface density, flax content and its fibre length.

Factor Affecting Mechanical Performance

As described by Khalilitabas et al. (2009), the application of pulp fibre in cement paste was shown to improve the bearing capacities of the cement composites. This behaviour depends on four main factors which are fibre types, mixture percentage, fabrication manner and additives. Instead of that, the fibres orientation is important to ensure that it disperses uniformly throughout the sample that leads to a good bond development between fibres and cement paste (Sangrutsamee, Srichandr, & Poolthong, 2012). They also proved that the main reason for decrease of strength of cement composites were (i) the thickening of the samples because of porosity and (ii) non uniform distribution of fibre in cement paste which the fibres orientation prone to twist inside the matrix (Khalilitabas et al., 2009). Chinga-Carrasco et al. (2011) stated that structural composites for load-carrying application are based on long fibres acting as reinforcement, whereas high-volume composites with particle reinforcement are preferred to reduce costs by using cheaper filler. It can be explained that the fibre length distribution is one of the practical interests in this issue. It was also reported by Suriani et al., (2013) that fatigue life and analyses on natural fibre reinforced composite provided better picture in predicting the life span of the materials. It was found that in composites, fatigue damage and failure mechanism were more complex compared to metal or steel.

Moreover, fibre length is also known to affect several important engineering properties of composites such as stiffness, tensile strength, fracture toughness and dimensional stability on moisture uptake. Kroling et al. (2014) found that the short fibre length in papers should reduce the composite strength. However, they also found that the actual single fibre strength, the shear

strength of the matrix-fibre interface and fibre orientation influence the tensile strength of the composites. The authors clearly showed that paper yields better composite properties than commercially available natural fibre reinforcements (Kroling et al., 2014; Nayak & Mishra, 2013). Instead of fibre dimensions, chemical components also play an important role in the mechanical properties of paper pulp and wood fibre-based composites. It shows that lignin which is one of the main components of wood cell walls has critical effect on mechanical properties of composites. It is found that the amount of lignin affects the tensile strength and elongation of single fibres (Zhang, Fei, Yu, Cheng, & Wang, 2013).

Basically, on the reinforcement part, the stiffness is relatively higher for nanofibrils compared with wood fibres, because of its orientation and distribution. Reinforcing cellulose is aligned to the direction of the nanofibrils, whereas wood fibres microfibrils cellulose have chiral orientation in the cell wall with an inclination angle (Chinga-Carrasco et al., 2011). Han et al. (2011) found that all physical properties of the composite paper made from the 30% by volume of slag wool fibres and 70% by volume of paper pulp and without any composite additives decreased sharply in comparison with common paper. They compared the same paper with the addition of composite additives, the tensile strength increased by 30% while the folding endurance by 40%. Thus, they proved that it was important to introduce composite additives in the process of composite papermaking. Praharaj et al. (2014) confirmed that the tensile strength of the fabricated composites gradually increased with the addition of filler (paper pulp), attained a maximum value at 40% of filler and then decreased. They found that as the filler content in the composite increased, the matrix becomes more and more dense which leads strengthens the composites, thus, enabling it to withstand stress efficiently. They also stated that at higher concentration, the filler was no longer able to mix thoroughly with the matrix which caused phase separation of the continuous and dispersed phase. Thus, this affects 40% reduction of toughness of the composites (Praharaj et al., 2014).

Moreover, by using specific resin type in paper composite materials, the surface coating quality is found to improve the bending strength, modulus of elasticity, and thickness swelling depending on the type and thickness of the coatings (Istek et al., 2010; Robillard & Lebrun, 2010). Istek et al. (2010) reported that phenolic-impregnated paper overlays resisted weathering better than do overlays impregnated with urea or melamine. They proved that after lamination process, the mechanical properties of the particleboards paper laminated with impregnated décor papers were increased. It showed that the highest bending strength was 16.61 N/mm^2 in wrenge-Urea-Formaldehyde as a resin. The bending strength of laminated particleboards paper ranged from 13.58 to 16.61 N/mm^2. Meanwhile, Ren et al. (2015) reported that the composites reinforced with pulps and using biodegradable polymers which was polyhydroxybutyrate (PHB) showed significant increase in tensile stiffness and impacted strength and that the addition of a coupling agent did not improve the mechanical performance of PHB biocomposites.

The defects which originated on the surface or sub-surface of the composite or at the composite interface show some significant effects on the mechanical performance of the materials. It was reported by Praharaj et al. (2014) that because of the phase separation in the

matrix composites, it resulted in the formation of the rough topography in the form of voids on the composite surface. They confirmed that the void led to easy water diffusion inside the substrate and thus weakening the interfacial bonding between BisGMA (Bisphenol-A glycidyldimethacrylate) and paper pulp. They also found that the high water uptake also weakened the composites mechanically. Meanwhile as observed by Ren et al. (2015), degradation occured on PLA (polylactic acid) and PHB (polyhydroxybutyrate) biocomposites during heating process was caused by the release of organic acid compounds such as crotonic and lactic acids which enhanced the existing interface between the pulps and the matrix polymers.

Ng et al. (2017) found that by using the mixture of natural fibre with synthetic fibre in composite structure, it could reduce the manufacturing cost as well as being safe for the environment.

Morphological Images

Khalilitabas et al. (2009) showed the microstructure of pulp-fibres represented the rough surface with good fibrillation (Figure 1). These well-arranged fibres form the numerous fibrils around the outer surface that could help the friction bond strength with cement matrix. They also proved the insertion of tiny particles which were associated with the fibres could be interpreted as defects that decreased the strength of the composite. Chinga-Carrasco et al. (2011) proved that the wall structure of cellulose fibres in wood pulp was mainly composed of microfibrils that were arranged differently in the various layers of a fibre wall structure in the nanometre-scale as shown in Figure 2.
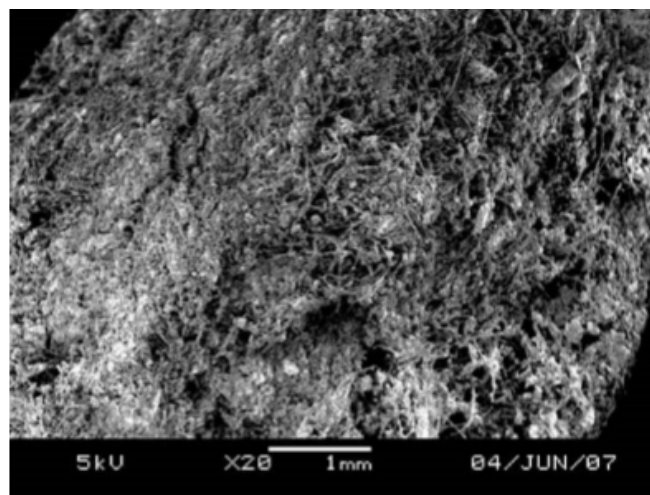
Figure 1. SEM image of cement paste with Kraft fibre (Khalilitabas et al., 2009)
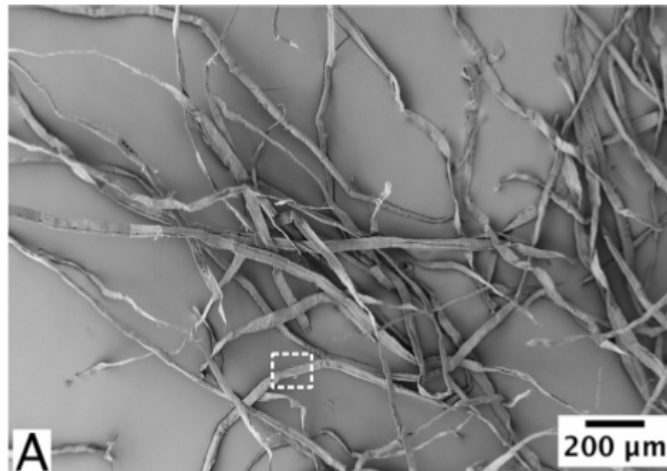
Figure 2. FESEM image of the Kraft pulp fibres structure exemplifying their high aspect ratio (Chinga-Carrasco et al., 2011)

MATERIALS AND METHODS

White paper sheets are cut and layered in specific dimension according to the size of the crack as shown in Figure 3 and Figure 4. Type of matrix used is epoxy resin. The lamination operation is carried out manually using hand lay-up process at ambient temperature. The epoxy resin is applied to a single paper sheet with a brush. After that, the next layer of paper sheet was laid upon the already impregnated sheet and then it is manually pressed onto it. The new sheet is
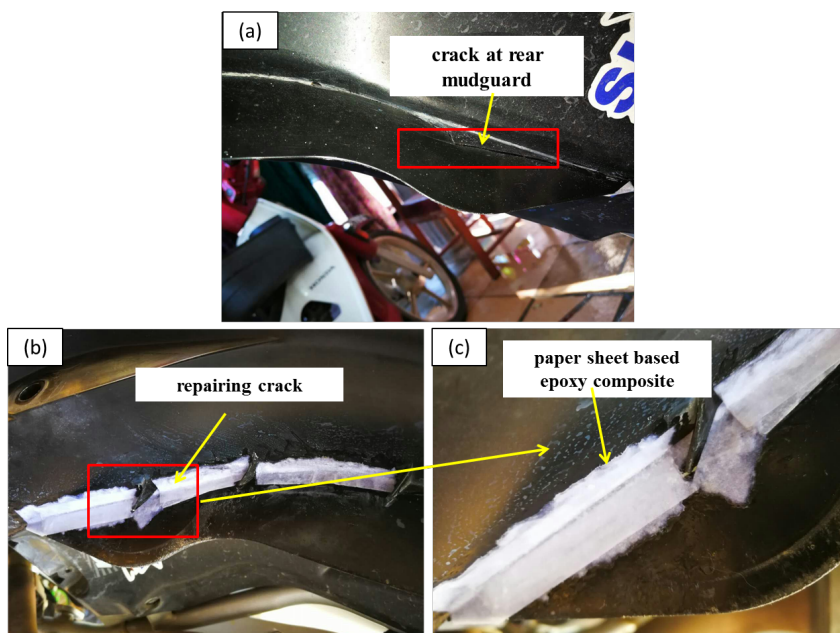
Figure 3. Repairing crack at motorbike rear mudguard, (a) crack location, (b) repairing crack using paper sheet-based epoxy composite, and (c) enlargement of repairing crack location in (b)

then also impregnated with the epoxy resin as described earlier. This procedure is repeated until three layers of both paper sheets and epoxy resins are laid. Samples are prepared for small repairing work at two locations which are (i) motorbike rear mudguard and (ii) battery container as shown in Figure 3 and Figure 4. Samples are then subjected to morphological analysis using scanning electron microscope (SEM).
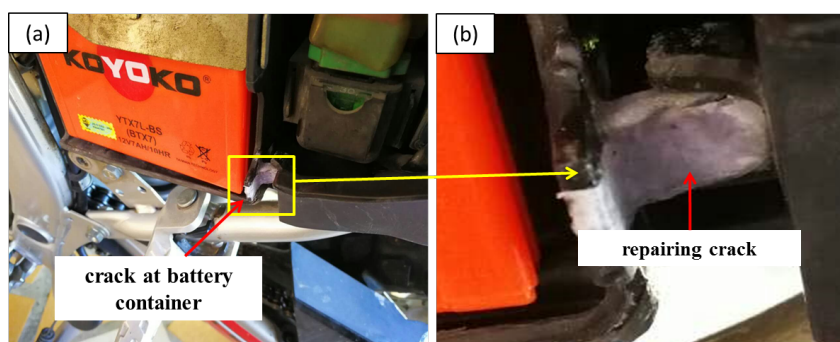
Figure 4. Repairing crack at motorbike battery container, (a) crack location, and (b) enlargement of repairing crack location in (b)

RESULTS AND DISCUSSION

Paper sheet is a complex material built from chemical and mechanical interactions between a large amount of single fibres, then forming a continuous fibre network arrangement (Habibi et al., 2016). The thickness of paper sheets is a result of the continuous deposition of individual fibres. As shown in Figure 5, the strength of the paper composite depends on its position, dimensions, and flexibility. Several images of the paper layer in Figure 5(b-d) taken with SEM, all at level of magnification of 640 x. It shows that the surface density is clearly dominated by paper sheets thickness layered with epoxy resin. In particular, this nature of layered process, lead to increase in paper sheet thickness.

In this work, to verify the behavior of paper sheet epoxy-based composite, SEM images are analysed and the morphologies of SEM images are shown in Figure 5. As shown in Figure 5(b), (c) and (d), the layers of epoxy and paper sheet arranged without any presence of impurities. The microstructure of paper sheet composite represents the rough surface with significant layer sequence. Well-arranged laminated layers could help the bond strength with epoxy matrix. Obviously, although the method applied for the composite structure is manual using conventional hand lay-up technique, this method can still promote the strength for the purpose of small repairing method. In other words, by adding and combining the paper sheet with epoxy, the paper strength has obviously improved and can be used to aid in small repair work purposes as shown in Figure 3 and Figure 4.
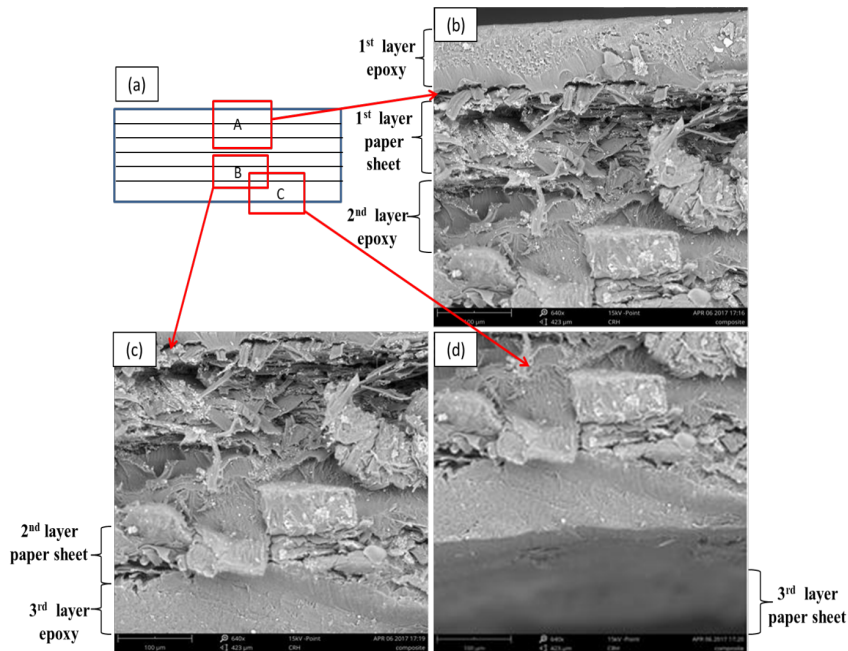
Figure 5. SEM images of cross sectional view of paper sheet epoxy composite sample, (a) point location where the SEM images are taken, (b) SEM image of epoxy and paper sheet layers taken at point A, (b) SEM image of epoxy and paper sheet layers taken at point B, and (c) SEM image of epoxy and paper sheet layers taken at point C

Moreover, paper has been utilised in everyday life for more than 20 decades and known to have good properties such as surface roughness, porous structure and optical opaqueness. As compared to other methods, it shows poor dimensional stability and bad water resistance which is due to the hydrophilic nature of the cellulose they are made of. Therefore, it is necessary to overcome these problems and produce new materials and increase the mechanical properties of regular paper or waste paper sheets by reinforcing it with other materials. As reported by Oladele and Okoro (2015), regular paper or white paper sheet was made of cellulose with diameter in the range of 20-50 mm. Their large surface roughness, porous structure and optical opaqueness act as an intrinsic barrier to hosting electronic devices on the surface of this material.

In addition, by using white paper sheets or recycled paper sheet waste materials which are increasing in our day by day usage, and because of our responsibility, it is important to take proper steps to remove this waste and safeguard our environment. Recycling a paper sheet waste and changing it (waste) into new products to prevent waste of potentially useful materials would reduce the consumption of fresh raw materials. Thus, this process would reduce energy usage, reduce air and water pollution, and lower greenhouse gas emissions (Shafiur Rahman et al., 2014).

CONCLUSION

Successfully fabricated paper sheet-epoxy-based composite made from waste paper sheet has been manually produced by simple hand lay-up technique. A cross-sectional analysis of the sample is evaluated and it shows a well-arranged laminated paper sheet layers could help to improve bond strength with epoxy matrix. Although the method applied for the composite structure is manual using conventional method, this still can promote strength by adding and combining paper sheet with epoxy. It can be stated that the fabrication of paper sheet epoxy-based composite is suitable for small repairing work purposes and it is found to be significantly responsible for crack repair. Thus, this paper sheet- based epoxy composite can be considered as easiest way and cheap that can be used for various small repair works in structural and automotive applications.

REFERENCES

Bajpai, P. (2012). Brief description of the pulp and paper making process. In *Biotechnology for Pulp and Paper Processing.* Springer Science.

Chinga-Carrasco, G., Miettinen, A., Luengo Hendriks, C. L., Gamstedt, E. K., & Kataj, M. (2011). Structural characterisation of Kraft pulp fibres and their nanofibrillated materials for biodegradable composite applications. In *Nanocomposites and Polymers with Analytical Methods* (pp. 243-260).

Habibi, M., Laperriere, L., Lebrun, G., & Chabot, B. (2016). Experimental investigation of the effect of short flax fibers on the permeability behaviour of a new unidirectional flax/paper composite. *Fibers, 4*(22), 1-16.

Han, Y., Feng, W. J., Cheng, W., Chen, F., & Chen, R. R. (2011). Application of composite addictives in paper-making using slag-wool fiber. *International Journal of Chemistry, 3*(1), 176-180.

Istek, A., Aydemir, D., & Aksu, S. (2010). The Effect of decor paper and resin type on the physical, mechanical, and surface quality properties of particleboards coated with impregnated decor papers. *BioResources, 5*(2), 1074-1083.

Khalilitabas, A. A., Khorrami, M., & Sobhani, J. (2009). Effects on wood-pulp fibers on the mechanical properties of cement composites. *In The 3rd International Conference on Concrete and Development,* Tehran, Iran.

Kiaei, M., Mahdavi, S., Kialashaki, A., Nemati, M., Samariha, A., & Saghafi, A. (2014). Chemical composition and morphological properties of canola plant and its potential application in pulp and paper industry. *Cellulose Chemistry and Technology, 48*(1), 105-110.

Kroling, H., Fleckenstein, J., Nubbo, N., Endres, A., Miletzky, F., & Schabel, S. (2014). Non-woven and paper based epoxy composites. *Science and Technology- Das Papier, 6*, 1-5.

Madsen, B., & Gamstedt, E. K. (2013). Wood versus plant fibers: Similarities and differences in composite applications. *Advances in Materials Science and Engineering, 2013*, 1-14.

Mohd Kassim, A. S. , Mohd Aripin, A., Ishak, N., Zainulabidin, M. H., & Abang Zaidel, N. F. (2016). Oil palm leaf fibre and its suitability for paper-based products. *ARPN Journal of Engineering and Applied Sciences, 11*(11), 7364-7369.

Moral, A., Aguado, R., TIjero, A., Tarres, Q., Delgado-Aguilar, D., & Mutje, P. (2014). High-yield pulp from Brassica napus to manufacture packaging paper. *BioResources.com, 12*(2), 2792-2804.

Nayak, N. C., & Mishra, A. (2013). Development and mechanical characterization of Palmyra fruit fiber reinforced epoxy composites. *Journal of Production Engineering, 16*(2), 69-72.

Ng, L. F., Sivakumar, D., Zakaria, K. A., Bapokutty, O., & Sivaraos. (2017). Influence of Kenaf fibre orientation effect on the mechanical properties of hybrid structure of fibre metal laminate. *Pertanika Journal Science and Technology, 25*, 1-8.

Oladele, I. O., & Okoro, M. A. (2015). Development of rattan (*Calamus longipinna*) particulate reinforced paper pulp based composites for structural application using waste papers. *Leonardo Journal of Sciences, 14*(27), 75-87.

Pal, L., & Joyce, M. (2017). Paper need not be: Paper and biomaterials industries need to converge to bring about true innovation. *BioResources.com, 1*2(2), 2249-2251.

Praharaj, A. P., Behera, D., & Bastia, T. K. (2014). Fabrication and mechanical properties of BisGMA/Amine functionalized paper pulp composites. *International Journal of Innovative Research in Science, Engineering and Technology, 3*(5), 12718-12723.

Qiu, R., Liu, W., & Li, K. (2015). Investigation of bamboo pul fiber-reinforced unsaturated polyester composites. *De Gruyter, 69*(8), 967-974.

Radif, Z. S., Ali, A., & Abdan, K. (2011). Development of a green combat armour from Rame-Kevlar-Polyester composite. *Pertanika Journal Science and Technology, 19*(2), 339-348.

Ren, H., Zhang, Y., Zhai, H., & Chen, J. (2015). Production and evaluation of biodegradable composites based on polyhydroxybutyrate and polylactic acid reinforced with short and long pulp fibers. *Cellulose Chemistry and Technology, 49*(7), 641-652.

Robillard, M., & Lebrun, G. (2010). Processing and Mechanical Properties of Unidirectional Hemp-Paper/Epoxy Composites. *In The 10th International Conference on Flow Processes in Composite Materials (FPCM 10),* Monte Verita, Ascona.

Sangrutsamee, V., Srichandr, P, & Poolthong, N. (2012). Re-pulped waste paper-based composite building materials with low thermal conductivity. *Journal of Asian Architecture and Building Engineering, 151*, 147-151.

Rahman, G. S., Al Mamun, M. A., Bashar, M. S., Mostafa, M. G., Ali, M. F., & Khan, M. A. (2014). A study on comparison between recycled waste paper reinforced polymer composite and hardboard. *Rajshahi University Journal of Environment Science, 3*, 9-15.

Suriani, M. J., Ali, A., Sapuan, S. M., & Khalina, A. (2013). Aspect of fatigue analysis of composite materials: A review. *Pertanika Journal Science and Technology, 21*(1), 1-14.

Zhang, S. Y., Fei, B. H., Yu, Y., Cheng, H. T., & Wang, C. G. (2013). Effect of the amount of lignin on tensile properties of single wood fibers. *Forest Science and Practice, 15*(1), 56-60.

Review Article

Platelet Transcriptome-Based Approaches in the Fight against Dengue and Other Diseases

Suppiah, J.[1], Sakinah, S.[1], Chan, S. Y.[1], Wong, Y. P.[1], Bala, J. A.[2], Lawal, N.[2], Benelli, G.[3], Subbiah, S. K.[1] and Chee, H. Y.[1]*

[1]*Department of Medical Microbiology and Parasitology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Selangor, Malaysia*
[2]*Department of Virology, Faculty of Veterinary Medicine, Universiti Putra Malaysia, 43400 UPM, Selangor, Malaysia*
[3]*Department of Agriculture, Food and Environment, University of Pisa, via delBorghetto 80, 56124 Pisa, Italy*

ABSTRACT

Human platelets are anucleate cells that lack in deoxyribonucleic acid (DNA), thus hampering genomic study on them. However, the presence of their own messenger ribonucleic acid (mRNA) transcript allows functional study via the transcriptome approach. Transcriptome not only allows profiling of platelet but also aids in studying gene regulation in virus infections and other diseases that have an impact on platelets. Some viruses are known to affect the platelet either by causing a reduction or destruction. Dengue virus is one of the most postulated virus having such effect and frequently linked to platelet reduction. The transcriptome approach has a pivotal role in providing a deeper insight to link certain diseases and their effect on platelets. This review critically discusses role of platelet in dengue and other viral diseases of public health relevance, with a specific focus on the methods currently used in platelet transcriptome profiling.

INTRODUCTION

Platelets are a component of blood that generally contributes to haemostasis, a process to stop bleeding by clumping and clotting blood vessel injuries. Platelets are anucleate cells derived from megakaryocytes in the bone marrow. Once endoplasmatic

maturation takes place, megakaryocytes develop pro-platelets, which bud off numerous platelets into the bloodstream. This mechanism is accentuated by cytokines such as interleukin 3 (IL-3), interleukin 6 (IL-6), granulocyte macrophage colony-stimulating factor (GM-CSF), stromal cell-derived factor-1 (SDF-1) and vitally by thrombopoietin (TPO) (Kuter, 2014). Platelet is uniquely found in mammals, whereas in other animals such as amphibians and avian, it circulates as intact mononuclear cells. The regulation of megakaryocyte and platelet production is done by TPO, a hormone produced in the kidneys and liver. In a healthy adult, approximately 10^{11} platelets are continuously produced and cleared daily to maintain $150–400 \times 10^9$/L of blood level (Balduni & Noris, 2014; Grozovsky, Giannini, Falet, & Hoffmeister, 2015).

To examine in depth the function of platelets, researchers have attempted to investigate the products of the genome, which are proteins. Proteomics provides details on protein quantity and diversity, but on the other hand, it may not depict a cell's entire story. Moreover, translational modification of proteins poses a technical challenge in characterising proteome (Gregorich & Ge, 2014). In order to overcome this, measuring the intermediate step between genes and proteins, called transcripts of messenger ribonucleic acid (transcriptome), alternatively bridges the gap between the genomics and proteomics.

Transcriptome profiling provides information on up-regulation or down-regulation of genes and quantity of gene expression in a cell. This information will be useful to understand the mechanism of certain diseases caused by thrombocytopenic condition involving platelets. For instance, transcriptional profiling of platelets reveals WD-40 repeat domain 1 (WDR1) vanquishes platelet activity and is associated with cardiovascular disease (Montenont et al., 2016). Other applications of platelet transcriptome profiling include identifying altered platelet function due to mutations (Noetzli et al., 2015), monitoring response of anti-platelet therapy in hyperthrombic patients (Mao et al., 2014) and potential biomarkers of disease (Best et al., 2015).

In addition to these applications, platelet transcriptome profiling has the potential to provide a breakthrough in viral infections that affect platelets. It is known that thrombocytopenia is the hallmark of dengue infection, however only in recent years, researchers have been able to prove the replication of dengue virus in platelet (Sutherland et al., 2016; Simon, Sutherland, & Pryzdial, 2015). Platelet transcriptome profiling, yet to be used in dengue research could be another stepping stone in understanding mechanism of dengue infection, similarly in other viral infections too.

HUMAN PLATELET TRANSCRIPTOME

Human platelets are known to be anucleate, lacking in DNA and produced as cytoplasmic buds from megakaryocytes, they retain megakaryocyte-derived messenger ribonucleic acid (mRNA) (Bahou & Gnatenko, 2004). There is a paucity of studies on platelet transcriptome; however, steady observations support the fact that platelet mRNA is biologically and patho-physiologically significant. Earlier studies show platelet transcriptome directly correlates with proteomic data indicating that transcriptional analysis can provide insight on platelet function and disorders. Adversely, one study reports a weak association of proteome data with human platelet transcriptome obtained from the healthy male individual (Londin et al., 2014). Having

said that, to date, it has been revealed that approximately 69% of secreted platelet-associated proteins are detectable at the mRNA level (McRedmond et al., 2004).

Additionally, 2000 mRNA transcripts are found to be present in unstimulated platelets, thus providing a better chance for evaluation of the transcriptome (Gnatenko et al., 2003). It is also believed that platelet transcriptome is fixed, with minimal changes indicating that significant changes in transcripts may need days compared to hours for nucleated cells. An interesting data to note, 58% of the mRNAs expressed by human platelets are found in mouse platelets while 83% vice versa (Rowley et al., 2011). In addition, transcriptome profiling reveal platelet has 50%-90% less ribosomal RNA, high levels mRNA and small RNAs and contains various isoforms of mRNA (Bray et al., 2010).

Prior investigation on platelet mRNA show integrins and glycoproteins are significantly higher compared with other components (Bugert & Kluter, 2006). The over-presentation of integrins may reflect their utmost importance in biological functions of the platelet. Transcriptome data have allowed novel discovery of transporters for gamma-aminobutyric acid (GABA), glutamate and dopamine molecules in platelets (Frankhauser et al., 2006; Rainesalo, Keranen, Saransaari, & Honkaniemi, 2005).

It is useful to characterise platelet transcriptome to identify the genetic aspect of disease and disease traits. Capturing primary megakaryocytes mRNA profile is not feasible when it involves many subjects. Thus, profiling platelet RNA is an alternative as it is easy to be procured compared with to megakaryocytes. There are several methods to study the platelet transcriptome of which some may become obsolete due to advancement in molecular methods (Figure 1).
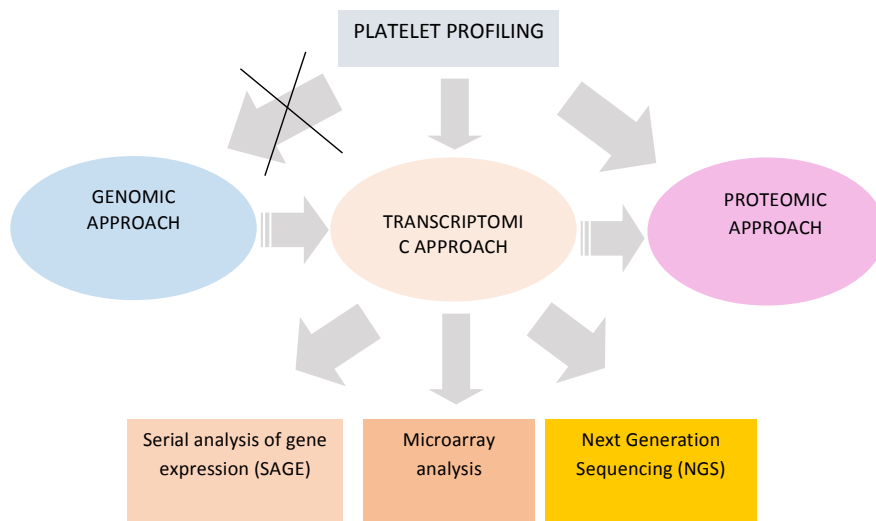
Figure 1. Methods currently available for platelet transcriptome profiling

Platelet profiling can be performed via the transcriptomic and proteomic approaches. On the other hand, genomic approach is not possible as the platelets do not have nuclear DNA. Transcriptomic profiling enables characterisation of platelet at the stage of mRNA transcription, a step ahead of protein translation (proteomic approach). Three identified methods of platelet transcriptome profiling are serial analysis of gene expression (SAGE), microarray analysis and next generation sequencing (NGS). Microarray analysis and NGS are advanced methods which are currently preferred by researchers as tools for platelet profiling.

THE LINK BETWEEN PLATELET AND DENGUE INFECTION

Dengue is a mosquito-borne viral disease of public health importance (Benelli & Mehlhorn, 2016). Platelet reduction, also known as thrombocytopenia is a hallmark of dengue virus infection. However, the actual mechanism on how dengue virus affects platelet during infection is not well understood. Besides, there are variations in research findings that make it difficult to come to a consensus conclusion on the link between platelet and dengue virus.

Recent studies demonstrated that dengue virus binds to lectin receptor and triggers platelet activation and apoptosis, which generates inflammatory responses in target monocytes (Hottz et al., 2014; Sun et al., 2007). Contradictory to this, a study observed the presence of dengue RNA in a highly augmented CD61 (+) cell population from infected rhesus monkey during the acute phase (Noisakran et al., 2012). This result may suggest that dengue virus utilises CD61 receptor for binding to the platelet. Another research reported the detection of dengue virus in platelets isolated from dengue patients (Noisakran et al., 2009). This raised the possibility of dengue virus replicating in platelet cells. In a most recent study in 2015, researchers provide the first evidence that dengue virus attacks platelets and utilise their translational machinery to replicate and produce infectious virus (Simon, Sutherland, & Pryzdial, 2015).

A prospective observational study in India found significant correlation between platelet and duration of hospitalisation of dengue patients. A decline in platelet level increases complication rate and prolonged hospitalisation, thus platelet count can be utilised as a predictive parameter (Jayanthi & Thulasi, 2016). In another study, correlation of bleeding manifestation and platelet count is assessed in patients with severe dengue. The study, however, finds poor correlation between thrombocytopenia and bleeding manifestations (Sreenivasa, Manjunatha, & Nivil, 2016). This perhaps indicates that abnormal platelet aggregation or disseminated intravascular coagulation may contribute to bleeding in severe dengue cases rather than number of platelets.

New evidence on correlation between activation and depletion of platelets in dengue patients with thrombocytopenia has surfaced. The findings a study show a substantial decrease in platelet count occur on day four of fever in dengue patients. Furthermore, high copy numbers of dengue virus genome in platelet pellet directly correlated with the platelet activation markers during day four, six and eight of fever are observed. Based on the observation that platelet activation is an important determinant of thrombocytopenia in dengue infections, the study suggests a controlled strategy of platelet inactivation may save them from rapid decline during dengue infections (Ojha et al., 2017).

PLATELETS AND OTHER VIRAL INFECTIONS

Platelet activation is not only limited to responding to injury but also indicates host response and virus survival. Thrombocytopenia, a condition due to either decreased in platelet synthesis or increased in platelet destruction, is common following certain viral infections (Table 1). Viruses' ability to diminish the levels of circulating platelets is a multifactorial event and differs based on the mechanism. For instance, simian immunodeficiency virus (SIV) alters TPO production in the liver by up-regulating tumour growth factor beta (TGF β) (Pate et al., 2014) whereas hepatitis C virus (HCV) induces bone marrow suppression and hypersplenism (Fouad, 2013). The spleen size is inversely correlated to platelet count in patients with chronic HCV infection (Medeiros, Domingues, Luna, & Lopes, 2014; van der Meer et al., 2016).

Platelet destruction, on the other hand, occurs via direct interaction of platelets and virus. Some viruses have the tendency to bind to platelet by utilising receptors, which enhance the interaction of platelets and neutrophils. Subsequently, platelets are phagocytised, thus, reducing the number of circulating platelets. Besides that, a reduction in platelet count sometimes is caused by self-destruction due to the presence of anti-platelet autoantibody. This leads to a condition called idiopathic thrombocytopenic purpura (ITP) that is commonly caused in HCV, cytomegalovirus (CMV), Human Immunodeficiency virus (HIV), Epstein Barr virus (EBV), Varicella Zoster virus (VZ) and herpes viruses (Assinger, 2014; Assinger et al., 2014; Goeijenbier et al., 2012). Studies have shown antibodies produced by B-lymphocytes against these viruses interrupt the surface integrins on the platelet such as the GPIIb/IIIa or GPIb-IX-V and diminish survival of platelets (Hamidpour, Behrendt, Griffiths, Partridge, & Lindsey, 2006; Najaoui et al., 2012).

HOW PLATELET TRANSCRIPTOME CAN BE USED TO STUDY MECHANISMS OF DENGUE INFECTION IN PLATELET

To establish an infection, almost all viruses exhibit the same modus operandi, which includes entry to the host's cells, replication via hosts' internal machinery and release of new virions (Li & Nagy, 2011). However, some viral replication mechanisms are less understood due to lack of laboratory evidences. For instance, replication of dengue virus in platelets remains unclear until recent study shows that dengue virus uses platelets to replicate (Simon, Sutherland, & Pryzdial, 2015). Thus, the key to unveiling further about dengue replication lies within the platelet, and platelet transcriptome can be utilised to study the mechanism of infection by dengue virus. To date, platelet transcriptome profiling has yet to be used as a tool to study the mechanism of dengue infection. One proposed method for future research includes *in vitro* study comparing the platelet transcriptome before and after dengue virus infection to identify genes, which are involved in virus transcription. Apart from that, it would be interesting to analyse platelet transcriptome profile at different phases of dengue infection or based on the severity of the disease.

Table 1

Mechanisms of action of selected viral infections on platelet

Virus	Mechanism of action on platelet	Study
Simian Immunodeficiency Virus (SIV)	SIV alters Hepatic thrombopoietin (TPO) production by up-regulating tumour growth factor.	Pate et al., 2014
Hepatitis C virus (HCV)	HCV interacts with platelets through collagen receptor GPVI. It stimulates bone marrow suppression and hypersplenism. The size of the spleen is inversely related to platelet count.	van der Meer et al., 2016
Cytomegalovirus (CMV)	CMV is known to replicate in megakaryocyte. CMV binds to platelet via the TLR2 receptor, initiating platelet activation and degranulation. This results in enhanced platelet interaction with neutrophils, phagocytosis, and drop in platelet count.	Assinger et al., 2014
Human immunodeficiency virus (HIV)	Multiple mechanisms: (1) Reduce survival of megakaryocytes and precursors; (2) HIV surface glycoprotein gp120 binds to lection receptor and leads to increased megakaryocyte apoptosis; (3) reduce TPO receptor (c-Mpl) expression.	Assinger, 2014
Hantavirus	Binds to $\alpha v \beta 3$ or $\alpha IIb \beta 3$ integrins expressed on platelets and endothelial cells, contributing to viral dissemination, platelet activation, and induction of endothelial cell functions.	Zapata, Cox, & Salvato, 2014
Epstein Barr Virus (EBV) Dengue virus	Interact with platelets by complementing receptor 2 (CR2) and induces platelet activation. Bind to lectin receptor and enhance platelet activation and apoptosis, which generates inflammatory responses in target monocytes. Anti-non-structural protein-1 (NS-1) triggers complement-mediated lysis of platelets which further accelerates thrombocytopenia. Dengue virus has been detected inside the platelet, raising the possibility of direct replication in platelet.	Ahmad & Menezes, 1997 Hottz et al., 2014 Sun et al., 2007 Noisakran et al., 2009
Rotavirus	Rotavirus binds to platelet using collagen receptor GPIa/IIa thus triggering platelet activation and destruction.	Assinger, 2014

METHODS OF PLATELET TRANSCRIPTOME PROFILING

Platelet Transcriptome Profiling Approach by SAGE

The serial analysis of gene expression (SAGE) is a quantitative method for measuring gene expression; it was first, developed and published in 1995 (Velculescu, Zhang, Vogelstein, & Kinzler, 1995). Though lower in throughput than microarray analysis, SAGE provides a more detailed transcriptome profiling due to its quantitative property while microarray has been just a qualitative method. The difference between SAGE and microarray has been widely

studied (Gnatenko et al., 2003). The study identified 89% of platelet mRNA transcripts are of mitochondrial origin, while 11% corresponds to unique genes which are not even detected by microarray. Therefore, SAGE is a good technique that may permit full characterisation of the mRNA transcripts including non-mitochondrial origin. Nevertheless, with the development of high-throughput sequencing technologies and more advanced microarray techniques, it appears possible that SAGE may be surpassed in future studies as detailed in Table 2.

Platelet Transcriptome Profiling by NGS

Next-generation sequencing (NGS) of platelets has enabled exceptional quantification and profiling of platelet transcriptome and discovered the diversity of mRNAs. The NGS has wide applications in this aspect, such as comparing platelet transcriptome between virus-infected and non-infected cells. One study uses NGS approach to evaluate various pathogen reduction systems (PR) for platelet storage (Osman, Hitzler, Ameur, & Provost, 2015). PR systems function to extend the shelf life of stored blood components, such as platelets and inhibit infections transmitted via transfusion. It is found that use of certain PR systems greatly reduces mRNA level in PR-treated platelet concentrates compared with non-treated platelet. This finding will be useful in alerting the blood bank authorities in implementing appropriate PR system for blood component storage (Osman, Hitzler, & Provost, 2016).

The NGS is also used to compare platelet transcriptome between mice and human (Rowley et al., 2011). Mouse platelets are commonly used as surrogates to study in vivo human platelet function. Still, uncertainty often arises regarding the functional differences and similarities between mouse and human platelets. This is the first study that utilises NGS to analyse human and mice platelet. The study discovered 95% conserved platelet transcriptome between human and mice. Conservation of platelet mRNA between two species enables mice model to be used to study haemostasis.

Though next-generation sequencing enables large scale platelet transcriptome profiling, its application for clinical studies is confined. NGS is foreseen to replace some of the old techniques used in the clinical setting. For instance, it may be used in assessing aspirin resistance in heart disease patients. To date, the influence of aspirin is commonly evaluated using light transmission aggregometry (LTA) and thrombelastography platelet mapping assay (TEG) (Liu et al., 2013). NGS approach to determine the platelet mRNA profile in association with aspirin resistance will provide more data in predicting genetic markers of aspirin responsiveness. The NGS generates a large transcript data per sample, and therefore, it is challenging to perform analysis with such data. The RNA sequencing in a much larger population will need to be performed so that a normal platelet transcript can be established (Freedman, 2011).

Microarray Analysis in Characterisation of Platelet Transcriptome

Apart from NGS, microarray analysis is another method commonly used for platelet transcriptome profiling. This method is beneficial in observing up-regulation and down-regulation of genes without the need for quantification as applied by NGS. One study had

attempted to compare platelet transcriptome from three unrelated donors by using leukocyte depleted platelet subjected to hybridisation to the Affymetrix HG-U95Av2 GeneChip, consisting of 12,600 probe sets. It found exceptional concordance, reproducibility and discrete profiles of human platelets among the three donors (Bahou & Gnatenko, 2004). In a most recent study, microarray analysis is performed to compare neonate and adult human platelet transcriptomes using pure platelet RNA obtained from adult and cord blood. Interestingly, 201 genes are found to be differentially expressed between these two groups of which neonatal platelets had higher amounts of mRNA that are associated with protein synthesis and lower levels of genes involved in calcium metabolism and cell signalling (Caparrós-Pérez et al., 2017).

Microarray approach seems to be more feasible for clinical application. Coronary heart disease is a condition in which platelet aggregation occurs and genetically modulated (Floyd, Ellis, & Ferro, 2014). Hereditary platelet glycoprotein polymorphisms are a known risk factor for coronary heart disease. The gene encoding glycoprotein GPIIIa has displayed a polymorphism (PlA2 allele) that is linked to vascular disease. Several case report studies identified the presence of this polymorphism in patients who died of myocardial infarction, stroke and related heart diseases (O'Donnell et al., 2001). It would be interesting to utilise microarray for comparison profiling of this platelet-related gene in patients who possess heterozygous or homozygous condition of the allele. In addition, platelet transcriptome analysis by microarray is useful in chronic kidney disease patients (CKD). An alteration in the uremic platelet mRNA is observed in CKD patients (Ple et al., 2012). However, dialysis appeared to have corrected this defect, which is consistent with the favourable effect of dialysis.

The microarray platelet profile is also helpful in providing insight into transcriptional signaling pathways in platelets in sickle cell anaemia disease. A study utilising microarray identified the absence of expression of CD45 and CD5 in sickle cell anaemic patients. These are lymphocytes and T cells markers. On the other hand, significant up-regulation of mRNAs encoding arginase II and ornithine decarboxylase antizyme in patients with sickle cell disease is observed in comparison with control subjects. It is known that arginine metabolic enzymes are involved in homeostasis (Raghavachari et al., 2007).

Microarray approach may have its limitations. There is no standardised method for microarray data analysis that, can contribute to inter-laboratory variations. Microarray assays also require a minimum of 5 mg of total cellular RNA as starting material, thus to prepare pure leukocyte depleted platelet RNA, a large volume of blood may be needed (Gnatenko et al., 2003). In addition, the microarray is no doubt a very costly method, which may be a prohibitive factor in a large-scale study of platelet transcriptome. This factor should be taken into stringent consideration when customising the microarray probe chips in regard to which gene should be spotted (Eicher et al., 2016). Prioritising the importance of genes to be spotted and prior knowledge is required to prevent wastage of data and prohibitive cost.

Table 2
Current knowledge on platelet profiling

Method	Key findings	Up-regulated/down-regulated	Reference
SAGE	Detected 89% mitochondrial transcripts and 11% of non-mitochondrial transcripts.	-	Gnatenko et al., 2003
Microarray	Detected mainly unknown function genes, genes involve in metabolism and receptors.	Majority of the up-regulated genes are unknown functional genes.	Gnatenko et al., 2003
Microarray	Detected abundant glycoproteins and integrins	Glycoprotein/integrins and receptors are majorly up-regulated.	Bugert & Kluter, 2006
Microarray	Human platelet transcriptome correlated well with proteome profile.	-	McRedmond et al., 2004
Microarray	Evaluation of platelet profile from 3 healthy donors revealed concordance, reproducibility and distinct profiles of human platelets.	-	Bahou & Gnatenko, 2004
Microarray	An alteration of the uremic platelet mRNA is observed in CKD patients. Dialysis appeared to have corrected this defect, which is consistent with the favourable effect of dialysis.	23 and 17 mRNAs are differentially expressed in platelets from dialysis and uremic patients respectively Defective genes such as PCTP, RAB1A, ZFAND5, and USP15 are corrected by dialysis.	Ple et al., 2012
Microarray	Demonstrated down-regulation of transcriptional signaling pathways in platelets in sickle cell anaemia patients.	CD45 and CD5 are down-regulated; arginase II and ornithine decarboxylase antizyme ae up-regulated.	Raghavachari et al., 2007
NGS	Abnormal transcript reads detected from an individual with autosomal recessive gray platelet syndrome.	Mutation in NBEAL2 gene responsible for causing gray platelet syndrome.	Kahr et al., 2011
NGS	Evaluated various pathogen reduction systems (PR) for platelet storage and finds that Intercept system alters platelet quality.	-	Osman et al., 2015
NGS	95% conserved platelet transcriptome between human and mice enables mice model to be used to study haemostasis.	-	Rowley et al., 2011

CONCLUSION

In sum, the study has highlighted that platelets are well suited for transcriptomic studies, as they lack nuclear DNA and their genome consists of megakaryocytes derived mRNA transcripts. There are many available transcriptomic approaches and analysis methods. While microarray is one of the widely-employed methods, next-generation sequencing is gradually becoming a favoured tool for transcriptome analysis. However, the limitations of modern transcript profiling include lack of reproducibility, difficulty in data processing and complexity of data

produced. Despite that, these methods are still reliable in clinical applications. In future, platelet transcriptome studies investigating a more diverse set of healthy and diseased samples will add value to the existing knowledge on platelet thrombotic and non-thrombotic functions.

REFERENCES

Ahmad, A., & Menezes, J. (1997). Binding of the Epstein-Barr virus to human platelets causes the release of transforming growth factor-beta. *The Journal of Immunology, 159*(8), 3984-3988.

Assinger, A. (2014). Platelets and infection - an emerging role of platelets in viral infection. *Frontiers in Immunology, 5*, 649-660. doi:10.3389/fimmu.2014.00649

Assinger, A., Kral, J. B., Yaiw, K. C., Schrottmaier, W. C., Kurzejamska, E., Wang, Y., … & Soderberg-Naucler, C. (2014). Human cytomegalovirus-platelet interaction triggers toll-like receptor 2-dependent proinflammatory and proangiogenic responses. *Arteriosclerosis, Thrombosis, and Vascular Biology, 34*(4), 801-809. doi:10.1161/ATVBAHA.114.303287

Bahou, W. F., & Gnatenko, D. V. (2004). Platelet transcriptome: the application of microarray analysis to platelets. *Seminars in Thrombosis and Hemostasis, 30*(4), 473-484. doi:10.1055/s-2004-833482

Balduini, C. L., & Noris, P. (2014). Platelet count and aging. *Haematologica, 99*(6), 953–955. doi:10.3324/haematol.2014.106260

Benelli, G., & Mehlhorn, H. (2016). Declining malaria, rising of dengue and Zika virus: Insights for mosquito vector control. *Parasitology Research, 115*(5), 1747-1754. doi:10.1007/s00436-016-4971-z

Best, M. G., Sol, N., Kooi, I., Tannous, J., Westerman, B. A., Rustenburg, F., … & Wurdinger, T. (2015). RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell, 28*(5), 666-676. doi: 10.1016/j.ccell.2015.09.018

Bray, P. F., Fortina, P. M., Nagalla, S., Delgrosso, K., Ertel, A., Rigoutsos, I., & McKenzie, S. E. (2010). High-throughput dequencing of the human platelet transcriptome. *Blood, 116*(21), 481-481.

Bugert, P., & Kluter, H. (2006). Profiling of gene transcripts in human platelets: An update of the platelet transcriptome. *Platelets, 17*(7), 503-504. doi:10.1080/09537100600901491

Caparrós-Pérez, E., Teruel-Montoya, R., López-Andreo, M. J., Llanos, M. C., Rivera, J., Palma-Barqueros, V., … & Ferrer-Marin, F. (2017). Comprehensive comparison of neonate and adult human platelet transcriptomes. *PLoS One, 12*(8), e0183042. doi:10.1371/journal.pone.0183042

Eicher, J. D., Wakabayashi, Y., Vitseva, O., Esa, N., Yang, Y., Zhu, J., … & Johnson, A. D. (2016). Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets, 27*(3), 230–239. doi:10.3109/09537104.2015.1083543

Floyd, C. N., Ellis, B. H., & Ferro, A. (2014). The PlA1/A2 polymorphism of glycoprotein IIIa as a risk factor for stroke: A systematic review and meta-analysis. *PLoS One, 9*(7), e100239. doi:10.1371/journal.pone.0100239

Fouad, Y. M. (2013). Chronic hepatitis C-associated thrombocytopenia: Aetiology and management. *Tropical Gastroenterology, 34*(2), 58-67.

Frankhauser, P., Grimmer, Y., Bugert, P., Deuschle, M., Schmidt, M., & Schloss, P. (2006). Characterization of the neuronal dopamine transporter DAT in human blood platelets. *Neuroscience Letters, 399*(3), 197-201. doi:10.1016/j.neulet.2006.01.062

Freedman, J. E. (2011). A platelet transcriptome revolution. *Blood, 118*(14), 3760-3761. doi:10.1182/blood-2011-05-356600

Gnatenko, D. V., Dunn, J. J., McCorkle, S. R., Weissmann, D., Perrotta, P. L., & Bahou, W. F. (2003). Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood, 101*(6), 2285-2293. doi:10.1182/blood-2002-09-2797

Goeijenbier, M., van Wissen, M., van de Weg, C., Jong, E., Gerdes, V. E., Meijers, J. C., .& van Gorp, E. C. (2012). Review: Viral infections and mechanisms of thrombosis and bleeding. *Journal of Medical Virology, 84*(10), 1680-1696. doi:10.1002/jmv.23354

Gregorich, Z. R., & Ge, Y. (2014). Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics, 14*(10), 1195-1210. doi:10.1002/pmic.201300432

Grozovsky, R., Giannini, S., Falet, H., & Hoffmeister, K.M. (2015). Regulating billions of blood platelets: glycans and beyond. *Blood, 126*(16), 1877-1884. doi: 10.1182/blood-2015-01-569129

Hamidpour, M., Behrendt, M., Griffiths, B., Partridge, L., & Lindsey, N. (2006). The isolation and characterisation of antiplatelet antibodies. *European Journal of Haematology, 76*(4), 331-338. doi:10.1111/j.1600-0609.2005.00614.x

Hottz, E. D., Medeiros-de-Moraes, I. M., Vieira-de-Abreu, A., de Assis, E. F., Vals-de-Souza, R., Castro-Faria-Neto, H. C., … & Bozza, P. T. (2014). Platelet activation and apoptosis modulate monocyte inflammatory responses in dengue. *The Journal of Immunology, 193*(4), 1864-1872. doi:10.4049/jimmunol.1400091

Jayanthi, H. K., & Tulasi, S. K. (2016). Correlation study between platelet count, leukocyte count, nonhemorrhagic complications, and duration of hospital stay in dengue fever with thrombocytopenia. *Journal of Family Medicine and Primary Care, 5*(1), 120-130. doi:10.4103/2249-4863.184635

Kahr, W. H., Hinckley, J., Li, L., Schwertz, H., Christensen, H., Rowley, J. W., … & Di Paola, J. (2011). Mutations in NBEAL2, encoding a BEACH protein, cause gray platelet syndrome. *Nature Genetics, 43*(8), 738-740. doi:10.1038/ng.884

Kuter, D. J. (2014). Milestones in understanding platelet production: a historical overview. *British Journal of Haematology, 165*(2), 248-258. doi:10.1111/bjh.12781

Li, Z., & Nagy, P. D. (2011). Diverse roles of host RNA-binding proteins in RNA virus replication. *RNA Biology, 8*(2), 305–315. doi:10.4161/rna.8.2.15391

Liu, X. F., Cao, J., Fan, L., Liu, L., Li, J., Hu, G. L., … & Li, X. L. (2013). Prevalence of and risk factors for aspirin resistance in elderly patients with coronary artery disease. *Journal of Geriatric Cardiology, 10*(1), 21-27. doi:10.3969/j.issn.1671-5411.2013.01.005

Londin, E. R., Hatzimichael, E., Loher, P., Edelstein, L., Shaw, C., Delgrosso, K., … & Rigoutsos, I. (2014). The human platelet: strong transcriptome correlations among individuals associate weakly with the platelet proteome. *Biology Direct, 9*(1), 3. doi:10.1186/1745-6150-9-3

Mao, Y., Lei, L., Su, J., Yu, Y., Liu, Z., & Huo, Y. (2014). Regulators of G protein signaling are up-regulated in aspirin-resistant platelets from patients with metabolic syndrome. *Pharmazi, 69*(5), 371–373. doi: 10.1691/ph.2014.3833

McRedmond, J. P., Park, S. D., Reilly, D. F., Coppinger, J. A., Maguire, P. B., Shields, D. C., & Fitzgerald, D. J. (2004). Integration of proteomics and genomics in platelets: A profile of platelet proteins and platelet-specific genes. *Molecular and Cellular Proteomics, 3*(2), 133-144. doi:10.1074/mcp.M300063-MCP200

Medeiros, T. B., Domingues, A. L., Luna, C. F., & Lopes, E. P. (2014). Correlation between platelet count and both liver fibrosis and spleen diameter in patients with schistosomiasis mansoni. *Arquivos de Gastroenterologia, 51*(1), 34-38.

Montenont, E., Echagarruga, C., Allen, N., Araldi, E., Suarez, Y., & Berger, J. S. (2016). Platelet WDR1 suppresses platelet activity and is associated with cardiovascular disease. *Blood, 128*(16), 2033–2042. doi:10.1182/blood-2016-03-703157

Najaoui, A., Bakchoul, T., Stoy, J., Bein, G., Rummel, M. J., Santoso, S., & Sachs, U. J. (2012). Autoantibody-mediated complement activation on platelets is a common finding in patients with immune thrombocytopenic purpura (ITP). *European Journal of Haematology, 88*(2), 167-174. doi:10.1111/j.1600-0609.2011.01718.x

Noetzli, L., Lo, R. W., Lee-Sherick, A. B., Callaghan, M., Noris, P., Savoia, A., & Di Paola, J. (2015). Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nature Genetics, 47*(5), 535–538. doi:10.1038/ng.3253

Noisakran, S., Gibbons, R. V., Songprakhon, P., Jairungsri, A., Ajariyakhajorn, C., Nisalak, A., & Perng, G. C. (2009). Detection of dengue virus in platelets isolated from dengue patients. *The Southeast Asian Journal of Tropical Medicine and Public Health, 40*(2), 253-262.

Noisakran, S., Onlamoon, N., Pattanapanyasat, K., Hsiao, H. M., Songprakhon, P., Angkasekwinai, N., & Perng, G. C. (2012). Role of CD61+ cells in thrombocytopenia of dengue patients. *International Journal of Hematology, 96*(5), 600-610. doi:10.1007/s12185-012-1175-x

Ojha, A., Nandi, D., Batra, H., Singhal, R., Annarapu, G. K., Bhattacharyya, S., & Gucchait, P. (2017). Platelet activation determines the severity of thrombocytopenia in dengue infection. *Scientific Report, 7*, 41697. doi:10.1038/srep41697

Osman, A., Hitzler, W. E., & Provost, P. (2016). Peculiarities of studying the effects of pathogen reduction technologies on platelets. *Proteomics Clinical Applications, 10*(8), 805–815. doi:10.1002/prca.201500124

Osman, A., Hitzler, W. E., Ameur, A., & Provost, P. (2015). Differential expression analysis by RNA-seq reveals perturbations in the platelet mRNA transcriptome triggered by pathogen reduction systems. *PLoS One, 10*(7), e0133070. doi:10.1371/journal.pone.0133070

O'Donnell, C. J., Larson, M. G., Feng, D., Sutherland, P. A., Lindpaintner, K., Myers, R. H., & Framingham Heart, S. (2001). Genetic and environmental contributions to platelet aggregation: the Framingham heart study. *Circulation, 103*(25), 3051-3056.

Pate, M. K. A., Lyons, C. E., Dorsey, J. L., Queen, S. E., Adams, R. J., Morrell, C. N., & Mankowski, J. L. (2014). TGFbeta-mediated downregulation of thrombopoietin is associated with platelet decline in asymptomatic SIV infection. *Journal of Acquired Immune Deficiency Syndrome, 65*(5), 510-516. doi:10.1097/QAI.0000000000000048

Ple, H., Maltais, M., Corduan, A., Rousseau, G., Madore, F., & Provost, P. (2012). Alteration of the platelet transcriptome in chronic kidney disease. *Thrombosis and Haemostasis, 108*(4), 605-615. doi:10.1160/TH12-03-0153

Raghavachari, N., Xu, X., Harris, A., Villagra, J., Logun, C., Barb, J., … & Gladwin, M. T. (2007). Amplified expression profiling of platelet transcriptome reveals changes in arginine metabolic pathways in patients with sickle cell disease. *Circulation, 115*(12), 1551-1562. doi:10.1161/CIRCULATIONAHA.106.658641

Rainesalo, S., Keranen, T., Saransaari, P., & Honkaniemi, J. (2005). GABA and glutamate transporters are expressed in human platelets. *Molecular Brain Research, 141*(2), 161-165. doi:10.1016/j.molbrainres.2005.08.013

Rowley, J. W., Oler, A. J., Tolley, N. D., Hunter, B. N., Low, E. N., Nix, D. A., & Weyrich, A. S. (2011). Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. *Blood, 118*(14), e101-111. doi:10.1182/blood-2011-03-339705

Simon, A. Y., Sutherland, M. R., & Pryzdial, E. L. (2015). Dengue virus binding and replication by platelets. *Blood, 126*(3), 378-385. doi:10.1182/blood-2014-09-598029

Sreenivasa, B., Manjunatha, B., & Nivil, J. (2017). Bleeding manifestations in dengue and their correlation with the platelet count. *Sri Lanka Journal of Child Health, 46*(3), 218-221. doi:10.4038/sljch.v46i3.8320

Sun, D. S., King, C. C., Huang, H. S., Shih, Y. L., Lee, C. C., Tsai, W. J., … & Chang, H. H. (2007). Antiplatelet autoantibodies elicited by dengue virus non-structural protein 1 cause thrombocytopenia and mortality in mice. *Journal of Thrombosis and Haemostasis, 5*(11), 2291-2299. doi:10.1111/j.1538-7836.2007.02754.x

Sutherland, M. R., Simon, A.Y., Serrano, K., Schubert, P., Acker, J. P., & Pryzdial, E. L. (2016). Dengue virus persists and replicates during storage of platelets and red blood cell units. *Transfusion, 56*(5), 1129–1137. doi:10.1111/trf.13454.

van der Meer, A. J., Maan, R., Veldt, B. J., Feld, J. J., Wedemeyer, H., Dufour, J. F., & Janssen, H. L. (2016). Improvement of platelets after SVR among patients with chronic HCV infection and advanced hepatic fibrosis. *Journal of Gastroenterology and Hepatology, 31*(6), 1168-1176. doi:10.1111/jgh.13252

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science, 270*(5235), 484-487.

Zapata, J. C., Cox, D., & Salvato, M. S. (2014). The role of platelets in the pathogenesis of viral hemorrhagic fevers. *PLoS Neglected Tropical Disease, 8*(6), e2858. doi:10.1371/journal.pntd.0002858

PERTANIKA
JOURNALS

Review Article

Forensic Body Height Estimation by Measuring Unsegmented Fingers of Javanese in Indonesia

Athfiyatul Fatati[1] and Myrtati D. Artaria[2]*

[1]*GRAMIK, Medical School, Universitas Airlangga, Jl. Dr. Moestopo, Surabaya, Indonesia*
[2]*Department of Anthropology, FISIP, Universitas Airlangga, Surabaya, Indonesia*

ABSTRACT

Estimation of stature is important in forensic examination. It is difficult to identify the deceased when the body has been mutilated and only fragmented remains are found . Body height estimation from fingers has been useful in establishing stature approximation. The present research attempts to construct a formula from measurements of the whole fingers from proximal to distal ends. The material for the present study comprises fifty (50) Mongoloid Javanese men between the ages of 21 and 25 years. Measurement of fingers and stature are taken on each subject based on internationally recommended standard methods and techniques. This study uses Pearson correlation test, and Regression analyses to determine the body height formula. The results indicate average height of the subject is 1675.96 mm. The results of index, middle, and ring finger measurements are positively correlated ($p < 0.001$) with stature. There are moderate correlations—r=0.4 to 0.5, between the length of fingers and body height. It can be concluded measuring each segment of finger can provide an estimation of stature. However, measuring the intact fingers is easier and more practical and is sometimes appreciated by the victims' families.

Keywords: Finger length, forensics, height estimation, identification, phalanx, stature

INTRODUCTION

The largest ethnic group in Indonesia is the Javanese who live in the most densely populated island of Java in Indonesia. Murderers often dismember the body of the victim to avoid identification and escape law enforcement agents. Forensic anthropology deals with identification, for legal purposes, of the skeleton in which the soft tissue have been lost partially or completely (Iscan &

Steyn, 2013). Identification of unknown skeletons of individual victims is very important, especially in cases of mutilation, natural disasters, accidents, and fires. Estimation of stature has obvious utility in the identification process. Body height estimation becomes important in cases where only body parts are available for examination (Özaslan et al., 2003). In recent years, there have been many cases that involved dismemberment or mutilated bodies. The presence of soft tissue on the human remains would usually necessitate dissection to expose skeletal elements to derive metric data for stature estimation. Many attempts have been made to find the best formula to estimate the body height of mutilated bodies, and this may differ from one population to another.

Body height is a cumulative or composite measurement, consisting of head and neck height, body height, and leg length. Body height is supported by several body elements, such as *head, vertebral column, pelvis, lower limbs* (consisting of the *femur-tibia*), and ankles. Body height varies among individuals, but what is common is the length of human bones, feet and hands are proportional to their height (Cheng et al., 1998).

In determining body height, it is advisable to utilise the long bones, instead of other bones, to get the best result. A regression formula based on the long bones has been formulated to calculate a person's body height. For Mongoloids, the Trotter and Gleser (Trotter & Gleser, 1958) and Stevenson (1929) regression formulas have been frequently employed. Stature estimation using length of fingers and hands have been done in other populations such as in Nigeria (Oladipo et al., 2015), Iran (Mahakizadeh et al., 2016), and Korea (Jee and Yun 2015).

Studies on body height estimation have been conducted in different countries, for example in India (Jasuda & Singh, 2004), and Iran (Habib & Kamal, 2010) by measuring separated phalanges. It would be useful to know which one is best to estimate body height - by using separated phalanges or by measuring the entire length of the phalanges. The present study, therefore, focuses on estimating body height based on the entire length of the left hand phalanges of Javanese men. Regression equations specific to Mongoloid Javanese population are computed separately for each finger and to study their reliability in estimating the living stature of the individual.

MATERIALS AND METHODS

The subjects of the present study were 50 Mongoloid Javanese males who consented to be part of the study. Inclusion criteria are: ages between 18 and 25, absence of visible shortening of any of the limbs; absence of bone pathology, bone surgery, and any other pathological process in the medical history that may have influenced the height of the individuals. A purposive sampling method was adopted by taking into consideration the specific characteristics of the subjects (Levy & Lemeshow, 2013). All procedures used in the study are in accordance with the ethical standards set by the Helsinki Declaration and Universitas Airlangga who made sure the research was bound by standards of human rights.

The subjects' body height and the length their left hand phalanges were measured. Height measurements were conducted using standardised anthropometer. The left hand phalanges were measured, since the left hands normally performed fewer activities compared with the right hand, which may cause variation in the length of the right fingers. Measurements were

conducted by sliding callipers. This method for measuring the phalanges is slightly different from that adopted by Jasuda and Singh (2004) on their Indian subjects. They conducted on the segment of the phalanges, while this study measures the entire phalange length from proximal phalanx to the distal phalanx.

Anthropometer was utilised to measure the height and length of the lower legs of the subjects. During the measurement, the subjects stand upright without any footwear.

The length of the fingers was measured using a sliding calliper. The hand was placed slightly open on a flat surface so that the fingers are apart from one another. The sliding calliper was held above the proximal phalanx of index finger while the other end of the needle is shifted to a distal phalanx position to determine the correct length of the finger. The same process applies for measuring other three fingers, except the thumb. The thumb is not included in this measurement because the shape of the thumb is a little bent, unlike the other fingers, and it only has two segments of phalanges.

The normality of numerical data distribution is tested using Kolmogorov-Smirnov test. Data are presented as means with standard deviation. Association of numerical parameters is assessed by Pearson's correlation coefficient r. Whenever r is significant, univariate regression equations are computed according to the equation $Y = ß0 + ß1 * x ± SEE$, where Y is the dependent variable (estimated height), ß0 is *intercept* (constant) and ß1 is *slope* (regression coefficient), x is the independent variable (length of fingers) and SEE is the *standard error of estimation* (Petrovecki, 2007). From the general equation of this linear regression, the formula of body height estimation based on the length of each finger is generated. Body height formula is generated using body height as the dependent variable (Y) and length of the finger as the independent variable (X).

RESULTS AND DISCUSSION

Human body morphology is influenced by hereditary, nutritional, climatic, and racial factors (Katzmarzyk & Leonard, 1998). The length of metacarpals and phalanges can be used for sex determination (El Morsi & Al Hawary, 2013). It means that it is most likely to influence variation between populations too. Studies have concluded that hands vary in different populations. Davies et al. (1980) found significant difference in the length of hand among west European, Indian, and West Indian women. A Nigerian female hand is wider and shorter than that of her foreign counterparts (Okunribido, 2000). This is influenced by the length of fingers.

Of the 50 male subjects, the average body height is 1676 mm (s ± 57.547). The average length of left index, middle, ring, and little fingers of the subject is 82.66 mm (s ± 4.614), 90.98 mm (s ± 4.373), 85.36 mm (s ± 4.814), and 67.56 mm (s ± 4.219) respectively.

The formula of body height based on body parts varies among the different populations (such as Karakas et al. (2011), Didia et al. (2009), and Hasegawa et al. (2009)). The formula for one sex cannot be applied to estimate stature of the other. This study results are only applicable to Mongoloid Javanese males only.

The data showed a significant correlation between height and the length of index, middle, and ring fingers; therefore, the formulas are calculated based on the length of those fingers. There is a moderate correlation between body height and the index finger (r = 0.4). The ß0 is

1268.506 and the ß1 is 4,929, resulting in the body height (BH) estimation formula (in mm) based on the length of index finger as follows:

BH = 1268.506 + 4.929 * Index Finger Length (r^2 = 0.16)

The data shows a moderately significant correlation between body height and the length of the middle finger (r = 0.5). The β_0 is 1085,601 while β_1 is 6.489 which results in the body height estimation formula (in mm) based on the length of the middle finger as follows:

BH = 1085.601 + 6.489 * Middle Finger Length (r^2 = 0.24)

The data shows a moderately significant correlation relationship between body height and the length of the ring finger (r = 0.4). The β_0 is 1274,860 while β_1 is 4,699, resulting in the body height estimation formula (in mm) based on the length of the ring finger as follows:

BH = 1274,860 + 4,699 * Ring Finger Length (r^2 =0,16)

The data, however, show no significant correlation between body height and the length of the little finger. This is in contrast with Jasuda and Singh (2004), who reported that all four fingers have significant correlations with body height. The different result may be due to the differences in the methods. On the other hand, Habib and Kamal (2010) showed a significant correlation between three fingers - index finger, middle finger, and ring finger - and stature. Their study also used similar sampling methods as the present study, examining subjects within the age group 18 to 25 years of age, though their sample size is bigger. Their findings support those of the present study in that, the little finger shows no correlation with body height.

The strength of correlation between body height and length of finger is less than that between body height and segments of bony fingers (2004). This indicates that bony landmarks are more reliable indicators of stature than when the skin and other soft tissue are attached to the fingers.

The study does not measure the thumb length to be correlated with stature. However, another study (Chandra et al., 2016) find a correlation between thumb and stature, so that it will be useful to do such a study for adding formulae in stature estimation.

Other studies report the predictive accuracy of stature estimation is higher for females (Sen et al., 2014). It will be very useful to do a similar study on female samples in the near future, so that it can be useful for forensic cases.

CONCLUSION

In general, based on the measurement of the subject of this study, the longer the finger is, the taller the person is. The greatest correlation with body height is found in the middle finger, followed by the index finger, and the ring finger. The little finger of the subject shows no correlation with body height.

Body height shows a positive and significant correlation with the length of fingers; however, the correlation is moderate when measured from proximal to distal ends. Therefore, it can be concluded that to estimate body height from mutilated hands, the measurement of the segments of fingers, rather than measuring the whole length of the phalanx, gives a better prediction of stature, although the latter method is easier and more practical.

However, this formulae will be useful, because sometimes in Indonesia when a human part is found, the family is grateful that the body is not subjected to further dissection. The formula of the intact finger is very useful, because it respect the feelings of the families.

Future studies should look at a bigger sample size and those below the height examined in this study.

REFERENCES

Chandra, P., Gupta, A., Singh, P., Kulshreshtha, V., Garg, S., & Prakash, G. (2016). A study of correlation between stature and thumb length. *Journal of Anatomy, 24*(1), 47-51.

Cheng, J. C., Leung, S. S. F., Chiu, B. S. K., Tse, P. W. T., Lee, C. W. Y., Chan, A. K. H., & Xu, Y. Y. Y. (1998). Can we predict body height from segmental bone length measurements? A study of 3,647 children. *Journal of Pediatric Orthopaedics, 18*(3), 387-393.

Davies, B. T., Abada, A., Benson, K., Courtney, A., & Minto, I. (1980). A comparison of hand anthropometry of females in three ethnic groups. *Ergonomics, 23*(2),179–182.

Didia, B. C., Nduka, E. C., & Adele, O. (2009). Stature estimation formulae for Nigerians. *Journal of Forensic Sciences, 54*(1), 20-21.

El Morsi, D. A., & Al Hawary, A. A. (2013). Sex determination by the length of metacarpals and phalanges: X-ray study on Egyptian population. *Journal of Forensic and Legal Medicine, 20*(1), 6-13.

Habib, S. R., & Kamal, N. N. (2010). Stature estimation from hand and phalanges lengths of Egyptians. *Journal of Forensic and Legal Medicine, 17*(3), 156-160.

Hasegawa, I., Uenishi, K., Fukunaga, T., Kimura, R., & Osawa, M. (2009). Stature estimation formulae from radiographically determined limb bone length in a modern Japanese population. *Legal Medicine, 11*(6), 260-266.

Iscan, M., & Steyn, M. (2013). *The Human Skeleton in Forensic Medicine*. USA: Charles C Thomas Publisher.

Jasuda, O. P., & Singh, G. (2004). Estimation of stature from hand and phalange length. *Journal of Indian Academy of Forensic Medicine, 26*(3), 100-106.

Jee, S. C., & Yun, M. H. (2015). Estimation of stature from diversified hand anthropometric dimensions from Korean population. *Journal of Forensic and Legal Medicine, 35*, 9-14.

Karakas, H. M., Celbis, O., Harma, A., & Alicioglu, B. (2011). Total body height estimation using sacrum height in Anatolian Caucasians: Multidetector computed tomography-based virtual anthropometry. *Skeletal Radiology, 40*(5), 623-630.

Katzmarzyk, P. T., & Leonard, W. R. (1998). Climatic influences on human body size and proportions: Ecological adaptations and secular trends. *American Journal of Physical Anthropology, 106*(4), 483-503.

Levy, P. S., & Lemeshow, S. (2013). *Sampling of Populations: Methods and Applications*. New Jersey: John Wiley & Sons.

Mahakizadeh, S., Moghani-Ghoroghi, F., Moshkdanian, G., Mokhtari, T., & Hassanzadeh, G. (2016). The determination of correlation between stature and upper limb and hand measurements in Iranian adults. *Forensic Science International, 260*, 27-30.

Okunribido, O. O. (2000). A survey of hand anthropometry of female rural farm workers in Ibadan, western Nigeria. *Ergonomics, 43*(2), 282–92.

Oladipo, G. S., Ezi, G., Okoh, P. D., & Abidoye, A. O. (2015). Index and ring finger lengths and their correlation with stature in a Nigerian population. *Annals of Bioanthropology, 3*(1), 18-21.

Özaslan, A., İşcan, M. Y., Özaslan, I., Tuğcu, H., & Koç, S. (2003). Estimation of stature from body parts. *Forensic Science International, 132*(1), 40-45.

Petrovecki, V., Mayer, D., Slaus, M., Strinovic, D., & Skavic, J. (2007). Prediction of stature based on radiographic measurements of cadaver long bones: A study of the Croatian population. *Journal of Forensic Sciences, 52*(3), 547-552.

Sen, J., Kanchan, T., Ghosh, A., Mondal, N., & Krishan, K. (2014). Estimation of stature from lengths of index and ring fingers in a North-eastern Indian population. *Journal of Forensic and Legal Medicine, 22*, 10-15.

Stevenson, P. H. (1929). On racial differences in stature long bone regression formulae, with special reference to stature reconstruction formulae for the Chinese. *Biometrika, 21*(1/4), 303-321.

Trotter, M., & Gleser, G. (1958). A re-evaluation of estimation of stature based on measurements of stature taken during life and long bones after death. *American Journal of Physical Anthropology, 16*(1), 79–123.

SCIENCE & TECHNOLOGY

The Association of FDG-PET (Suvmax) and Inflammatory Marker in Predicting Tumour Aggressiveness

Ahmad Saad, F. F.*, Abdullah, N. S., Shaharuddin, S. and Nordin, A. J.

Centre for Diagnostic Nuclear Imaging, Faculty of Medicine, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Chronic inflammation is associated with processes that contribute to the onset or progression of cancer. This study examined the correlation between dichotomised patients with malignant tumours and inflammatory markers based on the altered glucose metabolism measured by the FDG SUVmax that underpins the degree of tumour aggressiveness. Thirty-one patients underwent ^{18}F-FDG PET/CT for various carcinoma along with blood inflammatory markers such as C-reactive protein (CRP), interleukin-6 (IL6), lipid profile and fasting blood glucose (FBG) levels were obtained in retrospective study. Patients were dichotomised by the cut-off SUVmax value of 6.0 dl/ml derived from curve analysis (P=-0.025). The mean age of the subjects were 53.16 ± 12.06 years and mean SUVmax of 8.80 ± 6.27 g/ml. Significant correlation was noted between the SUVmax and CRP and IL6 (r=0.361; P<0.05) and IL-6 with BMI and FBG with r=0.38; p<0.05 and r=0.34; p<0.05 respectively. The odds ratio (95% confidence intervals) for patients with the SUVmax cut-off 6.0dl/ml was predicted by FBG (OR:0.385, p<0.05) and the SUVmax > 6.0 dl/ml was inversely related to IL-6 (OR: 0.049, with p<0.005). Serum inflammatory markers and endogenous glucose are associated with a potentially more aggressive malignant cancer. In particular, IL6 may be used as a useful surrogate marker for tumour aggressiveness with an important prognostic value.

Keywords: Inflammation, Computed tomography, positron emission tomography, [18F]-fluorodeoxyglucose, multimodality

INTRODUCTION

Development of cancer is significantly higher among patients with diabetes. In this regard, infection or chronic inflammation are associated with mutagenesis in 15-20% of all cancers among which are breast, colorectal and bronchus followed by cancer of liver, stomach and lung (Kuper, 2000; Balkwill, 2001; World Cancer Report, 2008).

Inflammatory factors, such as interleukin-6 (IL-6), plasminogen activator, inhibitor-1 (PAI-1), free fatty acids, monocyte chemo-attractant protein, leptin, fatty acids, adiponectin and tumour necrosis factor-α are produced by adipose tissue (Van Kruijsdijk et al., 2009). These factors are important in the progression of malignant cancer. In particular, the knowledge of IL-6 for immune homeostasis and how it induces profound activities in acute phase reaction has rapidly increased in chronic inflammation and carcinogenesis. The progression of cancer resulting from chronic inflammation is well known (Naugler et al., 2008). IL-6 is a commander marker in promoting inflammation to inflammation-associated cancer (Tangkijvanich et al., 2004). In most cases, the level of IL-6 is elevated which favours the development of carcinogenesis.

In recent decades, multimodality imaging was used to detect early progression of cancer, such as MRI, Single-Photon Emission Computed Tomography(SPECT) and Positron Emission Tomography PET-CT (Townsend et al., 2008). The PET-CT uses small amount of radiotracer to evaluate tissue functions, early onset of disease and lesion localisation via imaging sessions (Beyer et al., 2004; Nakamoto et al., 2003; An et al., 2007). The system complements each other to detect the lesion and hence help improve diagnostic yield by offering substantial advantages over anatomic and physiologic imaging techniques, unlike the standalone PET or CT system.

The study examined potential association of chronic inflammation and the molecular marker – 18 F FDG –PET-CT which contributes to the transformation and early onset of cancer.

MATERIALS AND METHODS

Patients

Thirty-one patients with lung carcinoma (11), oesophageal carcinoma (7), metastatic paraganglioma (3), colonic carcinoma (3), transitional cell carcinoma (3) and one each with lymphoma, hepatoma, nasopharyngeal carcinoma and breast carcinoma were investigated in a retrospective study. All patients underwent [18F] fluorodeoxyglucose (FDG)-positron emission tomography (PET)/computed tomography (CT) for various indications of disease progression.

Patient Preparation

All patients fasted for at least 6 hours prior to scanning session and only oral hydration with glucose-free water was allowed. Fasting blood glucose was recorded for all patients. Oral gastrografin solution (sodium meglumine diatrizoate; BerliMed S.A., Madrid, Spain) was given to patients in dilution in three parts before IV administration of radiopharmaceutical agent and immediately before scanning. Becquerel range from 290.5 – 415.5 mega (Mean 347.5 MBq) of 18F-FDG was injected intravenously. All patients were placed in a special room ranging from 23 – 190, for an average of mean 70.6 minutes before they were asked to empty their bladder prior to the PET/CT scan imaging study.

PET/CT Imaging Protocol

Studies were standardised using a dedicated integrated PET-CT system (Siemens Biograph–64, Germany). This device comprises a dedicated PET scanner with Optimum Performance in 3-D Imaging with Lutetium Oxyorthosilicate (LSO) scintillator crystal technology. The system is incorporated with a multislice CT scanner with capability for 64 slice CT and high spatial resolution. For the purpose for attenuation correction, a scout view was performed in cranio-caudal direction followed by low dose CT protocol in caudo-cranial. Contrast-enhanced CT (CECT) protocol, iohexol (Omnipaque 350 mgI/mL, GE Healthcare, Shanghai, China) ranging from 50 – 100 ml with mean 83.07 ml was injected intravenous by using dual head automatic pressure injector (Mallinckrodt, M.O, and USA) with flow rate at 2.5 ml per second and followed by 20 ml saline flush. At the start of the CT scan, CECT acquisition started in caudo-cranial direction with 80 seconds delay, ensuring optimised intravenous (IV) contrast in the circulation and tissue enhancement. Subsequently, in view of the higher sensitivity of the PET scan, the acquisition time for PET was 2 min per table position.

Figure 1. A circular region of interest (ROI) is drawn to evaluate the SUVmax value

Image Interpretation

PET/CT scans were read by an experienced radiologist with more than 5 years of clinical experience. The radiologist was blinded to the diagnosis and he/she was unaware of any biochemical findings or clinical information. The circular region of interest was drawn at the FDG-avid area (Figure 1) and the SUVmax value was automatically calculated.

Statistical Analysis

All patients with biopsy-proven cancer were dichotomised into groups based on the cut-off SUVmax value of 6.0 derived from the ROC statistical analysis (Figure 2-3). Paired t-test (two-tailed) and Wilcoxon signed rank test were used for parameters, which were normally

distributed and not normally distributed. Results were evaluated at a 95% confidence interval and the level of significance was set as $p<0.05$. Data was analysed using the Pearson correlation coefficient and binary logistic regression. All statistical tests were two-sided, and p values <0.05 were considered statistically significant. Data was analysed using Statistical Package for Social Sciences program (SPSS 21) (IBM Corp, Somers, New York).

RESULTS AND DISCUSSION

Patient Characteristics

The study investigated 31 retrospective patients (17 males) with mean age of 53.16±12.06 years. These patients who underwent FDG PET-CT l had biopsy-confirmed lung carcinoma (11), oesophageal carcinoma (7), metastatic paraganglioma (3), colonic carcinoma (3), transitional cell carcinoma (3) and one each with lymphoma, hepatoma, nasopharyngeal carcinoma and breast carcinoma (Table 1). Among them, there were subjects having NIDDM (10), Hypertension (16), Hypercholesterolemia (6) & history of smoking (8) with mean BMI of 23.56±5.31kg/m2 and FBS 5.30 ± 1.33 g/dl. The mean SUVmax values measured for the target lesions was 8.80±6.27 (Table 1).

Table 1
Patient characteristics

Patient	Age	Sex	BMI	Histology	Types	FBG	CRP	IL_6	Cholst	Trigl	HDL	LDL	SUVmax
P1	57	F	22	Left lung adenoca	Malignant	5.0	122.22	5.39	3.06	1.03	0.49	1.65	7.90
P2	59	F	19	Carcinoma esophagus	Malignant	5.0	0.98	6.08	4.27	0.78	1.06	2.56	13.51
P3	33	F	20	Carinoma of esophagus	Malignant	5.3	0.63	6.06	4.79	1.31	1.05	3.01	4.08
P4	59	F	21	Adenoca esophagus	Malignant	3.5	2.04	6.22	3.81	1.17	1.08	1.79	12.60
P5	50	M	22	Adenoca esophagus	Malignant	5.8	0.46	5.94	4.10	0.66	2.07	1.73	5.00
P6	68	F	21	Adenoca esophagus	Malignant	3.8	14.59	4.52	5.57	1.52	1.01	3.38	22.64
P7	45	M	29	Melanoma	Malignant	5.0	2.22	6.00	6.34	1.10	1.18	4.49	17.66
P8	47	F	23	Adenoca esophagus	Malignant	4.3	21.31	5.91	4.04	1.27	0.68	2.60	24.32
P9	56	F	22	Left adrenal phaeo	Benign	7.1	0.90	6.12	4.61	2.13	0.93	2.83	4.68
P10	59	F	19	Left lung adeno	Malignant	4.3	1.69	6.09	2.74	0.58	0.82	1.76	4.11
P11	60	M	19	Clear cell adenoca	Malignant	3.6	2.52	6.18	5.15	1.12	1.88	3.01	6.71
P12	42	F	29	Hepatocellular Ca	Malignant	5.3	10.05	5.87	3.50	0.97	0.68	2.63	4.16
P13	64	F	27	Rt chest wall tumour	Malignant	6.4	14.94	5.52	2.68	0.94	1.04	1.51	6.15
P14	38	M	33	Lt infraclavicular tumour	Malignant	6.7	0.32	6.13	3.86	1.11	0.79	2.85	3.02
P15	70	M	27	Rectal ca with	Malignant	6.0	5.65	6.26	3.28	0.97	1.12	2.42	4.50

Table 1 *(continue)*

Patient	Age	Sex	BMI	Histology	Types	FBG	CRP	IL_6	Cholst	Trigl	HDL	LDL	SUVmax
P16	57	F	26	Adrenal tumour	Benign	6.6	1.30	6.13	2.55	1.80	0.54	2.07	5.83
P17	61	F	16	Lung cancer	Malignant	4.7	184.41	5.36	3.01	0.74	0.62	2.54	10.40
P18	62	M	24	Primary clear cell Ca	Malignant	8.0	217.37	6.21	1.26	0.89	0.34	0.57	5.00
P19	63	M	20	Ca pancreas	Malignant	5.6	0.73	6.13	3.07	1.06	0.86	2.49	4.40
P20	67	M	22	Ca caecum with	Malignant	5.3	13.03	5.77	4.02	1`.48	0.96	3.22	11.40
P21	47	F	20	Adenoca esophagus	Malignant	4.4	0.52	6.25	3.15	0.99	0.79	2.57	3.64
P22	59	F	21	Adenoca esophagus	Malignant	4.3	1.45	5.26	3.94	1.68	0.86	3.33	3.04
P23	33	M	14	Stage iv NPC	Malignant	3.7	11.26	5.82	3.09	1.74	0.67	2.42	9.05
P24	32	F	29	Ca lung	Malignant	5.0	2.25	6.05	3.11	2.22	0.85	2.25	2.81
P25	58	M	26	Stage IV ca colon	Malignant	5.3	6.90	6.15	3.94	0.63	1.02	3.30	4.70
P26	18	M	23	Adrenal tumour	Benign	5.0	186.44	4.78	2.91	0.97	0.56	2.38	20.30
P27	44	M	22	Ca breast	Malignant	5.0	15.85	6.17	6.05	0.97	0.90	2.50	13.00
P28	62	M	13	Recurrent gastroc Ca	Malignant	2.8	0.51	4.94	3.19	1.91	0.92	2.23	6.12
P29	61	M	34	Recurrent ca cecum	Malignant	5.6	17.39	5.88	3.10	1.80	0.55	2.47	20.12
P30	35	F	27	Ca esophagus	Malignant	3.8	3.70	6.00	2.95	0.93	0.59	2.37	4.08
P31	61	F	34	Ca lung	Malignant	5.0	1.40	5.99	3.47	1.32	0.77	3.01	9.86

Notes: FBG: fasting blood glucose, CRP- C-reactive protein, Cholst: cholesterol, Trigl: triglyceride, HDL-high density lipoprotein, LDL: low density lipoprotein, SUVmax: PET-Ct marker (standardised uptake values)

Biochemical analysis for inflammatory markers revealed mean CRP of 25.59±53.25U/I, IL-6 5.87± 0.47U/I with Triglyceride 1.22± 0.52 U/I, HDL.89 ± 0.32 mg/dl and LDL 2.59 ±0.89 mg/dl (Table 2).

Table 2
Mean biochemical results

Characteristic	Mean ± SD
BMI (kg/m^2)	23.56±5.31
FBS (g/dl)	5.30 ± 1.33
CRP (U/l)	25.59±53.25
IL-6 (U/l)	5.87± 0.47
Tg (U/l)	1.22± 0.52
HDL(mg/dl)	0.89 ± 0.32
LDL(mg/dl)	2.59 ±0.89

Stratification of Patient by the Index of Tumour Aggressiveness (SUVmax) on FDG PET

The patients were divided into two groups based on the SUVmax cut-off. Using the Receiver operating curve (Figure 2), the cut-off SUVmax of 6.0 was found significant for which (P< 0.005).

ROC Curve

Diagonal segments are produced by ties.
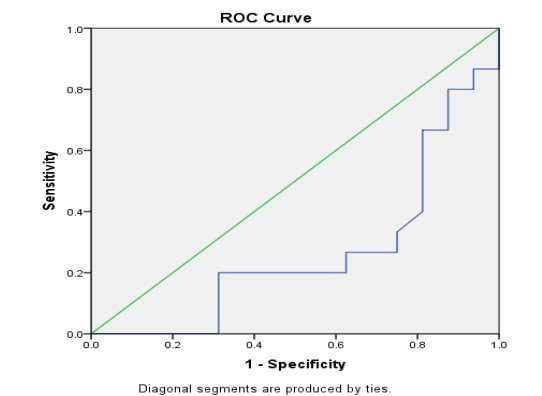
Area Under the Curve (AUC)

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.265	.092	.026	.084	.445

[a]Under the nonparametric assumption
[b]Null hypothesis: true area = 0.5

Figure 2. ROC curve with cut-off value of 6.0

(a)   (b)
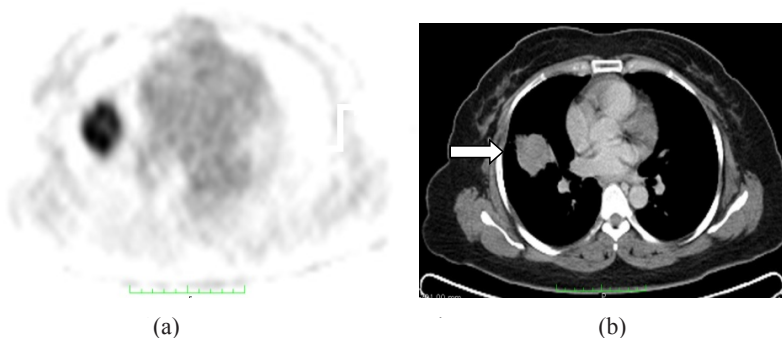
Figure 3. P31: (a) Axial FDG-PET and (b) CT images of a 61 years –old-male ex-smoker with high CRP diagnosed with lung carcinoma showing an FDG-avid mass in the right lung (arrowed) – SUVmax of 10.40
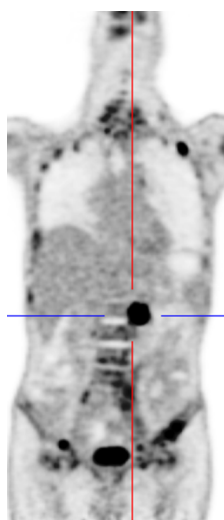
Figure 4. (P17); 18-year-old-male with NIDDM and abdominal paraganglioma (hairline marker) (SUVmax: 15.79)

Correlation Between Inflammatory Markers and SUVmax

Based on spearmen correlation (Table 3), there was strong correlation noted between the CRP and SUVmax (P<0.005); r: 0.527). The CRP was also strongly correlated with IL-6 (r=0.361; P<0.05). IL-6 was strongly correlated with the BMI (r: 0.38) and FBG (r: 0.34) with (P<0.05) respectively.

Table 3
Spearman Rhos to assess correlation between baseline inflammatory biomarkers and clinical marker and the SUVmax

Interleukin-6		P value
	Rho	
Age	0.23	0.88
BMI	0.38	*0.01
FBG	0.34	*0.03
CRP		p-value
	Rho	
SUVmax	0.527	*0.005
IL-6	0.361	*0.05

*statistical significant P<0.05

Predictor for the Carcinogenesis (FBS & IL-6)

The Figure 5 shows there is a one-unit increase in the FBS for patients with SUVmax of more than 6.0 are 0.38 times the odds of default for patients with SUVmax < 6.0. Conversely a 1-unit decrease in IL6 for patients with SUVmax more than 6 is associated with 0.049 times the odd of default for patient with lower SUVmax in this category.

Subgroup	OR	95% CI	^{P}heterogeneity
FBG	0.385	0.016-0.925	0.003
IL-6	0.049	0.003-0.948	0.048

Risk OR(95%) of malignant tumour for IL- 6 and FBG stratified by SUVmax (6).
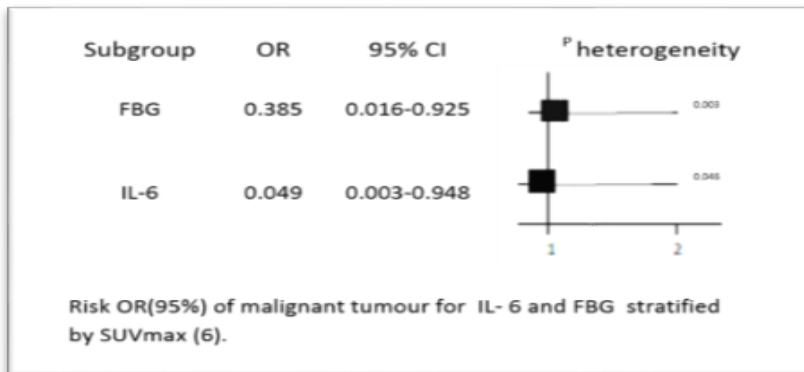
Figure 5. The odd ratios (95% confidence intervals) for patients with SUVmax cut-off of 6

The primary interest of this study was to evaluate the potential association of the inflammatory markers and the altered glucose metabolism in carcinogenic environment. The basis of inflammation in carcinogenesis has been well recognised especially in several known cancer i.e. endometrial cancer (Key et al.,1988). This study evaluated the imaging marker exploiting glucose analog (FDG) as a signal for altered glucose metabolism in different types of cancer and potential association with common inflammatory markers. The degree of altered glucose metabolism was measured using the standardised uptake value (SUVmax) of the PET image evaluated using PET-CT.

The higher value of SUVmax is correlated with the cellular proliferation and poor prognosis (Fathinul et al., 2015). Traditionally, the cut off value for SUVmax 2.5 was used to distinguished malignancy from benignity (Fathinul et al., 2015; Fathinul et al., 2014). Given all subjects were harbouring malignant cellular entity, the SUVmax of 6.0 was utilised to dichotomise groups of cancer patients for which rationale of higher SUVmax value raises the potential of cellular stress that underpins the carcinogenic reprogramming. Diabetes mellitus (with fasting blood sugar > 7.0 mmol/L) for instance is associated with increased tumour burden (Kaaks et al., 2002). The study found an association between patient with high glucose level and the altered glucose metabolism (SUVmax) which form the basis of strong signalling of the cancer-prone environment. The basis of this can be explained by the fact that isoform of the insulin receptor expression will induce cancer cells to produce insulin and IGF-I. The A receptor isoform can stimulate insulin-mediated mitogenesis and hence, the development of carcinogenic cellularity (Denley et al., 2007). There is also a strong correlation between SUVmax and the serum C-reactive protein among the cancers subjects. This is in line with other reports which found inflammatory process a potential precursor for cancer development (Love et al., 2005).

The potential correlation of clinical parameter of patients with evidence of chronic inflammation as evidenced by high glycaemic index evaluated by biomarkers i.e. FBS, level of total cholesterol and high BMI was investigated. The results showed obesity (high BMI) and NIDDM (FBG >7.0 mmol/L) was well correlated with the IL-6. These supported other reports which suggested elevated serum level of IL-6 in patient with insulin resistance and obesity and correlated with BMI (Kern et al., 2001). This shows cell stress and the alteration in cellular reprogramming for the progression of carcinogenesis.

Lower level of IL-6 in malignant cells of this study was an independent predictor for the altered glucose metabolism in malignancy. The basis of low level of IL-6 in association with cancer is poorly explained. One of the potential lower IL-6 levels has been described in younger age group but will rise with aging (Sarkar et al., 2006). Nevertheless, there was no association between low IL-6 with age in this study. In mouse model of colitis-associated cancer (CAC), trans-signalling was inhibited by tumour growth factor (TGF-B) to suppress the early onset or formation of cancer, which leads to low level of IL-6 in human sample of colon cancer (Atreya et al., 2000). The high level of IL-6 explains the carcinogenesis via classical IL-6 signalling. This phenomenon is well described in multiple myeloma patients (Lattanzio et al., 1997), which show that the high level of IL-6 is associated with vice-versa. Another study found high levels of IL-6 expression was correlated with high rates of myeloma protein secretion, low proliferative compartment, and low tumour mass. IL-6 induces an undifferentiated tumour cell compartment into terminal differentiation (high rate of immunoglobulin production, low proliferative fraction) and eventual tumour cell death (low tumour mass (Kishimoto,1989). Therefore, the lower IL-6 level in an altered glucose metabolism of mitogenesis can be explained by the probable the predominant expression of TGF-B which could have suppressed the trans-signalling cancer reprogramming pathway in carcinogenesis.

Data was obtained from outpatients and inpatients (a large nationally representative sample). The diagnoses were caught blindly and confirmed by the biopsy findings. Limitations included a relatively small number of subjects and a lack of other parameters, such as genetics, hormones, biochemical, lifestyle and diet. Additionally, the duration for follow-up the onset of clinical parameter and biological changes leading to cancer is probably too short.

CONCLUSION

Serum inflammatory markers and endogenous glucose are associated with a potentially more aggressive malignant cancer. In particular, IL-6 may be used as a useful surrogate marker for tumour aggressiveness with an important prognostic value. The low-level IL-6 indicates the cellular reprograming of cancer development has been signalled via different pathway of the IL-6 transduction in inflammation-associated cancers and hence, requires further understanding of various cancers types on the potential mechanism that regulates the cellular reprogramming. The understating of varied levels of IL-6 associated carcinogenesis is important and can be used as a marker for treatment efficacy.

REFERENCES

An, Y. S., Sheen, S. S., Oh, Y. J., Hwang, S. C., & Yoon, J. K. (2007). Nonionic intravenous contrast agent does not cause clinically significant artifacts to 18F-FDG PET/CT in patients with lung cancer. *Annual Nuclear Medicine, 21*(10), 585-92.

Atreya, R., Mudter, J., Finotto, S., Müllberg, J., Jostock, T., Wirtz, S., ... & Strand, D. (2000). Blockade of interleukin -6 trans signalling supresses T-cell resistance against apoptosis in chronic intestinal inflammation: evidence in Crohn disease and experimental colitis in vivo. *Nature Medicine, 6*(5), 583-588.

Balkwill, F., & Mantovani, A. (2001). Inflammation and cancer: Back to Virchow? *The Lancet, 357*(9255), 539-545.

Beyer, T., Antoch, G., Muller, S., Egelhof, T., & Freudenberg, L. S. (2004). Acquisition protocol considerations for combined PET/CT imaging. *Nuclear Medicine, 45*, 25-35.

Denley, A., Carroll, J. M., Brierley, G. V., Cosgrove, L., & Wallace, J. (2007). Differential activation of insulin receptor substrates 1 and 2 by insulin-like growth factor-activated insulin receptors. *Molecular Cell Biology, 27*(10), 3569–3577.

Fathinul, F. A. S., Nordin, A. J., & Immeran, M. Z. (2015). Application of PET-CT in oncology. *Pertanika Journal of Science and Technology, 23*(1), 1–12.

Fikri, A. S. F., Nordin, A. J., & Lau, W. F. E. (2014). 18(F) FDG-PET/CT is a useful molecular marker in evaluating tumour aggressiveness; A revised understanding of an in-vivo FDGPET imaging that alludes the alteration of cancer biology. *Cell Biochemistry and Biophysic, 55*(5), 631-640.

Kaaks, R., Lukanova, A., & Kurer, M. S. (2002). Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiology, Biomarkers and Prevention, 11*(12), 1531-1534.

Kern, P. A., Ranganathan, S., Li, C., Wood, L., & Ranganathan, G. (2001). Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *American Journal of Physiology-Endocrinology and Metabolism, 280*(5), E745-E751.

Key, T. J., & Pike, M. C. (1988). The dose-effect relationship "unopposed" oestrogens and endometrial mitotic rate: Its central role in explaining and predicting endometrial cancer risk. *British Journal of Cancer, 57*(2), 205-212.

Kinahan, P. E., Townsend, D. W., Beyer, T., & Sashin, D. (1998). Attenuation correction for a combined 3D PET/CT scanner. *Medical Physics, 25*(10), 2046-53.

Kishimoto, T. (1989). The biology of Interleukin-6. *The Journal of The American Society of Hematology, 74*(1), 1-10.

Kuper, H. (2000). Infection as a major preventable cause of human cancer. *Internal Medicine, 248*(3), 171-183.

Lattanzio, G., Libert, C., Aquilina, M., Cappelletti, M., Ciliberto, G., Musiani, P., & Poli, V. (1997). Defective development of pristane-oil-induced plasmacytomas in interleukin-6-deficient BALB/c mice. *The American Journal of Pathology, 151*(3), 689-696.

Love, C., Tomas, M. B., Tronco, G. G., & Palestro, C. J. (2005). FDG PET of infection and inflammation. *Radiographics, 25*(5), 1357–1368.

Nakamoto, Y., Chin, B. B., Kraitchman, D. L., Lawler, L. P., & Marshall, L. T. (2003). Effects of nonionic intravenous contrast agents at PET/CT imaging: Phantom and canine studies. *Radiology, 227*(3), 817-824.

Naugler, W. E., & Karin, M. (2008). The wolf in sheep's clothing: the role of interleukin-6 in immunity, inflammation and cancer. *Trends Molecular Medicine, 14*(3), 109-119.

Sarkar, D., & Fisher, P. B. (2006). Molecular mechanism of aging associated inflammation. *Cancer Lett, 236*(1), 13-23.

Tangkijvanich, P., Thong-ngam, D., Theamboonlers, A., Hanvivatvong, O., Kullavanijaya, P., & Poovorawan, Y. (2004). Diagnostic role of serum interleukin 6 and CA 19-9 in patients with cholangiocarcinoma. *Hepato-gastroenterology, 51*(55), 15-19.

Townsend, D. W. (2008). Positron emission tomography/computed tomography. *Seminars Nuclear Medicine, 38*(3), 152-66.

Van Kruijsdijk, R. C., van der Wall, E., & Visseren, F. (2009). Obesity and cancer: The role of dysfunctional adipose tissue. *Cancer Epidemiology Biomarkers Prevention, 18*, 2569–2578

World Cancer Report. (2008). *Boyle P, Bernard L Eds. Cedex, France, World Health Organization, International Agency for Research on Cancer.* Retrieved April 1, 2010, from http://www.iarc.fr/en/publications/pdfs-online/wcr/index.php.

Comparative Analysis of Contrast Enhancement Techniques for Medical Images

Randeep Kaur[1], Meenu Chawla[1], Navdeep Kaur Khiva[1] and Mohd Dilshad Ansari[2]*

[1]*Department of Computer Science and Engineering,*
Giani Zail Singh Campus Colleger of Engineering and Technology, Bathinda, Punjab151001, India
[2]*Department of Computer Science & Engineering, Jaypee University of Information Technology,*
Waknaghat, Solan 173234, India

ABSTRACT

Contrast enhancement is the focus of this paper namely use of digital for medical imaging. Five types of images were analysed, namely hand, brain, head, ankle and knee. Three techniques have been used such as INT Operator, Fuzzy Type-1 and Fuzzy Type-2 on five different images. The obtained results have been compared based on four quality parameters, namely mean square error (MSE), normalisation coefficient (NC), root mean square error (RMSE) and peak signal to noise ratio (PSNR).Results showed INT Operator provides the best resultant image compared with other techniques.

Keywords: Contrast enhancement, INT operator, Fuzzy Type-1, Fuzzy Type-2

INTRODUCTION

A digital image is a numeric picture of a 2-dimensional image and compises rows and columns. Each block is named pixels. Digital image plays a vital role in medical imaging.

Medical imaging technique generates visual structures of the internal organs. It shows the function of some organs or tissues. The important factor in any subjective evaluation of image quality is contrast enhancement. Contrast enhancement is mainly the difference in visual effects that makes an object separate from other objects and their background. It also improves the quality or clarity of images and increases the interpretability in images for human viewing.

Current Affilition:
Mohd Dilshad Ansari
Department of Computer Science and Engineering, CMR College of Engineering & Technology, Kandlakoya, Hyderabad 501401, India.

Contrast enhancement plays an important role in medical field or imaging. Contrast enhancement helps us to increase the brightness of an image. When some medical images have low contrast, contrast enhancement is the best method to enhance the quality of image. Three enhancement techniques have been used in this paper. The INT is an intensification operator to decrease the fuzziness of an image. Fuzziness means dullness of the pixels. This technique contains some parameters to analyse the enhancement of digital images (Mahashwari & Asthana, 2013).

Fuzzy type-1 and fuzzy type-2 techniques show good contrast of an image. The result of type-1 fuzzy logic system is represented by single numeric values. It means only one membership function is performed using this method. Membership values are computed with the help of fuzzy hyperbolisation approach. This approach has three steps, namely fuzzification, membership modification and defuzzification (Hartati et al., 2009). Fuzzification converts gray level into membership values. The new gray levels are generated by the defuzzification step. It has four components: fuzzifier, rulebase, inference engine and defuzzifier and work based on membership function (Kaur et al., 2017; Wu, 2014; Pal et al., 1983; Pal et al., 1981).

Fuzzy type-2 approach is a special case of fuzzy type-1 set. This set overcomes the limitations of type-1 fuzzy set. Fuzzy hyperbolisation technique is also applied in this approach for computing new gray levels. There are four components: fuzzifier, rulebase, inference engine and output processor. Rulebase component contains IF-THEN rules of the domain. type-2 fuzzy sets represent uncertainty based on two membership functions, namely lower membership function and upper membership function (Castillo et al., 2012; Castillo et al., 2007; Ensafi et al., 2005).

Four quality parameters such as Mean Square Error (MSE), Normalisation Coefficient (NC), Root Mean Square Error (RMSE) and Peak Signal to Noise Ratio (PSNR) are used for comparing contrast enhancement techniques. The MSE and RMSE stands for mean square error and root mean square error respectively while PSNR and NC means peak signal to noise ratio and normalisation coefficient respectively.

The paper is structured as follows. Section II contains literature review while Section III briefly discusses contrast enhancement techniques i.e. INT Operator, Type-1 Fuzzy & Type-2 fuzzy. Section IV describes performance analysis and simulation results by considering different performance quality parameters. Section V concludes the paper and discusses the future scope of work.

BACKGROUND

Kundra et al. (2009) presented an image enhancement based on fuzzy logic. The main goal of this paper is to remove the noise and improve the contrast of an image using digital imaging. There two main steps: removal of impulse noise. This begins with gray scale image before applying the filter. The second step is to improve contrast of the image. According to this step, set shape of membership function and the value of fuzzifier beta according to the actual image is computed based on membership values. Further, membership values are changed using linguistic values to generate new gray-levels.

Hassanien et al. (2011) proposed the contrast for breast MRI images. Fuzzy type-2 technique is applied on different MRI breast images. This technique provides a higher accuracy compared with other techniques. The results of the fuzzy type-2 are compared with type-1 fuzzy technique. The type-2 Fuzzy approach provides better results compare to Fuzzy type-1 technique.

Tizhoush and Fochem (1995) has developed a hybrid technique for image contrast enhancement using fuzzy histogram hyperbolisation approach. This method provides better results for image enhancement. In this method they have taken X-ray image and satellite image for their result analysis. The main idea of this approach is to improve the contrast of the input images.

Preethi et al. (2013) has proposed the function modification using fuzzy logic. This paper represents the function modification using fuzzy logic and membership functions are modified to enhance finger prints. They have taken medical images for result analysis. This algorithm is also used to enhance the video images. Square and cube operator are applied for modifying new membership functions.

Lakshmi et al. (2013) have proposed an image contrast enhancement using the fuzzy technique. The existing algorithms manage uncertainties. The algorithm is used to calculate the parameters and is also applied on different types of images. The fuzzy technique is found to be a better technique compared with other techniques for contrast enhancement.

Mahashwari and Asthana (2013) proposed image processing theory based on a fuzzy technique using three steps, such as fuzzification, membership value and defuzzification. The proposed method was able to enhance the quality of image successfully.

Sesadri and Nagaraju (2015) proposed type 2 fuzzy technique for image enhancement. First, fisher criterion function was used to generate membership values of type-I fuzzy. Further, fuzzy rules are applied to generate enhanced image.

Khandewal and Kaur (2016) provides comparative study of different image enhancement technique. In this paper, six enhancement techniques were used to improve information in images. Erosion technique has produced best result with highest value PSNR and lower MSE value.

Kaur and Kaur (2016) compared enhancement techniques for medical images. In this paper, five enhancement techniques were used, namely average filter, bilateral ratinex, neighbourhood operation, imadjust and sigmoid function. Sigmoid function and neighbourhood operation produced the best result with low value of MSE & RMSE and high value of PSNR.

CONTRAST ENHANCEMENT TECHNIQUES

Three contrast enhancement techniques are used in this paper.

INT Operator

The INT is an intensification operator. This operator is used to decrease the fuzziness of an image. Fuzziness means dullness of the pixels. This technique contains some parameters to

analyse enhancement of the image. Three steps are used to analysis this approach (Mahashwari & Asthana, 2013).

Step 1: Defining membership functions

$$\mu_{mn} = G(g_{mn}) = \left[1 + \frac{g_{max} - g_{mn}}{F_d}\right]^{-F_e}$$

Where
g_{mn} = Intensity value of pixel
g_{max} = maximum gray level
F_d = Denomination fuzzifier
F_e = Exponential fuzzifier

Denomination fuzzifier helps to increase the value in range [0, 1] and decrease fuzzified value.

Step 2: Membership modification

$$\mu'_{mn} = \begin{cases} 2 \cdot [\mu_{mn}]^2 & 0 \leq \mu_{mn} \leq 0.5 \\ 1 - 2 \cdot [1 - \mu_{mn}]^2 & 0.5 \leq \mu_{mn} \leq 1 \end{cases}$$

INT Operator modifies the membership values on a fuzzy set μ_{mn}. Cross over point is 0.5. The intensification operator steadily increases the value of μ_{mn} for increasing the values of cross over point from 0 to 1. This operator is applied on image again and again.

Step 3: Generate new gray levels

$$g'_{mn} = G^{-1}(\mu'_{mn}) = g_{max} - F_d\left((\mu'_{mn})^{\frac{-1}{F_e}} - 1\right)$$

The contrast depends on the increasing value of cross over point.

Fuzzy Type-1

The result of type-1 fuzzy logic system is represented by single numeric values. It means only one membership function is performed in this method. Membership values are computed with the help of fuzzy hyperbolisation approach. This approach has three steps, namely fuzzification, membership modification and defuzzification (Hartati et al., 2009). Fuzzification transforms gray level into membership values. The value of membership functions is changed after the fuzzification process. The new gray levels have generated in defuzzification step. Four components are used for processing type-1 fuzzy set, namely fuzzifier, rulebase, inference engine and defuzzifier. Fuzzifier transforms crisp value into fuzzy value. Rulebase component contains rules of the domain. Inference engine performs actions for fuzzy control. Defuzzifier transforms fuzzy value into crisp output value (Hassanien et al., 2011).

Step 1: Compute the type-1 fuzzy membership value using following equation:

$$\mu(g_{mn}) \ = \ \frac{g_{mn} - g_{min}}{g_{max} - g_{min}}$$

<div align="right">(1)</div>

Where g_{mn} = Intensity value

g_{min} = Minimum gray level

g_{max} = Maximum gray level

Step 2: Compute the new gray levels

$$\tilde{g}_{mn} \ = \ \frac{L-1}{e^{-1}-1} \times \left[e^{-\mu(g_{mn})^\beta} - 1 \right]$$

<div align="right">(2)</div>

Where the parameter β is set to 0.8 and L is a number of gray levels.

Fuzzy Type-2

Fuzzy type-2 approach is the special case of type-1 fuzzy set. It overcomes the limitations of type-1 fuzzy set. Fuzzy hyperbolisation technique is also applied in this approach for computing new gray levels (Hartati et al., 2009). Fuzzification transforms gray level into membership values which are changed during the fuzzification process.

The new gray levels have been generated by the defuzzification step. Type-2 fuzzy sets represent uncertainty. Four components are used for processing, namely fuzzifier, rulebase, inference engine and output processor. Fuzzifier transforms crisp value into fuzzy value. Rulebase component contains IF-THEN rules of the domain. Inference engine performs actions for fuzzy control. Output processor contains type reducer and defuzzifier. Type-1 fuzzy set output and crisp number are generated from type reducer and defuzzifier respectively. Two membership functions are performed in this fuzzy set, such as upper membership function and lower membership function (Hassanien et al., 2011).

Step 1: Compute the upper and lower membership values by using the following equation:

$$\mu_{Lower}(x) \ = \ \mu(x)^2$$

and

$$\mu_{Upper}(x) \ = \ \mu(x)^{0.5}$$

Compute a window of size 21×21.

Determine type-2 fuzzy membership function using following equation

$$\mu_{T_{II}(g_{mn})} \ = \ (\mu_{Lower} \times \alpha) + \left(\mu_{Upper} \times (\alpha - 1) \right)$$

Where $\alpha = \frac{g_{mean}}{L}$

Step 2: Calculate the new gray levels using equation (2)

PERFORMANCE ANALYSIS AND SIMULATION RESULTS

Five types of experiment on given images were conducted and the various performance parameters were calculated to check the robustness of the algorithm.

The following quality parameters are considered:

1) **Mean Square Error:** It measures the average of the squares of the errors.

$$MSE = \sum \frac{\sum(Y - \hat{Y})}{m \times n}$$

Where Y= actual value, $\hat{Y}$ = predicted value, m= actual size and n= predicted size

2) **Root Mean Square Error:** It measures the difference between two similar images.

$$RMSE = \sqrt{MSE}$$

3) **Peak Signal to Noise Ratio:** It is a ratio between the maximum signal power and noise power.

$$PSNR = 10 \log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right)$$

3) **Normalisation Coefficient:** It means range of values.

$$NC = \frac{\sum_{s \in ref} M_s}{n_{ref}},$$

Where M_s= mean value for reference

Experiment 1

This is a hand image experiment. Three techniques were applied on first image shown in Figure 1. Various quality parameters have been calculated such as MSE, RMSE, PSNR and NC. These three techniques are compared based on four parameters shown in Table 1 and comparison of MSE and PSNR values are shown in Figure 2. It shows minimum value of MSE and maximum value of PSNR. It means the method have increased clarity of the image.

| | | | |
| (a) | (b) | (c) | (d) |
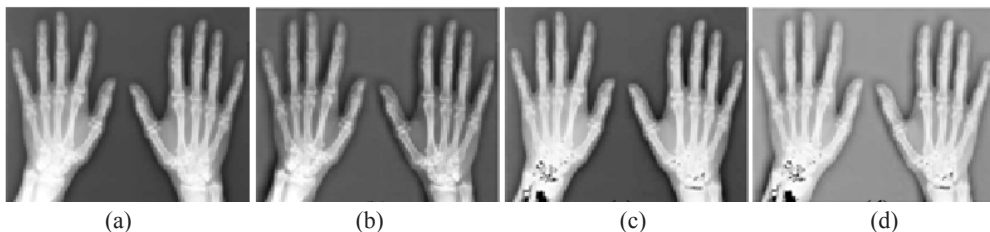
Figure 1. (a) Original image; (b) INT Operator; (c) Fuzzy Type-1; and (d) Fuzzy Type-2

Table 1
Techniques comparison for hand image

Parameters	MSE	RMSE	PSNR	NC
F2	25.5444	0.4143	57.6458	0.00033
F1	4.3789	0.1681	89.5659	0.0012
INT	2.8159	0.1337	100	0.0019
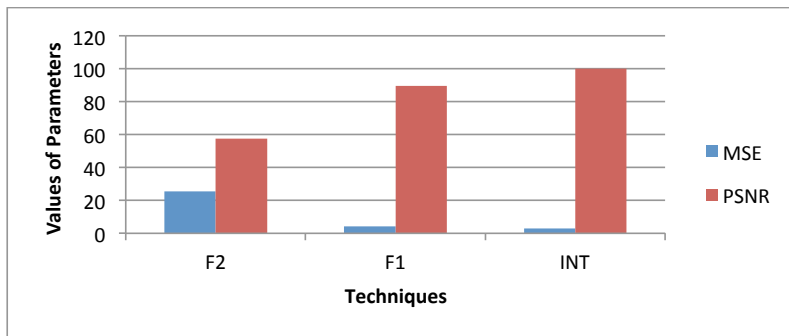
Figure 2. MSE and PSNR values comparison graph

Experiment 2

Three techniques were applied on brain image as shown in Figure 3. Various quality parameters have been calculated such as MSE, RMSE, PSNR and NC. These three techniques are compared and are shown in Table 2. Comparison of MSE and PSNR values are shown in Figure 4. Table 2 shows a minor value of MSE and RMSE which means maximum pixels are bright in the image. Figure 4 shows a greater value of PSNR which means they are brighter.

| (a) | (b) | (c) | (d) |
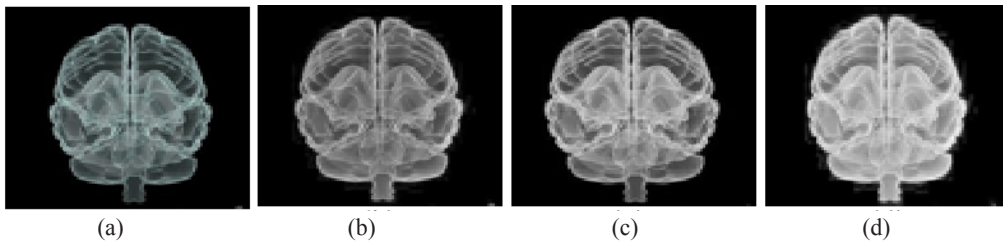
Figure 3. (a) Original image; (b) INT Operator; (c) Fuzzy Type-1; and (d) Fuzzy Type-2

Table 2
Comparison of techniques for brain image

Parameters	MSE	RMSE	PSNR	NC
F2	10.6655	0.2419	57.0233	0.00040
F1	2.0279	0.1023	86.307	0.0014
INT	1.1226	0.0747	100	0.0028
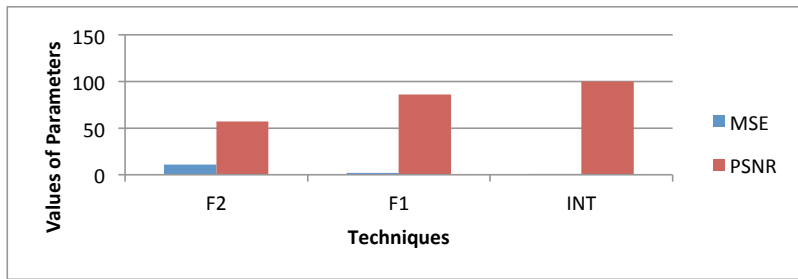
Figure 4. MSE and PSNR values comparison Graph

Experiment 3

The third experiment is a head image. Three techniques were applied and shown in Figure 5. Various quality parameters are used such as MSE, RMSE, PSNR and NC. These three techniques are compared based on four parameters shown in Table 3. The pixels are brighter based on parameters PSNR and NC. Figure 6 shows the comparison of PSNR and MSE values.

| (a) | (b) | (c) | (d) |
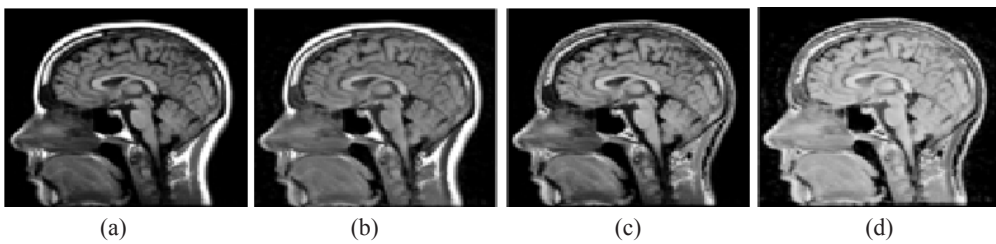
Figure 5. (a) Original image; (b) INT Operator; (c) Fuzzy Type-1; and (d) Fuzzy Type-2

Table 3
Technique comparison for head image

Parameters	MSE	RMSE	PSNR	NC
F2	40.7719	0.5405	55.3084	0.0012
F1	14.7564	0.3225	71.3044	0
INT	3.81	0.1607	100	0.0017
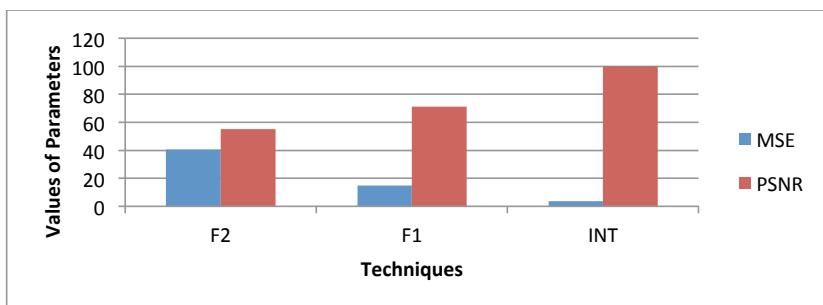
Figure 6. MSE and PSNR values comparison graph

Experiment 4

The fourth experiment is shown in Figure 7. Three techniques have been applied and compared with four quality parameters. Figure 8 shows the comparison of MSE and PSNR values. To check the robustness, various quality parameters have been calculated such as MSE, RMSE, PSNR and NC. Comparisons of these three techniques are based on four parameters shown in Table 4. Higher NC values mean good quality of image. The INT Operator always has higher NC values compared with the rest.
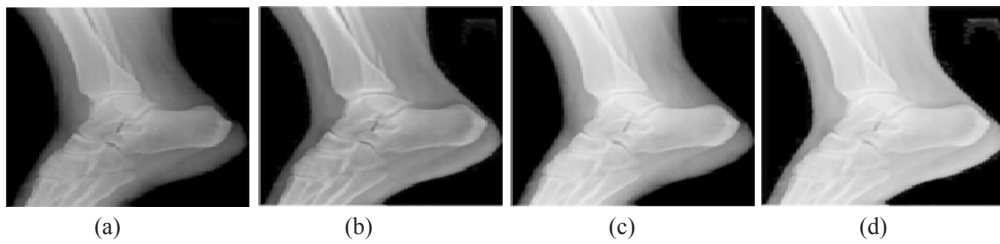
Figure 7. (a) Original image; (b) INT Operator; (c) Fuzzy Type-1; and (d) Fuzzy Type-2

Table 4
Technique comparison for ankle image

Parameters	MSE	RMSE	PSNR	NC
F2	26.5421	0.4278	58.6169	0.0026
F1	10.7424	0.2704	73.4866	0.0043
INT	3.1294	0.1437	100	0.0049
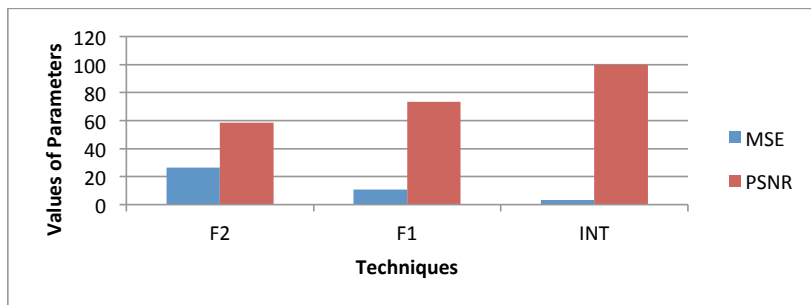
Figure 8. MSE and PSNR values comparison graph

Experiment 5

The fifth experiment is on knee image. Three techniques were used as shown in Figure 9. Various quality parameters are calculated such as MSE, RMSE, PSNR and NC. Table 5 shows a higher value of PSNR and NC for brightness.
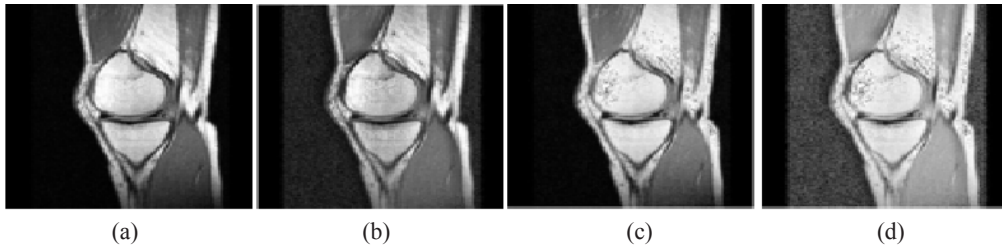
Figure 9. (a) Original image; (b) INT Operator; (c) Fuzzy Type-1; and (d) Fuzzy Type-2

Table 5
Techniques comparison for knee image

Parameters	MSE	RMSE	PSNR	NC
F2	26.0675	0.435	63.3894	0.00036
F1	5.0747	0.1882	95.4083	0.0012
INT	4.2038	0.1707	100	0.0018
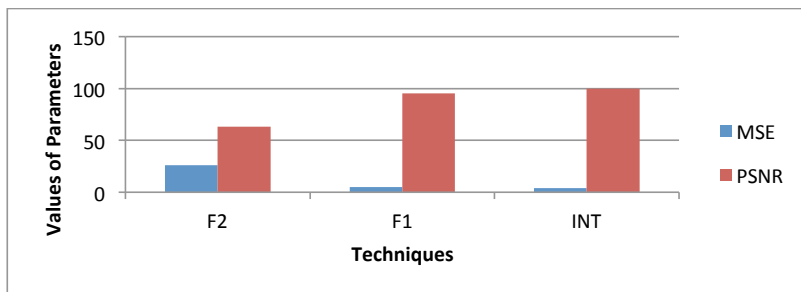
Figure 10. MSE and PSNR values comparison graph

COMPARATIVE STUDY AND RESULTS

Table 5 shows results of all 3 operators based on all the 5 images. We can clearly observe that INT Operator has the greatest value of NC in all images. First four parameters i.e. MSE1, RMSE1, PSNR1 and NC1 represent result for hand image. Next four parameters are expressing in table representation for brain image. MSE3, RMSE3, PSNR3, NC3 and MSE4, RMSE4, PSNR4 and NC4 are signifying outcome for head image and ankle image respectively. The rest of the parameters show result for knee image.

The result shows that the INT Operator is better than other techniques to increase the contrast of the images without edge detection. It means this technique improves the brightness of the image.
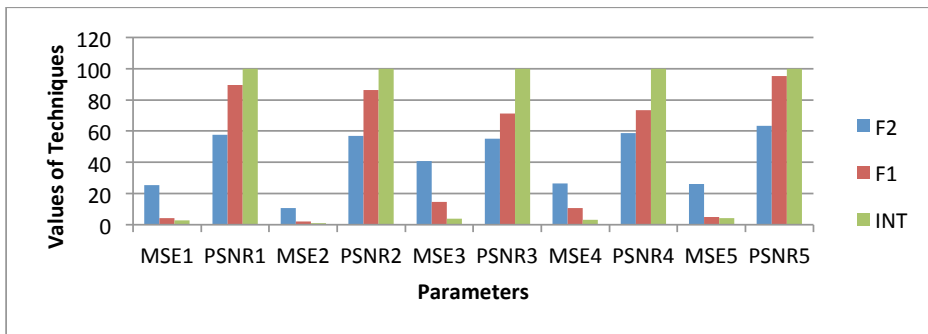
Figure 11. MSE and PSNR values comparison graph

Table 6
Technique comparison of five images

Image Name	Parameters	F2	F1	INT
Hand	MSE1	25.5444	4.3789	2.8159
	RMSE1	0.4143	0.1681	0.1337
	PSNR1	57.6458	89.5659	100
	NC1	0.00033	0.0012	0.0019
Brain	MSE2	10.6655	2.0279	1.1226
	RMSE2	0.2419	0.1023	0.0747
	PSNR2	57.0233	86.307	100
	NC2	0.00040	0.0014	0.0028
Head	MSE3	40.7719	4.7564	3.81
	RMSE3	0.5405	0.3225	0.1607
	PSNR3	55.3084	71.3044	100
	NC3	0.0012	0	0.0017
Ankle	MSE4	26.5421	10.7424	3.1294
	RMSE4	0.4278	0.2704	0.1437
	PSNR4	58.6169	73.4866	100
	NC4	0.0026	0.0043	0.0049
Knee	MSE5	26.0675	5.0747	4.2038
	RMSE5	0.435	0.1882	0.1707
	PSNR5	63.3894	95.4083	100
	NC5	0.00036	0.0012	0.0018

CONCLUSION

In this paper, various contrast enhancement techniques have been discussed for medical images. Three techniques, namely INT operator, Type-1 Fuzzy and Type-2 Fuzzy have been used on simple images (without edge detection). These techniques were compared based on

four parameters: MSE, RMSE, PSNR and NC. Findings indicated INT Operator showed the best enhancement for simple medical images. It had a higher value of PSNR and NC.

Future research should focus on edge detection image.

REFERENCES

Ansari, M. D., Ghrera, S. P., & Wajid, M. (2017). An approach for identification of copy-move image forgery based on projection profiling. *Pertanika Journal of Science and Technology, 25*(2), 507-518.

Bansal, R., Arora, P., Gaur, M., Sehgal, P., & Bedi, P. (2009, August). Fingerprint image enhancement using type-2 fuzzy sets. In *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on* (Vol. 3, pp. 412-417). IEEE.

Castillo, O., & Melin, P. (2012). Type-2 Fuzzy System. In *Springer briefs in Computational Intelligence* (pp. 7-12).

Castillo, O., Melin, P., Kacprzyk, J., & Pedrycz, W. (2007, November 2-4). Type-2 fuzzy logic: Theory and applications. In *Granular Computing, 2007, GRC 2007, IEEE International Conference on* (pp. 145-145). IEEE, USA.

Ensafi, P., & Tizhoosh, H. (2005). Type-2 fuzzy image enhancement. In *International Conference Image Analysis and Recognition* (pp. 159-166). Springer, Berlin, Heidelberg.

Hartati, S., Harjoko, A., & Nickerson, B. G.(2009). Image enhancement using fuzzy hyperbolization and artificial neural network for anomaly detection. *International Journal of Computer, Electrical, Automation, Control and Information Engineering, 3*(8), 2089-2092.

Hassanien, A. E., Soliman, O. S., & El-Bendary, N. (2011). Contrast enhancement of breast MRI images based on fuzzy type-2. In *Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011* (pp.77-83). Springer Berlin, Heidelberg.

Kamra, A., & Rani, K. (2012). An improved method for image enhancement using fuzzy approach. *International Journal of Computer Science and Information Technology and Security (IJCSITS), 2*(6), 1093-1095.

Kaur, R., & Kaur, S. (2016, March 4-5). Comparison of contrast enhancement techniques for medical images. In *Emerging Devices and Smart Systems (ICEDSS), Conference on* (pp. 155-159) IEEE, India.

Kaur, R., Chawla, M., Khiva, N. K., & Ansari, M. D. (2017). On Contrast Enhancement Techniques for Medical Images with Edge Detection: A Comparative Analysis. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9*(3-6), 35-40.

Khandewal, P., & Kaur, G. (2016). Comparative study of different image enhancement technique. In *International Journal of Electronics and Communication Technology, 7*(2), 116-121.

Kundra, H., Aashima, E., & Verma, E. M. (2009). Image enhancement based on fuzzy logic. *International Journal of Computer Science and Network Security, 9*(10),141-145.

Mahashwari, T., & Asthana, A. (2013). Image enhancement using fuzzy technique. *International Journal of Research Review in Engineering Science and Technology (IJRREST), 2*(2), 1-4.

Nirmala, D. (2015). Review: Medical image contrast enhancement techniques. *Research Journal of Pharmaceutical, Biological and Chemical Sciences, 6*(3), 321-329.

Pal, S. K., & King, R. (1981). Image enhancement using smoothing with fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics, 11*(7), 494-500.

Pal, S. K., & King, R. A. (1983). On edge detection of X-ray images using fuzzy sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*(1), 69-77.

Patel, P. D., Trivedi, V. K., & Mishra, S. (2014). Image enhancement using fuzzy techniques: Survey and overview. *International Journal of Science, Technology and Management, 3*(12), 154-160.

Preethi, S. J., & Rajeswari, K. (2013). Membership function modification for image enhancement using fuzzy logic. *International Journal of Emerging Trends and Technology in Computer Science (IJETTCS)*, 114-118.

Raju, A., Dwarakish, G. S., & Reddy, D. V. (2013). A comparative analysis of histogram equalization based techniques for contrast enhancement and brightness preserving. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 6*(5), 353-366.

Reshmalakshmi, C., & Sasikumar, M. (2013, March). Image contrast enhancement using fuzzy technique. In *Circuits, Power and Computing Technologies (ICCPCT), 2013 International Conference on* (pp. 861-865). IEEE.

Sesadri, U., & Nagaraju, C.(2015). Type2 fuzzy computing technique for image enhancement. *International Journal of Computer Science and Information Security,13*(11), 94-105.

Sharma, S., & Bhatia, A. (2015). Contrast enhancement of an image using fuzzy logic. *International Journal of Computer Applications, 111*(17), 14-20.

Shrivastava, D., & Richhariya, V. (2014). Adaptive contrast image enhancement based on fuzzy set theory. *International Journal of Advanced Research in Computer Science and Software Engineering, 4*(2), 453-462.

Singh, A., Shibu, S., & Dubey, S. (2014). Recent image enhancement techniques: A review. *International Journal of Engineering and Advanced Technology (IJEAT), 4*(1), 40-45.

Srilakshmi, M., Rao, P. V., & Prasad, K. S. (2014). Comparison of fuzzy contrast enhancement techniques. *International Journal of Computer Applications, 95*(22), 26-31.

Tizhoosh, H. R., & Fochem, M. (1995). Fuzzy histogram hyperbolization for image enhancement. *Proceedings EUFIT, 95*, 1695-1698.

Tusia, A., & Kumar, N. (2014). Performance analysis of type-2 fuzzy system for image enhancement using optimization. *International Journal of Enhanced Research in Science Technology and Engineering, 3*(7), 108-116.

Wu, D. (2014). A brief tutorial on interval type-2 fuzzy sets and systems. In *Signal and Image Processing Institute, University of Southern, California, Los Angeles* (pp. 1-13). USA.

Preparation and Thermal Properties of Cellulose Acetate/Polystyrene Blend Nanofibers via Electrospinning Technique

Rosdi, N. H., Mohd Kanafi, N. and Abdul Rahman, N.*

Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Cellulose acetate (CA) is an interesting material due to its wide spectrum of utilities across different domains ranging from absorbent to membrane filters. In this study, polystyrene (PS) nanofibres, and cellulose acetate/polystyrene (CA/PS) blend nanofibres with various ratios of CA: PS from 20: 80 to 80: 20 were fabricated by using electrospinning technique. The SEM images show that the nanofibres exhibited non-uniform and random orientation with the average fibre diameter in the range of 100 to 800 nm. It was found that the incorporation of PS had a great effect on the morphology of nanofibre. At high proportion of PS, no or less beaded CA/PS nanofibres were formed. Thermal properties of the composite nanofibres were investigated by using thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) techniques. The TGA results showed thermal stability of CA/PS nanofibres were higher than pristine CA.

Keywords: Cellulose acetate, electrospinning, nanofibres, polystyrene

INTRODUCTION

Cellulose is the most abundant renewable natural polymer material and resource. Most of the plant cells are made up of cellulose.

Just like polysaccharides, cellulose is widely used in medical applications because it is non-toxic, has a high swelling ability and stable temperature and pH variations (Raucci et al., 2014). Cellulose acetate (CA) is part of cellulose derivatives and produced by reacting cellulose with acetic anhydride with the presence of sulfuric acid as a catalyst (Yan & Yu, 2012). Cellulose acetate is well known as a biodegradable and biocompatible polymer.

Recently, cellulose acetate based nanofibres have been gaining attention because its fabrication and disposal does not damage the environment (Konwarh, Karak, &

Misra, 2013b). Cellulose acetate nanofibre has high potential to be used in various applications like in bone regeneration (Yamaguchi et al., 2016), gas sensor (Qingqing et al., 2016), supercapacitor (Yang et al., 2015), filtration (Sehaqui et al., 2016) and drug delivery (Konwarh, Karak, & Misra, 2013a). Composites/blend polymers consists of synthetic and biodegradable polymers will provide novel properties to the materials. Researchers have prepared cellulose acetate blend with other synthetic polymer to form nanofibre (Zhijiang, Yi, Haizheng, Jia, & Liu, 2016; Gopiraman, Fujimori, Zeeshan, Kim, & Kim, 2013; Zhou, He, Cui, & Gao, 2011). According to Konwarh et al., (2013b) electrospun CA in the presence of other polymer such as poly(vinyl pyrrolidone) (PVP), poly(ethylene oxide) (PEO) and poly(butyl acrylate) (PBA) have improved tensile strength and increased its stability (Konwarh, 2013a). However, there has not been much focus on how to improve thermal properties of CA based nanofibres.

Polystyrene (PS) is a cheap and strong polymer that is widely used in the plastic industry. The PS nanofibre have been widely used as filter media, ion exchanger and separator (An, Shin, & Chase, 2006). Additionally, numerous studies have focused on PS as a composite material and the result showed an improvement in material performance compared to that without PS Kaya, Kaynak, & Hacaloglu, 2016; Jia, Chen, Yu, Zhang, & Dong, 2015; Zhang, Wen, Hu, Zhang, & Liu, 2010). The PS is also able to increase thermal stability of composites material due to an increase in the mobility of the macromolecule chains which results in a to more ideal crystalline pattern (Kaya et al., 2016; Meireles, Filho, Assunc, & Zeni, 2007). In addition, Mathew et al., and Kaya et al., showed PS enhanced the thermal stability of the composite PS with natural rubber (NR) (Mathew, Packirisamy, & Thomas, 2001) and nanocomposite with organoclay (Kaya et al., 2016) respectively.

In general, nanomaterials are defined as materials that are manufactured and used at a scale less than 100 nm. However, for polymer nanofibres, it has been widely accepted for fibre diameter less than 1000 nm (Lubasova, Niu, & Lin, 2015; Stone, Gosavi, Athauda, & Ozer, 2013; Zhou et al., 2011; An et al., 2006). Many techniques can be used to produce polymer nanofibres including electrospinning, self-assembly, template synthesis and phase separation (Huang, Zhang, Kotaki, & Ramakrishna, 2003). Among these techniques, electrospinning is considered simple, cost effective, and suitabile to produce very long fibres from various polymers and composite materials (Abdul Rahman et al., 2014), (Konwarh, 2013a). Polymer nanofibres produced high surface area to volume ratio that make the electrospun fibrous membranes suitable for use in many applications such as filtration, sensors, catalysis and protective clothing (Huang, Zhang, Kotaki, & Ramakrishna, 2003).

This study focused on producing PS nanofibres and CA/PS composite nanofibres by using electrospinning technique. PS was chosen in this study to improve thermal stability of the CA. The thermal properties of the composite were investigated using differential scanning calorimetry (DSC) and thermogravimetric analysis (TGA).

MATERIALS AND METHODS

Polystyrene and cellulose acetate (Mn = 30,000) was supplied by BDH Chemicals. Dimethylformamide (DMF) and tetrahydrofuran (THF) used as solvents were purchased from the Fischer Chemical. All chemicals were used as received without further purification.

Preparation of PS nanofibres

PS nanofibres were prepared using mixed solution of DMF and THF (1:1). PS solution was filled in a 5 mL syringe fitted with a steel needle and was placed on the syringe pump whereas the collector was 8.0 cm away from the tip of needle. By using 35 to 45 kV of voltage and 2 mL/h for flow rate, PS nanofibres were synthesised. The PS nanofibre was fabricated based on five different concentrations from 0.5 to 10 w/v%.

Preparation of cellulose acetate/ polystyrene (CA/PS) nanofibres

Electrospinnable solution of CA/PS nanofibre was prepared by dissolving appropriate amount of CA and PS in the mixed solution of DMF and THF (1:1) with total concentration of 2.5 w/v%. The CA/PS solution was filled in a 5mL syringe attached to a steel needle. An aluminum foil was placed 8.0 cm distance from the needle. The electrospinning process was carried out at room temperature in a horizontal spinning configuration, using the fixed applied voltage of 40 kV and pump rate of 2 ml/h. The humidity was kept at around 40%. The electrospun fibres were collected directly on an aluminum foil which was used to collect substrate.

Characterisations

The morphology of the electrospun nanofibres was observed using Jeol JSM-6400 scanning electron microscopy (SEM). The sample for SEM observations was put on the aluminium stubs and was coated with gold prior image the nanofibres morphology. The electrospun fibre diameter was measured using SEM images. The average of diameter for each sample was calculated from at least 30 measurements. For beaded fibres, only the areas of the fibres between beads were measured. Thermal properties of the composite nanofibres were determined using differential scanning calorimetry (DSC) (Model DSC Mettler Toledo 822°) from 50°C to 250°C with a heating rate of 10°C/min under nitrogen flow at 50 mL/min. Thermogravimetric analysis (TGA) (Mettler Toledo thermogravimetric model TGA/SDTA 851) measurement was done under nitrogen gas. The sample was heated from 35°C to 500°C with a heating rate of 10°C/min.

RESULTS AND DISCUSSION

Morphology of the Nanofibres

Electrospun PS nanofibres 2.5 w/v% and CA/PS composite nanofibres with various CA to PS ratios were prepared using electrospinning technique. The total concentration of CA/PS was kept constant at 2.5 w/v%, but the ratio of CA to PS was varied between 20% and 80%. Figure 1 shows the SEM images of electrospun PS and CA/PS nanofibre. PS nanofibres have average fibre diameter of 249 ± 70 nm.

As shown in Table 1, the average fibre diameter decreased from 229 ± 200 nm to 111 ± 50 nm as the ratio of PS in CA/PS nanofibre reduced from 80% to 20%. However, the formation of beads was increased as CA content in the composite nanofibres increased. The optimum morphology of CA: PS nanofibres was found at 40% of CA. However, in this study, CA alone cannot be electrospun by using electrospinning technique.
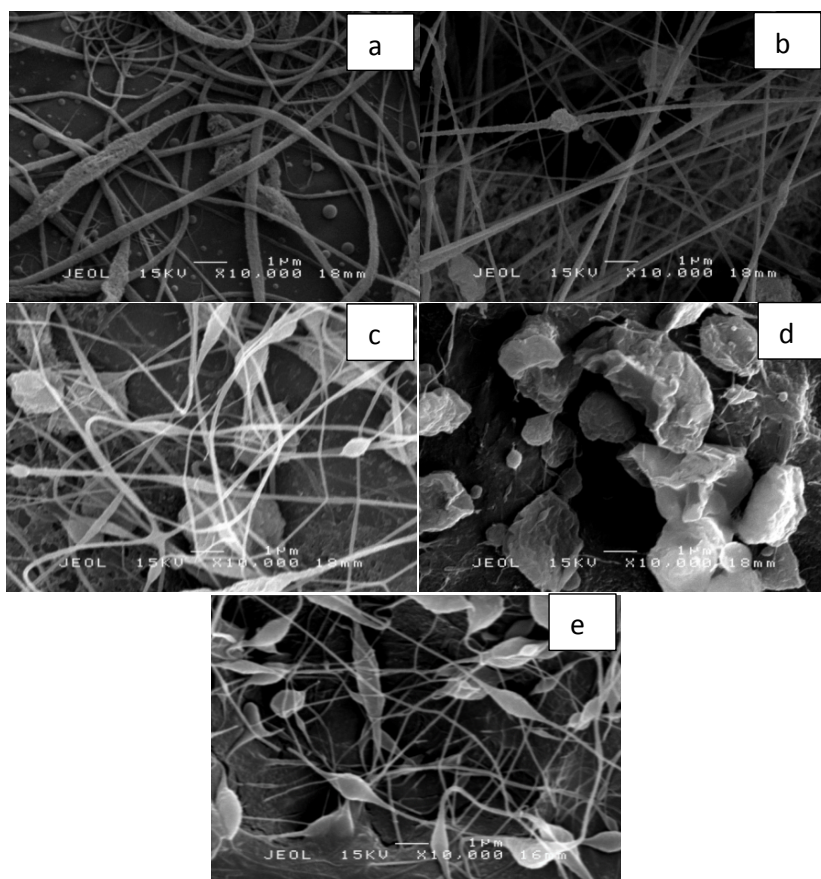
Figure 1. SEM images of CA/PS blend nanofibre with CA: PS ratios of (a) PS (b) 20:80 (c) 40:60 (d) 60:40 and (e) 80:20

Table 1
Fibre diameter of PS and CA/PS composite nanofibre

CA:PS	Average fibre diameter
	(nm) ± Standard Deviation
20:80	229 ± 200
40:60	186 ± 60
60:40	-
80:20	111 ± 50
PS	249 ± 70

Spectroscopic study of the nanofibres

Figure 2 shows IR spectra of PS nanofibre, pristine CA and CA/PS nanofibres with different ratios of CA and PS. The PS nanofibres have three main peaks in the spectrum corresponding to PS structure that consists of alkyl group and aromatic ring. The absorption bands of IR

spectrum of PS nanofibres at wavenumber 3000 to 3100 cm^{-1} belong to the aromatic C-H stretching vibration of CH$_2$ groups on the PS main chain. A peak at 1598 cm^{-1} shows the present of aromatic C=C skeletal and a peak at 1448 cm^{-1} which corresponds to aromatic C-H bending. Absorption peaks at around 1100 cm^{-1}, 765 cm^{-1} and 700 cm^{-1} corresponding to vibrations of C-H bending of the benzene ring, C–H out-of-plane bend, and CH$_2$ rocking mode respectively (Nair, Hsiao, & Kim, 2008). The IR spectrum of pristine CA consists of a broad absorption band at 3300-3500 cm^{-1}, which is belongs to stretching O-H group. The absorption bands at 2700- 2900 cm^{-1} are attributed to C-H stretching of alkane. A peak at 1035 cm^{-1} corresponds to C-O-C (ether linkage) from glycosidic units.

The IR spectrum of the composite CA/PS nanofibres shows combination peaks of CA and PS. The O-H peaks that belong to CA appear very small and become clearly visible when the ratio of CA was increased from 40% to 80%. The stretching O-H vibration peak slightly shifted to 3029 cm^{-1} from 3417 cm^{-1} and the intensity of the peak reduced as the concentration of CA is reduced. This is due to the hydrogen bonds formed between the OH groups in CA are reduced as CA content becomes less in the composite nanofibres. The C-H aromatics peak was clearly observed in the IR spectra in Figure 2(a) to (d). However, the peak disappeared in Figure 2(e) and (f). A peak in between 1704.42 cm^{-1} to 1741.42 cm^{-1} that belong to carbonyl group shifted to higher wavenumber indicates there is an interaction between CA and PS in CA/PS nanofibres (Chen, Wang, & Huang, 2011).
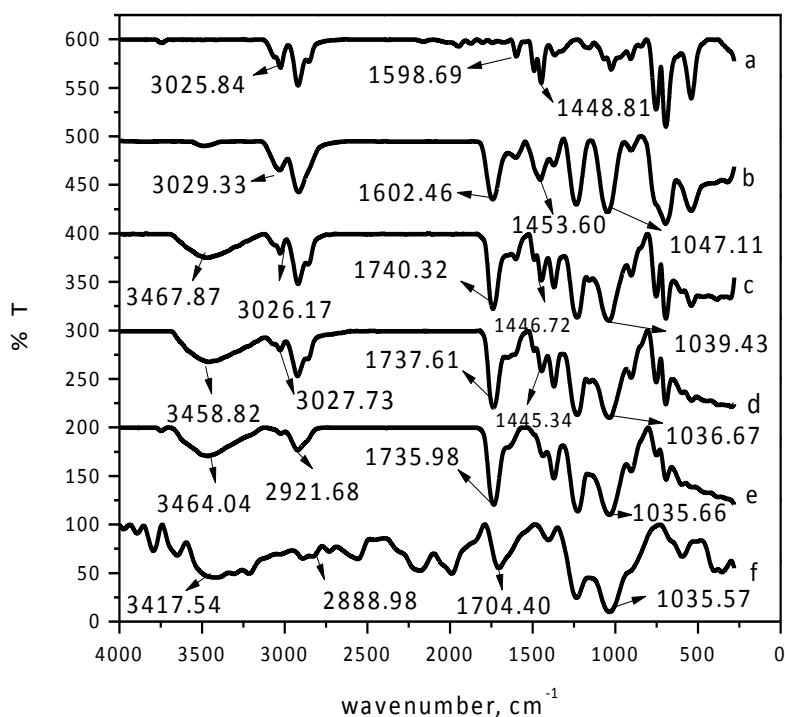
Figure 2. FTIR spectra of (a) PS nanofibre and CA/PS blend nanofibre; (b) 20/80; (c) 40/60; (d) 60/40; (e) 80/20; and (f) CA pristine

Thermogravimetric Analysis of the nanofibres

Figure 3 shows TGA thermograms of CA pristine and PS nanofibre. As can be seen from the TGA results and PS nanofibres show only a single degradation step. According to Uyar et al., (2008) polystyrene degrades in a single step starting at 250°C until 500°C under nitrogen atmosphere. The degradation range for PS nanofibres was between 210°C and 463°C is due to rupture of PS chains (Meireles, 2006). CA pristine undergoes two steps of degradation. The first peak of degradation started at 76°C and the second decomposition peak was at 342°C. 10% of weight loss was recorded at first peak and almost 80% weight loss was occurred at second peak. Similar result was obtained by Meireles et al., where CA underwent major degradation between 330°C and 450°C (Arthanareeswaran, Thanikaivelan, Srinivasn, Mohan, & Rajendran, 2004).
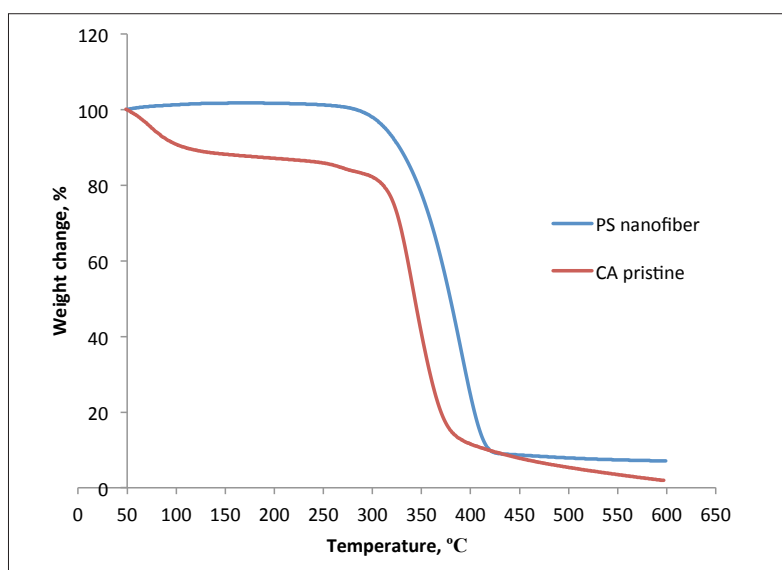
Figure 3. TGA thermograms of PS nanofibre and CA pristine

Figure 4 shows TGA thermograms of PS nanofibres, CA pristine and CA/PS nanofibres. DTG peaks of CA and the nanofibres are shown in Figure 5 and summarised in Table 2. The first step is the elimination of moisture, and the second, third and fourth steps are decomposition of cellulose acetate and polystyrene chains. The elimination of moisture was occurred at between 30°C to 120°C. The moisture mostly comes from CA due to present of hydroxyl group that can have hydrogen bond with water. The second and third degradation temperature of composite nanofibre was shifted to the higher temperature as the PS content in the composite nanofibre was increased (Table 2). This is due to PS having higher thermal stability than CA (Arthanareeswaran, 2004). Pure CA has lower thermal stability than PS nanofibre and CA/PS nanofibres. At high ratio PS, the blend nanofibres have higher thermal stability than CA/PS nanofibre with high ratio CA. This is probably due to the presence of PS that could increase the crystallinity of PS composite nanofibres (Ding et al., 2003).
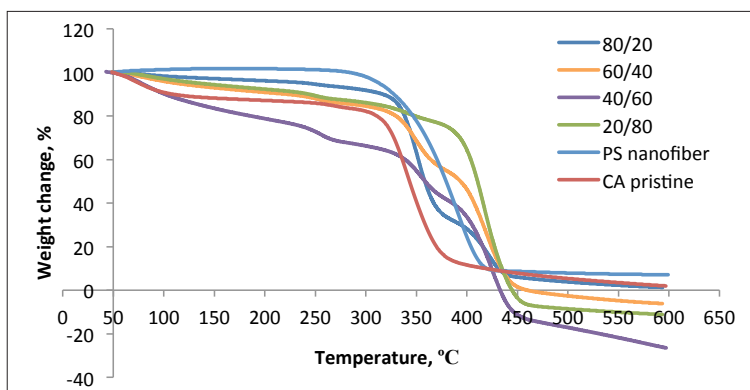
Figure 4. TGA thermograms of CA/PS blend nanofibre with CA: PS ratios of 80/20, 60/40, 40/60, 20/80, CA pristine and PS nanofibre
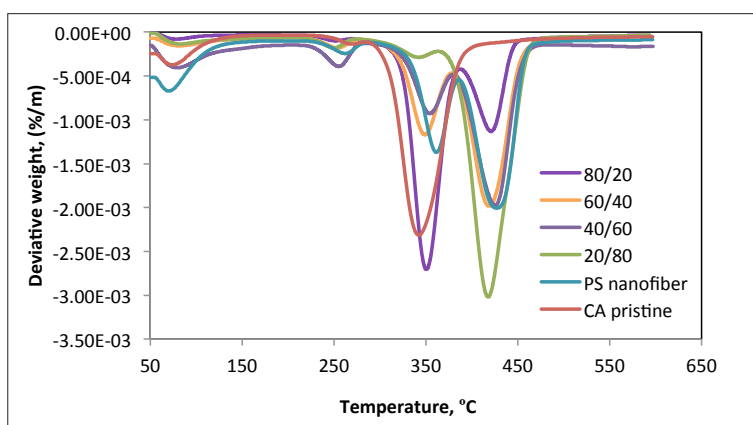
Figure 5. DTG thermograms of CA/PS blend nanofibres with CA: PS ratios of 20/80, 40/60, 60/40 and 80/20, CA pristine and PS nanofibres

Table 2
DTG results for CA/PS composite nanofibre, CA pristine and PS nanofibre

CA/PS composite nanofibres ratio	1st step (°C)	2nd step (°C)	3rd step (°C)	4th step (°C)
CA pristine	76	-	342	
PS nanofibre	-	-	390	
20/80	81	249	-	417
40/60	79	255	-	425
60/40	81	253	348	419
80/20	76	-	350	421

DSC Analysis of PS nanofibres

Figure 6 shows the first scan DSC thermogram of CA pristine, PS nanofibres and CA/PS nanofibres with different ratios of CA to PS. The DSC thermograms of CA/PS have similar pattern as previously reported for CA/PS blend film (Meireles et al., 2006). All the composite

nanofibres show an endothermic peak at temperature from 50°C to 100°C. The endothermic peak of CA is due to the outflow of water. This in agreement with observed TGA thermogram. The endothermic peak is shifted to lower temperature for CA/PS nanofibres compared to CA pristine. Incorporation of PS to CA could increase the hydrophobicity of the composite nanofibres and prevent water absorption.
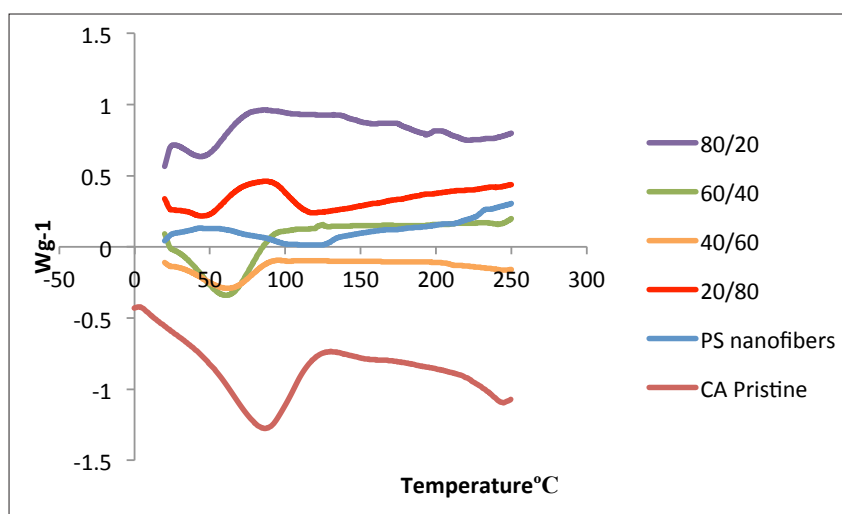
Figure 6. DSC thermograms of PS nanofibres and CA/PS blend nanofibres with CA: PS ratios of 20/80, 40/60, 60/40, 80/20and PS nanofibres

According to Holmes et al. (2010), glass transition (T_g) of PS occurs between 74°C – 109°C and the melting temperature is at 240°C to 250°C. However, in Figure 6, neither peaks of T_g nor the melting peak was observed for PS and CA/PS nanofibres in the thermograms. The absence of T_g peak was possibly because certain polymer only showed the T_g value after second run of DCS test (Meireles et al., 2006), (Hyon, Cha, & Ikada, 1987) or it overlapped with endothermic peak of CA at around 100°C.

According to Meireles et al., (2006), temperature of fusion increased sharply as the amount of PS added was increased indicating the increase in crystallinity (Zhijiang et al., 2016). This is due to the movement of the macromolecule chains caused by PS. This leads to a perfect crystalline pattern. However, the peak is not observed in the CA/PS nanofibres, possibly due to compact structure of the nanofibres preventing the movement of molecular chains to crystallise.

CONCLUSION

CA/PS composite nanofibres were successfully prepared using electrospinning technique. The SEM images showed the prepared composite nanofibre had smooth nanofibres morphology as the PS ratio in blend polymer solution increased. All electrospun nanofibre and composite nanofibre exhibited submicron sized fibre diameters, ranging from 100 to 800 nm. The thermal

degradation properties of CA/PS nanofibres were investigated using TGA and DSC techniques. CA/PS composites nanofibre has higher thermal stability than CA pristine. The DSC results showed mixing CA with PS reduced outflow of water. Unlike the CA/PS film, the absence of crystallisation peak in DSC thermograms of composite nanofibres indicated the nanofibres have low tendency to crystallise.

ACKNOWLEDGEMENTS

REFERENCES

An, H., Shin, C., & Chase, G. G. (2006). Ion exchanger using electrospun polystyrene nanofibers. *Journal of Membrane Science, 283*(1–2), 84–87. http://doi.org/10.1016/j.memsci.2006.06.014

Arthanareeswaran, G., Thanikaivelan, P., Srinivasn, K., Mohan, D., & Rajendran, M. (2004). Synthesis, characterization and thermal studies on cellulose acetate membranes with additive. *European Polymer Journal, 40*(9), 2153–2159. http://doi.org/10.1016/j.eurpolymj.2004.04.024

Chen, C., Wang, L., & Huang, Y. (2011). Electrospun phase change fibers based on polyethylene glycol/cellulose acetate blends. *Applied Energy, 88*(9), 3133–3139. http://doi.org/10.1016/j.apenergy.2011.02.026

Ding, R., Hu, Y., Gui, Z., Zong, R., Chen, Z., & Fan, W. (2003). Preparation and characterization of polystyrene/graphite oxide nanocomposite by emulsion polymerization. *Polymer Degradation and Stability, 81*(3), 473–476. http://doi.org/10.1016/S0141-3910(03)00132-0

Gopiraman, M., Fujimori, K., Zeeshan, K., Kim, B. S., & Kim, I. S. (2013). Structural and mechanical properties of cellulose acetate/graphene hybrid nanofibers: Spectroscopic investigations. *Express Polymer Letters, 7*(6), 554–563. http://doi.org/10.3144/expresspolymlett.2013.52

Huang, Z. M., Zhang, Y. Z., Kotaki, M., & Ramakrishna, S. (2003). A review on polymer nanofibers by electrospinning and their applications in nanocomposites. *Composites Science and Technology, 63*(15), 2223–2253. http://doi.org/10.1016/S0266-3538(03)00178-7

Hyon, S., Cha, W., & Ikada, Y. (1987). Polymer Bulletin 9. *Polymer Bulletin, 29*(3), 119–126. http://doi.org/10.1007/BF00310794

Jia, Y., Chen, L., Yu, H., Zhang, Y., & Dong, F. (2015). Graphene oxide/polystyrene composite nanofibers on quartz crystal microbalance electrode for the ammonia detection. *RSC Advances, 5*(51), 40620–40627. http://doi.org/10.1039/C5RA04890G

Kaya, H., Kaynak, C., & Hacaloglu, J. (2016). Thermal degradation of polystyrene composites. Part II. The effect of nanoclay. *Journal of Analytical and Applied Pyrolysis, 120*, 194–199. http://doi.org/10.1016/j.jaap.2016.05.005

Konwarh, R., Karak, N., & Misra, M. (2013a). Electrospun cellulose acetate nano fibers: The present status and gamut of biotechnological applications. *Biotechnology Advances, 31*(4), 421–437. http://doi.org/10.1016/j.biotechadv.2013.01.002

Konwarh, R., Misra, M., Mohanty, A. K., & Karak, N. (2013b). Diameter-tuning of electrospun cellulose acetate fibers: A Box-Behnken design (BBD) study. *Carbohydrate Polymers, 92*(2), 1100-1106. http://doi.org/10.1016/j.carbpol.2012.10.055

Lubasova, D., Niu, H., & Lin, T. (2015). RSC Advances Hydrogel properties of electrospun polyvinylpyrrolidone and polyvinylpyrrolidone. *RSC Advances, 5*, 54481–54487. http://doi.org/10.1039/C5RA07514A

Mathew, A. P., Packirisamy, S., & Thomas, S. (2001). Studies on the thermal stability of natural rubber/polystyrene interpenetrating polymer networks: Thermogravimetric analysis. *Polymer Degradation and Stability, 72*(3), 423–439. http://doi.org/10.1016/S0141-3910(01)00042-8

Meireles, C. D. S., Filho, G. R., de Assunção, R. M., Zeni, M., & Mello, K. (2007). Blend compatibility of waste materials—cellulose acetate (from sugarcane bagasse) with polystyrene (from plastic cups): diffusion of water, FTIR, DSC, TGA, and SEM study. *Journal of Applied Polymer Science, 104*(2), 909-914. http://doi.org/10.1002/app

Nair, S., Hsiao, E., & Kim, S. H. (2008). Fabrication of electrically-conducting nonwoven porous mats of polystyrene-polypyrrole core-shell nanofibers via electrospinning and vapor phase polymerization. *Journal of Materials Chemistry, 18*(42), 5155–5161. http://doi.org/Doi 10.1039/B807007e

Qingqing, Z., Fenglin, W., Qufu, H., Yang, Z., Chen, Y., Zhang, J., … & Aspects, E. (2016). A room temperature ammonia gas sensor based on cellulose / TiO_2 / PANI composite nanofibers. *Colloids and Surface A: Physicochemical and Engineering Aspects, 494*, 248–255. http://doi.org/10.1016/j.colsurfa.2016.01.024

Rahman, N. A., Srinivas, A. G., & Travas-Sejdic, J. (2014). Spontaneous stacking of electrospun conjugated polymer composite nanofibers producing highly porous fiber mats. *Synthetic Metals, 191*, 151–160. http://doi.org/http://dx.doi.org/10.1016/j.synthmet.2014.03.006

Sehaqui, H., Mautner, A., de Larraya, U. P., Pfenninger, N., Tingaut, P., & Zimmermann, T. (2016). Cationic cellulose nanofibers from waste pulp residues and their nitrate, fluoride, sulphate and phosphate adsorption properties. *Carbohydrate Polymers, 135*, 334–340. http://doi.org/10.1016/j.carbpol.2015.08.091

Yamaguchi, K., Prabakaran, M., Ke, M., Gang, X., Chung, I. M., Um, I. C., … & Kim, I. S. (2016). Highly dispersed nanoscale hydroxyapatite on cellulose nanofibers for bone regeneration. *Materials Letters, 168*, 56–61. http://doi.org/http://dx.doi.org/10.1016/j.matlet.2016.01.010

Yan, J., & Yu, D. G. (2012). Smoothening electrospinning and obtaining high-quality cellulose acetate nanofibers using a modified coaxial process. *Journal of Materials Science, 47*(20), 7138–7147. http://doi.org/10.1007/s10853-012-6653-2

Yang, C., Chen, C., Pan, Y., Li, S., Wang, F., Li, J., … & Li, D. (2015). Flexible highly specific capacitance aerogel electrodes based on cellulose nanofibers, carbon nanotubes and polyaniline. *Electrochimica Acta, 182*, 264–271. http://doi.org/10.1016/j.electacta.2015.09.096

Zhang, X., Wen, S., Hu, S., Zhang, L., & Liu, L. (2010). Electrospinning preparation and luminescence properties of Eu(TTA)3phen/polystyrene composite nanofibers. *Journal of Rare Earths, 28*(3), 333–339. http://doi.org/10.1016/S1002-0721(09)60108-3

Zhijiang, C., Yi, X., Haizheng, Y., Jia, J., & Liu, Y. (2016). Poly(hydroxybutyrate)/cellulose acetate blend nanofiber scaffolds: Preparation, characterization and cytocompatibility. *Materials Science and Engineering C, 58*, 757–767. http://doi.org/10.1016/j.msec.2015.09.048

Zhou, W., He, J., Cui, S., & Gao, W. (2011). Preparation of electrospun silk fibroin/Cellulose Acetate blend nanofibers and their applications to heavy metal ions adsorption. *Fibers and Polymers, 12*(4), 431–437. http://doi.org/10.1007/s12221-011-0431-7

Experimental Analysis of Condensation in Helical Coil Tube

Rashed Ali[1,2]* and Nitin P Gulhane[1]

[1]Department of Mechanical Engineering, Veermata Jijabai Technological Institute, Mumbai, Maharashtra 400031, India
[2]Pillai College of Engineering, New Panvel, Mumbai, Maharashtra 410206, India

ABSTRACT

A helically coiled tube enhances heat transfer rate due to development of secondary flows. Use of helical tube as heat exchanger will enhance heat transfer coefficient. Correlations available in the literature for calculating condensation heat transfer coefficient for straight tube are used to evaluate the performance of helically coiled tube-in-shell heat exchanger. Experiments were performed on helically coiled tube of 175 mm coil diameter. Measurements were taken in 10 steps at inlet rate of mass flow of steam. Mean value of the coefficient of heat transfer at inside surface of the tube for condensation was evaluated by experimental methods and results were compared with three correlations available in literature and the deviations are reported. A new empirical correlation is proposed which is based on the experimental investigations.

Keywords: Condensation, heat transfer coefficient, helical coil

INTRODUCTION

Helical coils are coiled tubes used in many industries for enhancing the heat transfer. It provides many advantages over the straight tube heat exchanger (Prabhajan et al., 2004). Helical coils are also used for condensation and boiling. Condensation phenomenon inside tubes is essentially needed for condensers (Colorado et al., 2011). Condensers are employed in most of the chemical, petroleum, processing and power industries, for distillation, refrigeration & air-conditioning industries, most importantly for power generation. Generally, in chemical process industries, water-cooled shell and tube heat exchangers or condensers are employed. Condensation may occur drop-wise or film wise. While drop-wise condensation is most desirable, it is difficult to maintain. On the other hand, film-wise condensation offers resistance to heat transfer (Kern, 1965). In case of single phase fluid flow when fluid flows through helically coiled tube, the flow

pattern is different in comparison to flow pattern through a straight tube (Bae et al., 1969). In helical tubes, fluid is under influence of centrifugal mass forces producing secondary flow of fluid which is responsible for the enhancement of heat transfer coefficient inside helical tubes (Vashisth et al., 2008). The same phenomenon was reported by Owhadi et al. (1968) when two-phase flow occured in helical tube. The only difference lies in the fact that in case of condensation heat transfer, secondary flow causes increased contact surface area inside the tube between vapour and tube wall which ultimately leads to significant increase in heat transfer coefficient of condensation. Therefore, the purpose of this investigation was to experimentally evaluate heat transfer coefficient in condensation inside vertical helical coil tube by condensing water steam. Experimentally determined values of heat transfer coefficients were compared with predicted values using equations reported in the literature which were experimentally tested and were based on conduction resistance method.

LITERATURE REVIEW

Nusselt theory of film-wise condensation heat transfer for plane surface and horizontal cylindrical surfaces was published in 1916. In the today's world where revolutionary developments are happening in the field of engineering, his theory is still relevant, even after a century. Nusselt's theory explains elementary physics of film condensation (Tanasawa, 1990). He analysed laminar film condensation on vertical flat plate and then extended his work to laminar film condensation inside vertical tube. He assumed the thickness of film was small and it had no effect on condensing process and shear stress on edge of condensate film was proportional to pressure drop. He successfully obtained heat transfer correlation for laminar flow condensate.

Early Research on Condensation in Tubes

Chato (1926) had theoretically and experimentally studied condensation in inclined and horizontal tubes and Nitheanandan et al. (1993) carried out an experimental study for an inclined tube for the study of flow regime transitions for steam condensation. They have concluded that transition between wavy and slug flow regimes of condensate is because of inclination of the tube. It was shown the wavy flow was more dominant on downward inclinations whereas slug flow of condensate was prominent on upward inclinations. Tandon et al. (1995) carried out experiments for forced convection condensation inside the horizontal tube. They had observed that, the effect of various parameters on condensing heat transfer coefficient was the same as reported earlier by other investigators. They found vapour Reynolds number of 3×10^4 at which change over from annular flow/semi-annular flow to wavy flow occurred. Berenson et al. (1968) investigated the condensation of pure R-134a vapour inside a single micro-fin tube subjected to different inclinations. They had used inclination angle from –90 to +90 under refrigerant mass velocity 54 to 107 kg/m²s for each inclination. They had concluded that value of heat transfer coefficient was dependent on the inclination of tube. It decreased as the vapour quality and mass velocity decreased. It was also concluded that clockwise inclination of 3° gave good rise in value of heat transfer coefficient in comparison with 0° inclination.

They also proposed correlation for evaluating heat transfer coefficient. Wang and Du (2000) conducted condensation heat transfer study for laminar film on tube-in-tube condensers having counter flow arrangement on four small diameter (less than 5 mm) tube with water steam as the working medium for low range of mass flux that is in the range of 10 to 100 kg/s/m². It was tested for various inclination angles for assessing the performance of the condenser under the effect of gravity. They had developed an analytical model for predicting the condensation heat transfer characteristics. They had concluded that the effect of gravity in case of small diameter tube decreased. They also found that proposed analytical model gave the good predictions. Akhavan-Behabadi et al. (2007) had conducted an experimental study on R134a for observing condensation inside a micro finned, 8.92 mm in inner diameter tube. They had observed the high value of heat transfer coefficient at 30° inclination. Wang et al. (1991) had proposed semi-empirical correlation for the purpose of designing heat pipe heat exchangers. Further, they observed that inclination of thermosyphon had a strong effect on the value of heat transfer coefficient due to condensation. Saffari et al (2010) theoretically and numerically studied stratified condensation heat transfer mechanism in an inclined tube and reported high value heat transfer coefficient for 30° incline in comparison to 60° and 90° incline tube.

Recent Literature on Condensation in Helical Coil

Han et al. (2005) investigated the condensation heat transfer coefficient of R-134a at various vapor saturation temperatures. He had used annular helical pipe tube-in-shell heat exchanger. He observed that condensation heat transfer coefficient increased with the mass flux in the annular helical pipe. Wongwises et al. (2006) experimentally investigated two-phase pressure drop and condensation of HFC 134a inside tube-in-tube helical coil heat exchanger. They had reported 33-53% increase in the value of heat transfer coefficient in helical coil condenser as compared to straight tube condenser. Shao et al. (2007) examined condensation of R-134a in helical tube-in-tube and horizontal straight tube condenser. Refrigerant R-134a at various saturation temperature and vapor quality were used in experimentation. They reported 4-13.8% higher heat transfer coefficients in helical coil compared with straight tube condenser. Mosaad et al. (2009) reported condensation and pressure drop at 815 kPa pressure in coiled double tube heat exchanger. They concluded that heat transfer coefficient in condensation increased with mass flux and decreases with rising saturation temperature difference. Gupta et al. (2014) investigated condensation and pressure drop characteristics of R-134a helical tube in shell heat exchanger in a horizontal orientation. They studied the flow regimes during condensation on Taitel and Dukler flow maps, and proposed new correlation and enhancement factor for helical coil condenser.

Mozafari et al (2015) studied condensation and pressure drop behaviour of R600a in the helical tube in tube exchanger at different inclination angles. They had reported the highest value of heat transfer coefficient at 30ºC inclination compared with horizontal and vertical position. Salimpour et al (2017) studied heat transfer coefficient for condensation of R-404A vapor. They reported high heat transfer coefficient for small curvature coils compared to large coils. They had proposed new correlation for calculating heat transfer coefficient in condensation for the helical coil.

Yu et al. (2018) studied condensation of hydrocarbon refrigerant propane in a helical tube. Experiments were conducted for mass flux from 200 to 400 kg/m²s and saturated temperature from 40 to 27°C. The effects of flow parameters were analysed. They concluded heat transfer coefficients increased with the increase in vapour quality and mass flux, while the influence of heat flux on heat transfer at the same quality was observed as insignificant.

Gap in Literature and Novelty of Work

- It is evident from literature review that there is no study on condensation of steam in the vertical helical tube-in-shell heat exchanger. Majority of the studies were oriented towards the condensation of refrigerants, particularly for tube-in-tube helical type heat exchangers.

- There is much application in industries where the tube-in-shell type of heat exchangers are used, for example in chemical reactors, steam generators etc.

- In this context, there is no study for condensation of steam considering the effect of gravity on condensation in the vertical helical tube-in-shell heat exchanger.

In the present work, condensation of steam is carried out in the helically coiled tube. Steam flows inside a helically coiled tube and the helical coil is placed in the center of the shell surrounded by cool water. Experimental analysis is carried out and development of the dimensionless equation similar in kind to that of Nusselt's equation is presented. The experimentally obtained results are compared with three chosen equations which predict coefficient of heat transfer in condensation.

EXPERIMENTATION

Analysis on the experimental setup by conducting experiments in 10 steps on vertically mounted single helical coil was done. The helical coil was mounted in the shell. The coil dimensions are given in Table 1. The coil was made using the SS304 material on bending machine. The thickness of the tube of the coil was selected keeping in mind that there should not be a change in tube cross-section which was circular here. The vertical helical coil was surrounded by the continuous flow of cold water. Sufficient amount of insulation was provided everywhere to avoid the heat losses. The experimental setup was divided into two loops, one was water steam loop and other was cooling water loop.
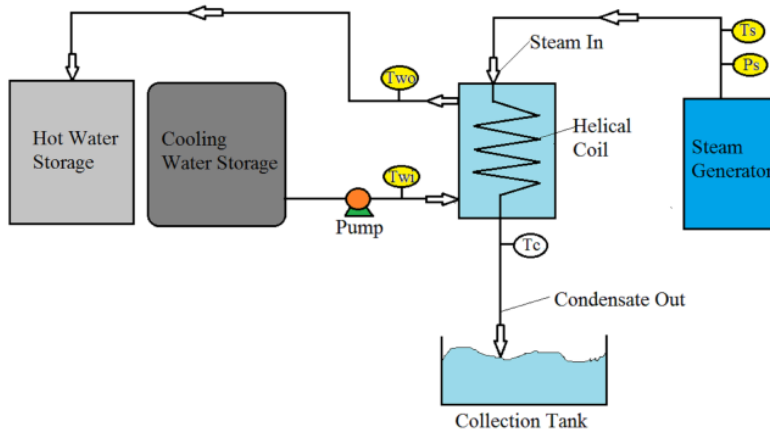
Figure 1. Schematic representation of experimental setup

Figure 2. Actual view of the experimental setup

In the water-steam loop, a steam generator (mini-boiler, Non-IBR, capacity: 27 kW) produced the steam continuously at known temperature T_s (°C) and pressure P_s (bar). The steam entered in the helical coil and condensation took place. Condensate volume flow rate and its temperature were measured at the outlet of the helical coil before the storage or collection tank. Cold water was circulated through shell where the coil is mounted. Volume flow rate and temperature of cold water circulated was monitored at the entrance of shell. The temperature of cold water at the outlet of the shell was measured. The condensate forming in the coil was directly measured by using calibrated measuring flask. The water steam inlet temperature was measured in 10 steps during experimental exercise in the range from T_s = 101.1°C to 109.9°C. Concurrently steam mass flow rate was also changed from $(\dot{m}_s m)$ = 0.00348 kg/s to $(\dot{m}_s)$ = 0.00528 kg/s. The selection of this range was restricted by the availability of the small size mini steam generator. Inlet parameters of steam are maintained constant for a sufficient long span of time in order to ensure thermal steady state as per the monitored parameters. The mass flow rate of cold water which was at 0.0834 ± 0.001 kg/s and inlet temperature 30±1°C was kept constant during experimentation.

Table 1
Physical dimension of the coil under investigation

Do	do	di	p	n	β
mm	mm	mm	mm		Deg
175	13.7	9.22	20	5.5	3
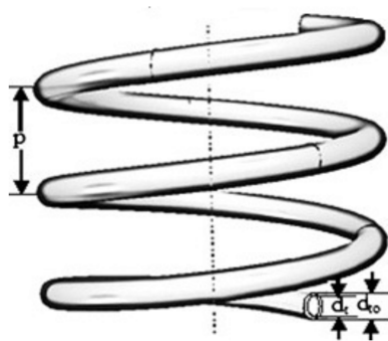
Figure 3. Schematic of helical coil

Uncertainty Analysis

Experimental uncertainty was estimated as described in Holman (2012) using Equation (3.2). In addition to this method, Salimpour (2017) was also referred. All uncertainties are tabulated in Table 2.

Table 2
Uncertainty in measurements

Measurements	Uncertainty
Heat Transfer rate	4%
Heat transfer coefficient	10%
Temperature sensor	±0.1

METHODS

Experimental heat transfer coefficient

Average inside coefficient of heat transfer is defined as:

$$h_i = \frac{Q_l}{A_i(T_s - T_{wi})} \qquad [1]$$

Where h_i is the average inside coefficient of heat transfer, A_i is inside tube surface area, T_s is saturation temperature of the steam and T_{wi} is temperature at inner side of wall of tube which given by

$$T_{wi} = T_{wo} + \frac{Q\, di \ln\left[\frac{d_o}{d_i}\right]}{2A_i k} \qquad [2]$$

Where T_{wo} is outside wall temperature which is measured directly using thermocouples.

The heat transfer rate, Q_1 released by the water steam inside the experimental test section (helical coil) can be obtained as per the following relation:

$$Q_1 = \dot{m}_{si} h_{si} - \dot{m}_{co} h_{co} \qquad [3]$$

Where $(\dot{m}_{si})$ is rate of flow of steam at inlet, h_{si} is enthalpy of the dry saturated steam at the inlet, $(\dot{m}_{co})$ is mass rate of condensate and hc_o is enthalpy of condensate. The amount of heat transfer rate absorbed by the cold water in the annular drum can be obtained as given below:

$$Q_2 = \dot{m}_{cw} C_{pw}\left(T_{cwi} - T_{cwo}\right) \qquad [4]$$

Where T_{cwi} and T_{cwo} are temperatures of cold water at entry and exit from shell respectively, $(\dot{m}_{cw})$ is flow rate of water mass through the annular space in the drum and C_{pw} is specific heat of water at prevailing temperature of hot water.

Prediction of the coefficient of condensation heat transfer

It is predicted using various equations available in the literature. The first equation used for comparison is given by Nusselt (1916). His equation is based on the basic approach for finding mean heat transfer coefficient using condensate film thermal resistance.

$$h = 0.943 \left(\frac{k_f^3 \rho_f^2 g\lambda}{(T_s - T_{wi})L\mu_f}\right)^{0.25} \qquad [5]$$

Next, chosen equation is proposed by Wang et al (1991) for mean heat transfer coefficient

$$\frac{h}{h_{Nu}} = \left[\frac{2L}{D}\right]^{\frac{cos\beta}{4}} \left(0.54 + 5.86 \times 10^{-3}\beta\right) \qquad [6]$$

Where h_{Nu} is Nusselt heat transfer coefficient given by the following equation.

$$h_{Nu} = 0.943\left[\frac{\rho_f^2\, g\, h_{fg}\, k_f^3\, sin\beta}{L\, \mu_f\, \Delta T}\right]^{0.25} \qquad [6a]$$

Next correlation (equation 7) is reported by Tandon et al. (1995) which take into consideration the change in the value of heat transfer coefficient for the change of annular flow or semi-

annular flow to wavy flow. Tandon's equation is used here as it satisfies the applicability limits of Reynolds number. In the present work, Reynolds number is greater than 30000, hence, this equation may be used for comparison.

$$\mathbf{Nu} = \frac{hd_i}{k_f} = 0.084 Pr_l^{1/3} \left(\frac{\lambda}{Cp_l \Delta T}\right)^{1/6} Re_v^{0.67}$$ [7]

For $Re_v > 3 \times 10^4$

Tandon's equation is based on liquid Prandtl Number, Reynolds number and Jacob number. Jacob Number is a dimensionless number which is used in phase change heat transfer calculations. Second, this equation is simple to analyse and compare.

RESULTS

The experimentation is performed on the single helical coil (physical dimensions are given in Table 1). The inlet mass flow rate and temperature of the steam is changed in 10 steps. The inside heat transfer coefficient is determined from the experimental observations and compared with three equations which are described above.

Energy Balance

Energy balance study was carried out to ascertain the reliability of experimental observations. It was done on the quantity of condensate collected and rate of flow of cooling water. Cold water heat gain was calculated using equation 4. The heat energy released by water steam was calculated using equation 3. It can be seen from Figure 4 that error of heat balance between cold water sensible heat and latent heat is less than 10%.
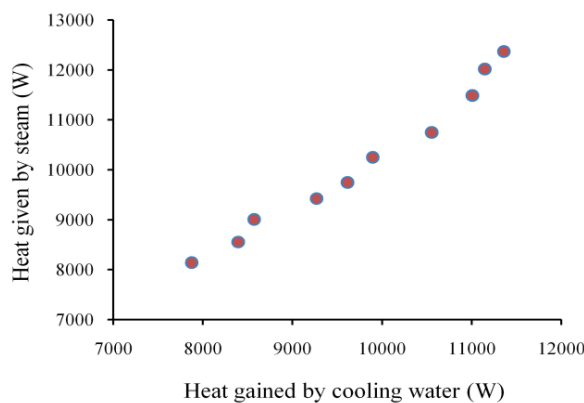
Figure 4. Results of energy balance for experimental observations

Comparison of experimental results with predicted values

The experimentally determined heat transfer coefficient was compared with predicted value using equations 5, 6 and 7. For the purpose of comparing the results, experimental heat transfer coefficient was calculated using equation 1 and it was found to be $h_{exp} = 19599\pm8460$ W/m^{2}K with a variation of 43% among the observations for mass flow rate ranging from 0.00348 kg/s to 0.00528 kg/s.

Comparison with Nusselt's Equation. The first equation selected for comparison was one which was theoretically determined by Nusselt (1916). By calculation, it is $h = 5377\pm411$ W/m^{2}K with fluctuations of 7.65%. Low value of heat transfer coefficient is inherent as the Nusselt equation is based on the assumption of stationary steam which does not take into consideration the effect of waviness in the condensate liquid film.

Comparison with Wang's Equation. Second equation chosen is equation 6 proposed by Wang et al (1991) which takes care for the inclination of the tube by incorporating angle of inclination 'β'. By calculating, the value of heat transfer coefficient is $h = 7087\pm542$ W/m^{2}K with fluctuations of 7.65%. Low value of heat transfer coefficient may be attributed to low inclination angle of helical tube in present case. Wang et al. had reported marginal effect on heat transfer coefficient in condensation at optimum angle of inclination which was 20 to 50 degrees.

Comparison with Tandon's Equation. The third and most important equation considered was proposed by Tandon et al (1995) and it predicted the value of $h = 22267\pm3871$ W/m^{2}K with fluctuations of 17.39%.

Influence of mass flow rate

The steam mass flow rate was changed during the experimentation in the range from 0.00348 kg/s to 0.00528 kg/s with an average increase of 5%. The influence of mass flow rate of steam ($\dot{m}_s$) on the heat transfer coefficient calculated using all three equations (Equation 5, 6 and 7) is illustrated in Figure 5. This graph shows that there is significant increase in the heat transfer coefficient with the rise in flow rate. This is observed in case of heat transfer coefficient determined experimentally and predicted by using Tandon's equation 7. On the other hand, equations 5 and 6, under-predicts the values of heat transfer coefficient in case of helical coil whereas the equation proposed by the Tandon et al (1995) predicts the values on average 18% deviations. The large deviation in the predicted values from the experimental heat transfer coefficient is justified by the presence of secondary flow in the coiled tube. In helically coiled tubes, secondary flow is due to curvature of the tube. During condensation of vapour, condensed liquid is pushed to wall of the tube and this increases interfacial shear area between vapour and liquid. This shear causes the waviness on the liquid surface and increases the heat transfer rate. The same phenomenon was reported by Mosaad et al. (2009) and Salimpour et al. (2017).

There is significant improvement in the heat transfer coefficient in helical coiled tube which is one of the biggest advantages of using helical coiled tube as heat exchangers. Here, authors have attempted to correlate the experimental data of the helical coil with the equations available in literature for straight tube.
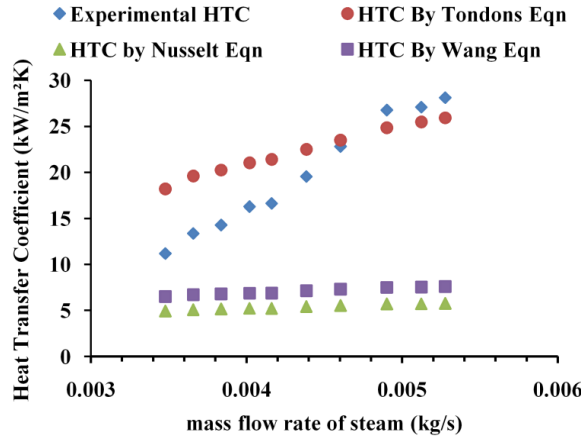
Figure 5. Influence of vapor mass flow rate on heat transfer coefficient

Correlation of the experimental data

The experimental data is correlated as per the coordinates suggested by the Equation 8 and are used to correlate the data obtained experimentally as reported in the McAdams (1958).

$$\text{h}\left(\frac{\mu_f^2}{k_f^3 \rho_f^2 g}\right)^{1/3} = a\left(\frac{4\tau}{\mu_f}\right)^b \tag{8}$$

The constants 'a' and 'b' in the above equation are determined by using the least square power-law fitting technique, based on the experimental data collected. A log-log plot of $\left(\frac{4\tau}{\mu_f}\right)$ against $\text{h}\times\left(\frac{\mu_f^2}{k_f^3 \rho_f^2 g}\right)^{1/3}$ is illustrated in Figure 6. The trend line on the plot is the curve fitting line through the experimental readings, using the least square power law fitting technique. The Pearson's correlation coefficient is 0.981.

Following correlation is obtained from the graph.

$$\text{h}\left(\frac{\mu_f^2}{k_f^3 \rho_f^2 g}\right)^{1/3} = 3\times10^{-7}\left(\frac{4\tau}{\mu_f}\right)^{1.967} \tag{9}$$

the values of constants 'a' and 'b' are 3×10^{-7} and 1.967 respectively.
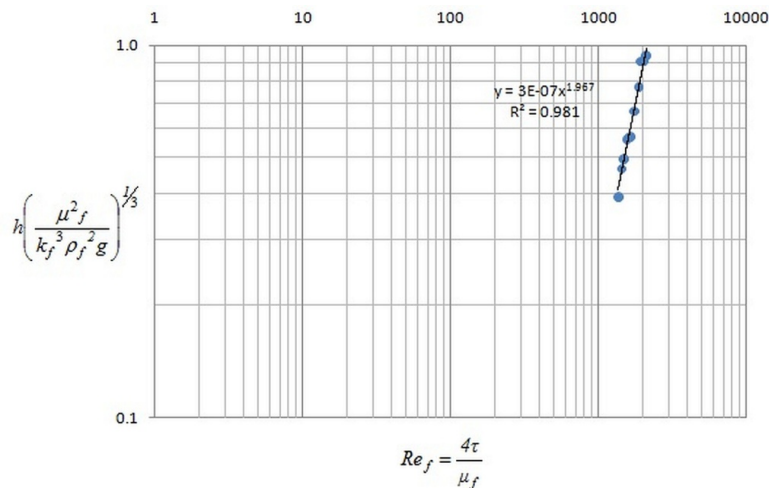
Figure 6. Representation of variation in magnitude of heat transfer coefficient as a function of Reynolds Number

CONCLUSION

An experimental investigation was carried out on the 175 mm coil diameter vertical helical coiled tube under variable mass flux conditions. Mass flux was varied in 10 steps. Experimental heat transfer coefficient was evaluated from observations recorded. The experimental heat transfer coefficient was compared with three equations reported in the literature for prediction of heat transfer coefficient. An attempt had been made to calculate the heat transfer coefficient using Jacob number in the equation of heat transfer coefficient for in-tube condensation in the vertical tube-in-shell heat exchanger. It was observed that basic Nusselt equation under-predicts the condensation heat transfer coefficient. This is quite obvious as the Nusselt equation is based on simplified assumptions of stationary vapour and laminar flow and neglects the shear stress. A similar trend is observed for the Wang equation. Tandon's equation has less deviation compared with the other two equations. Experimentally determined value of heat transfer coefficient is more on account of the secondary flow pattern in the the helical coil which enhances the condensation heat transfer coefficient significantly. Based on the experimental data new correlation is developed.

ACKNOWLEDGEMENT

REFERENCES

Akhavan-Behabadi, M. A., Kumar, R., & Mohseni, S. G. (2007). Condensation heat transfer of R-134a inside a microfin tube with different tube inclinations. *International Journal of Heat and Mass Transfer, 50*(23), 4864-4871.

Bae, S., Maulbetsch, J. S., & Rohsenow, W. M. (1969). *Refrigerant forced-convection condensation inside horizontal tubes.* Cambridge, Mass.: MIT Heat Transfer Laboratory.

Berenson, P. J., Schuster, J. R., & Soliman, M. (1968). A general heat transfer correlation for annular flow condensation. *Journal of Heat Transfer 90*(2), 267-274.

Chato, J. C. (1962). Laminar condensation inside horizontal and inclined, tubes. *ASHRAE Journal, 4*, 52–60.

Colorado, D., Papini, D., Hernández, J. A., Santini, L., & Ricotti, M. E. (2011). Development and experimental validation of a computational model for a helically coiled steam generator. *International Journal of Thermal Sciences, 50*(4), 569-580.

Holman, J. P. (2012). *Experimental Methods for Engineers* (8th Ed.). New York: McGraw Hill Publications.

Kern, D. Q., (1965). *Process Heat Transfer.* New York: McGraw Hill Publications.

Li, S. H. A. O., Han, J. T., Su, G. P., & Pan, J. H. (2007). Condensation heat transfer of R-134A in horizontal straight and helically coiled tube-in-tube heat exchangers. *Journal of Hydrodynamics, Ser. B, 19*(6), 677-682.

McAdams, W. H. (1958). *Heat transmission.* New York: McGraw Hill Publications.

Mosaad, M. E. S., Al-Hajeri, M., Al-Ajmi, R., & Koliub, A. M. (2009). Heat transfer and pressure drop of R-134a condensation in a coiled, double tube. *Heat and Mass Transfer, 45*(8), 1107-1115.

Mozafari, M., Akhavan-Behabadi, M. A., Qobadi-Arfaee, H., & Fakoor-Pakdaman, M. (2015). Condensation and pressure drop characteristics of R600a in a helical tube-in-tube heat exchanger at different inclination angles. *Applied Thermal Engineering, 90*, 571-578.

Nusselt, W. (1916). The surface condensation of water vapor, Zeitschr. *V.D.I,* (60), 541-569.

Nitheanandan, T., & Soliman, H. M. (1993). Influence of tube inclination on the flow regime boundaries of condensing steam. *The Canadian Journal of Chemical Engineering, 71*(1), 35-41.

Owhadi, A., Bell, K. J., & Crain, B. (1968). Forced convection boiling inside helically-coiled tubes. *International Journal of Heat and Mass Transfer, 11*(12), 1779-1793.

Prabhanjan, D. G., Rennie, T. J., & Raghavan, G. V. (2004). Natural convection heat transfer from helical coiled tubes. *International Journal of Thermal Sciences, 43*(4), 359-365.

Saffari, H., & Naziri, V. (2010). Theoretical modeling and numerical solution of stratified condensation in inclined tubes. *Journal of Mechanical Science and Technology, 24*(12), 2587-2596.

Salimpour, M. R., Shahmoradi, A., & Khoeini, D. (2017). Experimental study of condensation heat transfer of R-404A in helically coiled tubes. *International Journal of Refrigeration, 74*, 582-589.

Tanasawa, I. (1990). Recent progress of Japanese research on condensation heat transfer. *Applied Mechanics Reviews, 43*(1), 1-11.

Tandon, T. N., Varma, H. K., & Gupta, C. P. (1995). Heat transfer during forced convection condensation inside horizontal tube. *International Journal of Refrigeration, 18*(3), 210-214.

Vashisth, S., Kumar, V., & Nigam, K. D. (2008). A review on the potential applications of curved geometries in process industry. *Industrial and Engineering Chemistry Research, 47*(10), 3291-3337.

Wang, B. X., & Du, X. Z. (2000). Study on laminar film-wise condensation for vapor flow in an inclined small/mini-diameter tube. *International Journal of Heat and Mass Transfer, 43*(10), 1859-1868.

Wang, J. C. Y., & Ma, Y. (1991). Condensation heat transfer inside vertical and inclined thermosiphon. *Journal of Heat Transfer, 113*(3), 777-780.

Wongwises, S., & Polsongkram, M. (2006). Condensation heat transfer and pressure drop of HFC-134a in a helically coiled concentric tube-in-tube heat exchanger. *International Journal of Heat and Mass Transfer, 49*(23-24), 4386-4398.

Yu, J., Chen, J., Li, F., Cai, W., Lu, L., & Jiang, Y. (2018). Experimental investigation of forced convective condensation heat transfer of hydrocarbon refrigerant in a helical tube. *Applied Thermal Engineering, 129*, 1634-1644.

SCIENCE & TECHNOLOGY

Modified Wiener Filter for Restoring Landsat Images in Remote Sensing Applications

Kalaivani K[1,2]* and Asnath Victy Phamila Y[1]

[1]*School of Computing Science and Engineering, VIT University, Chennai, India*
[2]*Department of Computer Science and Engineering, VELS University, Chennai, India*

ABSTRACT

Images are inherently affected by different noises such as, speckle, salt and pepper, Gaussian, poisson, and white noise during its acquisition or transmission, hence, their true intensities cannot be reflected from the pixel values of an image. The main focus of this work is on designing a filtering method that preserves the edges of an image while removing the noise. The modified wiener filter is proposed to denoise satellite images. A detailed study on different noises and the filtering techniques, such as mean, median, wiener, Gaussian filter and the proposed modified wiener filter is done. The performance of these filtering methods is assessed by the image quality metrics, such as mean squared error, peak signal to noise ratio, and correlation coefficient. The results show the choice of a filter for denoising depends on the type of noise present in the image. The proposed modified wiener filter performs relatively well for most of the noise models compared with the existing linear and non-linear filtering methods. This technique can be widely used during the pre-processing of satellite images in remote sensing applications.

Keywords: Image denoising, filtering, remote sensing applications, peak signal to noise ratio, satellite images, Modified wiener filter

INTRODUCTION

Images are prone to degradation by noise, distortion and artefacts from various sources either at the time of image acquisition or transmission which are mainly due to improper imaging system, inappropriate focusing, movement etc (Jain & Tyagi, 2014). Employing feature extraction and classification algorithms on noisy images paves way for inaccurate conclusion. Hence, these noises in images must be removed in the pre-processing phase. The ratio of the

corrupted pixels in the image decides the quantification of the noise. Image denoising is an essential task in many applications where the image is restored by retaining as much as possible information with the use of filtering techniques (Ansari & Budhhiraju, 2016; Wang, Ziou, Armenakis, Li, & Li, 2005). Hence, reliable filtering techniques are required to restore the image with the essential information for accurate image analysis.

A number of noise removal algorithms is available and a detailed study on different noises and the filters are required to choose a suitable technique for any particular application. Different noises that may present in images include impulse noise, Gaussian noise, Poisson noise, Speckle noise and so on (Bovik, 2010). The impulse noise occurs mainly due to the sudden and severe disturbances in an image signal. This noise is also referred as salt and pepper noise because the majority of the pixels are in black and white. For an 8-bit image, the pixel intensity value 0 and 255 refers to the pepper and salt noise respectively.

The Gaussian noise model is additive in nature and follows Gaussian distribution. In other words, it is referred to as a statistical noise having the probability density function close to normal distribution. The intensity of each pixel in the noisy image is the sum of the Gaussian distributed noise value and the true pixel value. The communication channels suffer from Gaussian noise because of the thermal vibrations, particularly in telecommunication and networking.

Poisson or Photon noise refers to the uncertainty in the measurement of light during the image acquisition. In digital and film-based sensors, the photoelectric effect on image scene brightness leads to photon noise. Speckle noise is a multiplicative noise triggered by consistent processing of backscattered signals from several objects of focus. It is a signal dependent form of noise whose magnitude is associated with the value of the original pixel. This noise follows a gamma distribution. Images captured or acquired through laser, acoustics, Thematic Mapper (TM) and Synthetic Aperture Radar (SAR) results in multiplicative noise. A generic radiometric calibration cannot be used optimally for detecting targets in various applications; hence, a specific filtration based on the requirement is essential.

Surface water is one of the invaluable assets for human survival and social development. True data about the spatial distribution of surface water is essential for various applications, such as water resource assessment, environment monitoring, climate modelling, water survey and management among others. Water features can be extracted from different bands of Landsat imagery. Hence, the images must be free from noises to further detect and map the spatiotemporal changes using various surface water extraction techniques. Lake Urmia has a total area of approximately 51,876 km^2 in the north west of Iran (Eimanifar & Mohebbi, 2007). The lake has been in a catastrophe for recent years because of decreasing surface water and increasing salinity.

This study aims to enhance and restore the Landsat TM images of Lake Urmia. The images are taken from USGS (US Geological Survey) Global Visualisation Viewer. The noise is added to the test image in a controlled manner to obtain a corrupted image. The original and noisy images of Lake Urmia are shown in Figures 1(a-f).
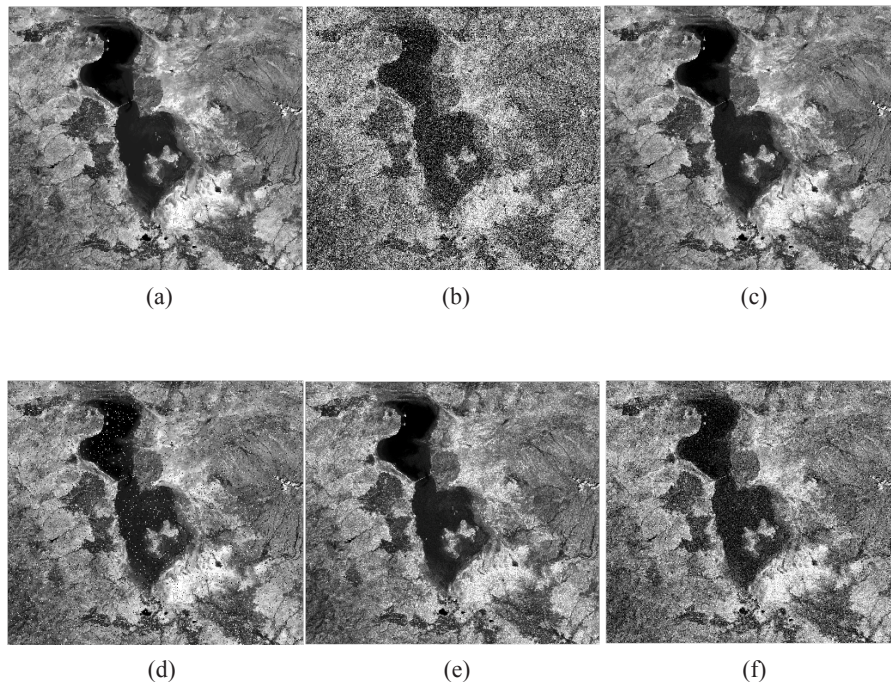
Figure 1. (a) Lake Urmia, Landsat 5 Original image, (b) Image with Gaussian Noise, (c) Image with Poisson Noise, (d) Image with Salt and Pepper Noise, (e) Image with Speckle Noise, (f) Image with White Noise

FILTERING TECHNIQUES

Image filtration is also known as image denoising, an operation to clean the image from a degraded noise affected image. Filtering is the process of altering pixel intensity values to know certain image characteristics. Filtering is done on images for various purposes such as:i) to enhance by improving the contrast; ii) to smoothen by removing the noises; and iii) to do template matching by detecting the known patterns.

Filtering is a neighbourhood process, in which the value of every pixel in the output image is found by applying an algorithm to all the neighbourhood pixel intensities of the corresponding input pixel. Hence, the boundary pixels require additional rows and columns to be padded before applying the filter. The amount of padding depends upon the kernel size and generally it can be done in four different ways known as zero padding, reflection, replication and extrapolation. Zero padding is done by placing zeros in the outer padded region. Pixel replication is easily done by copying the border pixel intensities and placing the same in the outer padded region. In reflection padding, the outer padded region is filled with mirror reflection of boundary to inner pixels based on the kernel size. Extrapolation can be done by padding with a linear extension of the outmost two border values.

The filtering techniques are broadly classified as linear and non-linear (Plataniotis & Venetsanopoulos, 2013). Linear filtering is a process in which the value of an output pixel is a linear combination of the pixel intensities of the input pixel's neighbourhood. Linear filtering techniques removes noise within a short span of time but it does not preserve edges. Non-linear

filters are the vice versa of linear filters, wherein their output is not a linear function of their input and these techniques are good at preserving the edges of an original image.

The corrupted image is denoised using various existing filters, such as mean, median, Gaussian, wiener and the proposed modified wiener filter. The performance of the denoised image is measured based on the quality assessment metrics such as Peak Signal to Noise Ratio (PSNR) and Correlation Coefficient (CORR) (Kalaivani & Phamila, 2016). The filters achieving the highest PSNR, CORR and less computational complexity can be assumed as the best filter for a particular noise model. Peak Signal to Noise Ratio measures the quality of the filtered image and the value will be high when the resultant image is identical to the original image. The PSNR is calculated using the formula:

$$PSNR = 10 \log_{10} \frac{r^2}{MSE}$$

where MSE is the Mean Squared Error and r is the peak value of the original / reference image. Higher value of PSNR indicates better filtration. The correlation coefficient (Roche, Malandain, Pennec, & Ayache, 1998) represents the similarity between the corrupted image and original image with respect to the pixel intensity. It can be computed using the formula:

$$CORR(A, B) = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (A(m,n) - \mu_A) (B(m,n) - \mu_B)}{\sqrt{(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (A(m,n) - \mu_A))^2)(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (B(m,n) - \mu_B)^2)}}$$

where μ_A and μ_B are the mean values of filtered and original images respectively. The value of correlation coefficient lies between 0 and 1, if the value is 0, then it indicates the complete loss of information and 1 indicates maximum similarity.

Mean Filter

Mean filter is an averaging linear filter, which uses a spatial mask over each pixel in an image. Usually the size of the mask will be odd so that the central pixel can be found easily. The value of the pixels which are grouped under the mask is averaged and this is known as the centre pixel intensity (Verma & Jahid, 2013). This averaging eliminates pixel values which are unrepresentative of their surroundings. Mean filter is a convolution filter as they are represented using a matrix multiplication. It works as follows:

i.   Pad the outer region of boundary pixels.

ii.   Apply a mask over a pixel of a noisy image.

iii. Find an average: add the pixel intensities of all the pixels which fall under the mask and divide by the total number of elements in the mask.

iv.   Assign the average value to the centre pixel of the kernel.

v.   Repeat the steps ii, iii and iv for all the pixels of an image.

It reduces the amount of intensity variation between adjacent pixels thereby restoring the affected image. An averaging filter of size 3×3 is employed to filter the multidimensional array of the input noisy images of Lake Urmia (Figure 1 b-f). The PSNR and Correlation coefficients are computed for all the filtered images of different noises and their performance are shown in Figure 2.
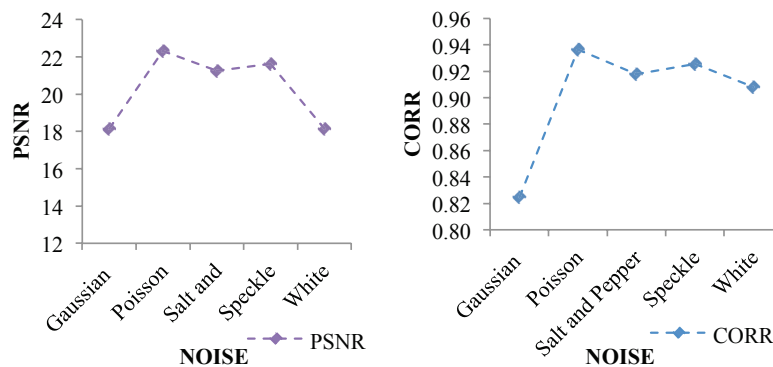
Figure 2. PSNR and CORR results of Mean filter

Median Filter

Median filter is a non-linear robust filter, which preserves the edge details than the mean filter. This filtering technique also uses a mask over each pixel and replaces the centre value with the median of its local neighbourhood pixel intensities. The pixel intensities, which fall under the mask, are sorted and the middle pixel value is considered as the median. Instead of replacing the centre pixel value with the average of neighbouring pixel values, it replaces with the median (Chen, Church, & Rice, 2008). The working principles are as follows:

i.   Pad the outer region of boundary pixels.

ii.   Apply a mask over a pixel of a noisy image.

iii. Sort the pixel intensities of all the pixels which fall under the mask.

iv.   Find the median and assign it to the centre pixel of the kernel.

v.   Repeat the steps ii, iii and iv for all the pixels of an image.

Salt and pepper noise is highly reduced by median filters without any loss of fine edge details (Chan, Ho, & Nikolova, 2005). The input matrix of all noisy images is median filtered in two dimensions using the default filter size 3×3. This is done for all the considered noisy images (Figure 1b-f) and their performance based on PSNR values and correlation coefficient is shown in Figure 3.
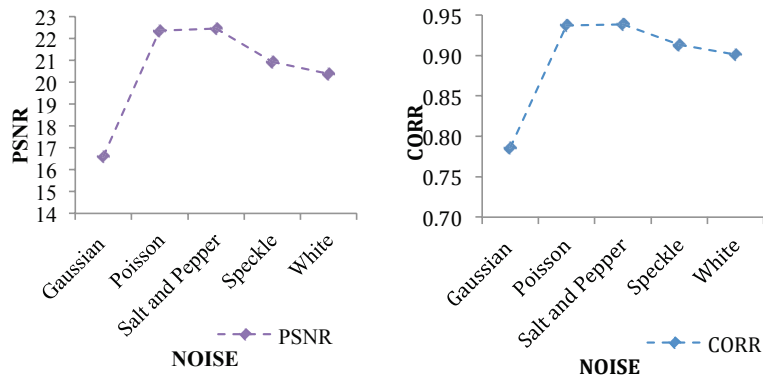
Figure 3. PSNR and CORR results of Median filter

Gaussian Filter

It is a linear smoothing filter with the weight chosen according to the form of a Gaussian function (Kumar, Murugan, & Rajalakshmi, 2015). A Gaussian filter can be generated from PASCAL's triangle, and these filters can be applied recursively to create a Gaussian pyramid. The 3 × 3 kernel / mask is created from third row of the Pascal's triangle. It is a particular class of averaging filter. The filter works as follows:

i. Pad the outer region of boundary pixels.

ii. Generate and apply a Gaussian mask over a pixel of a noisy image.

iii. Perform an element-by-element multiplication with this pixel neighbourhood and sum up all the elements.

iv. Assign the resultant sum to the centre pixel of the kernel.

v. Repeat the steps ii, iii and iv for all the pixels of an image.

The test images are filtered with a 2-D Gaussian smoothing kernel with 0.5 as standard deviation. The noisy images (Figure 1 b-f) are filtered using the Gaussian kernel and their performance based on PSNR and correlation coefficient is shown in Figure 4.
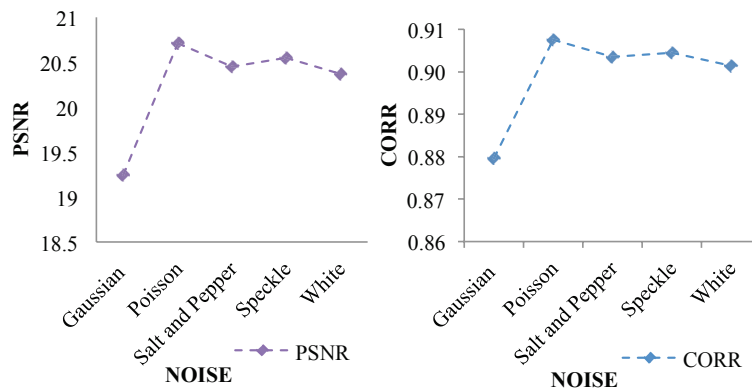
Figure 4. PSNR and CORR results of Gaussian filter

Wiener Filter

It is an additive noise removal filtering process. The low pass wiener filters use the pixel wise adaptive method based on facts assessed from a local neighbourhood of each pixel. It computes the local mean and variance in its process of filtering. It performs deconvolution by inverse filtering and removes the noise by a compression operation (Jiang, Yang, Wang, & Hou, 2014; Yoo & Ahn, 2014). The wiener filter works as follows:

i.   Estimate the power spectra (Fourier transform of the auto correlation function) of the original and noisy image.

ii.   Apply a mask over a pixel of a noisy image.

iii. Calculate the local mean (μ) and variance (σ^2).

iv.   Compute the new pixel value using the mean, variance and noise power.

v.   Repeat the steps ii to iv for all pixels of a noisy image.

Wiener filtering acts as the optimal trade-off between noise smoothing and inverse filtering. The wiener filter is applied to all noisy images (Figure 1(b-f)) and the performance is evaluated based on the PSNR values and correlation coefficient obtained. The results are shown in Figure 5.
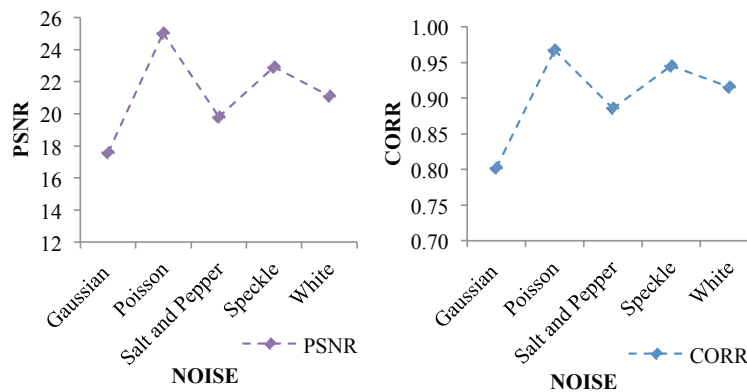
Figure 5. PSNR and CORR results of Wiener filter

Modified Wiener Filter

The proposed modified wiener filter is an additive noise removal technique where median is computed for all the pixels using a 3×3 filter. The resultant matrix is further processed similar to the wiener filter. This non-linear adaptive spatial filter is proposed with the aim to combine the features of median filter and wiener filter reciprocally nullifying their defects. The working of the proposed filter is shown as flow diagram in Figure 6. The modification of the wiener filter is very significant due to the non-linear behaviour of the median operator in adaptive contest. The main objective is to preserve the edges without modifying the shape and structure while reducing the noise in an image.
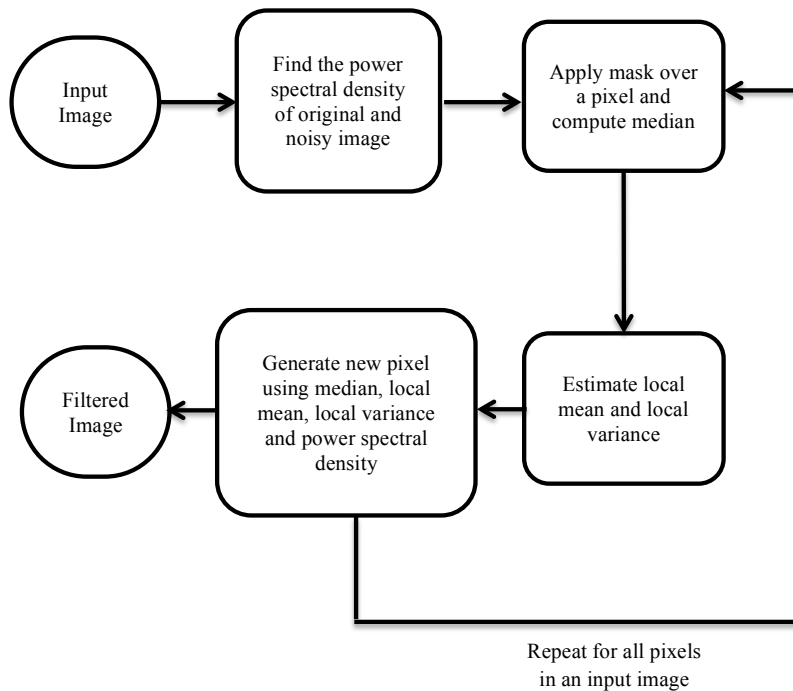
Figure 6. Flow diagram of Modified Wiener filter

Calculating median instead of mean reduces noise without blurring the edges and therefore, this edge preserving nature makes the proposed filter useful in applications where edge blurring is undesirable. The algorithm of modified wiener filter is as follows:

i.   Estimate the power spectral density (Fourier transform of the auto correlation function) of the original and noisy image.

ii. Apply mask over a pixel of the noisy image.

iii. Sort the pixel intensities of all the pixels which fall under the mask.

iv. Find the median (med) and assign it to the centre pixel of the mask.

v.   Estimate the local mean (μ) and variance (σ^2)

   a. $Mean\ (\mu) = \frac{1}{mn}\sum_{mn} A(m,n)$, where m and n represent the row and column of an image A.

   b. $Variance\ (\sigma^2) = \frac{1}{mn}(A(m,n) - \mu)^2$, where m and n represent the row and column of an image A.

vi. Compute the new pixel value *B (m, n)*

   $B\ (m,n) = med + \frac{\sigma^2 - v^2}{\sigma^2}\ (A(m,n) - med)$, where v^2, med represents the noise variance and median value of the local window respectively.

vii. Repeat the steps ii to vi for all pixels of a noisy image.

The proposed filter is applied on different types of noisy images and the resultant filtered images are shown in Figure 7(a-j).
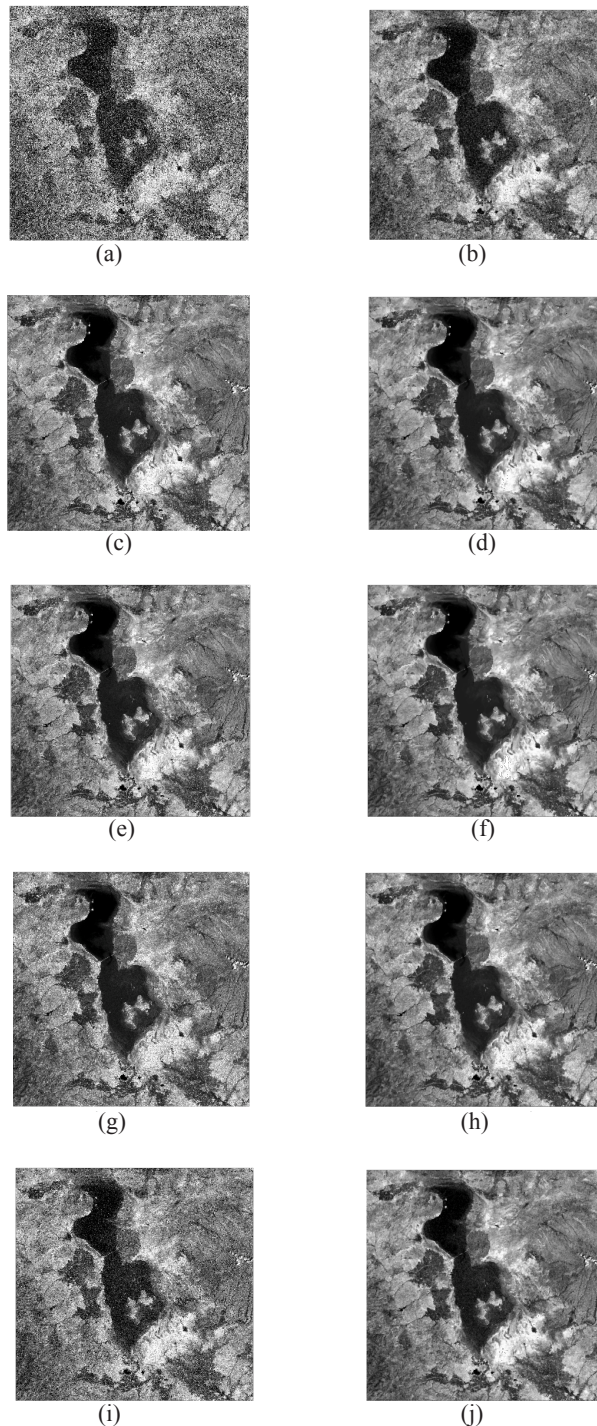
Figure 7. (a) Image with Gaussian noise, (b) Modified wiener filter on Gaussian noise affected image, (c) Image with Poisson noise, (d) Modified wiener filter on Poisson Noise affected image, (e) Image with Salt and Pepper noise (f) Modified wiener filter on Salt and Pepper noise affected image, (g) Image with Speckle noise, (h) Modified wiener filter on Speckle noise affected image, (i) Image with White noise, (j) Modified wiener filter on White noise affected image

The performance of modified wiener filter is evaluated based on PSNR values and correlation coefficient of the processed image in the presence of various noises (Figure 1(b-f)) and the results are represented in Figure 8.
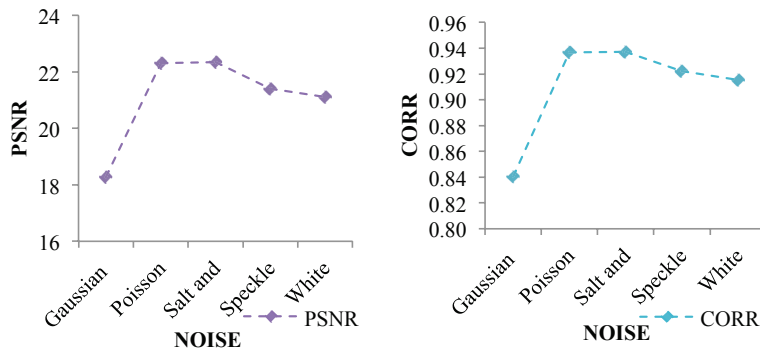
Figure 8. PSNR and CORR results of Modified Wiener filter

RESULTS AND DISCUSSION

The performance and appropriate use of a particular filter for any application is still an ongoing research (Darus et al., 2017; Verma & Jahid, 2013). The mean filter is the simplest linear spatial filter, which assigns the average value of the pixels in the window (mask) as the centre value. A new denoised value is computed for each pixel by moving the window across the pixels of an image, hence, this filter is called as moving average filter. This averaging filter is good at removing the grain noise in an image. Because of the averaging of neighbouring pixels, local variations caused by grain noise are greatly reduced.

Median filter is the non-linear spatial filter, which can be widely used to remove impulsive noises like salt and pepper noise effectively. It can be employed in applications where there is a necessity to preserve edges while removing noise (Ismail, Adnan, Malek, & Bebakar, 2008). Median filter is more effective in preserving edges, the reason is that it is less sensitive to extreme values called outliers (Chen et al., 2008) when compared with mean. The median filter erodes the edges of isolated spots when the window size is large while the Gaussian filter blurs the image and remove noises and some essential information.

Images from US Geological survey, such as Lake Turkana, Huang he Delta (Yellow River), Lake Urmia, Capecod USA and the images from Remote Pixel Earth Observation Dataset are considered for evaluating the performance of existing and modified wiener filters in the presence of various noises. The PSNR values of different filtration techniques on various noises of study image Lake Urmia are shown in Table 1 and Figure 9.

Table 1
PSNR values of different filtration methods on various noises

Noise	Mean	Median	Gaussian	Wiener	Modified Wiener
Gaussian	18.1445	16.5898	19.2459	17.5559	18.2747
Poisson	22.3117	22.3462	20.7104	25.0223	22.318
Salt and Pepper	21.2359	22.4474	20.4513	19.8091	22.355
Speckle	21.628	20.9377	20.5515	22.938	21.4071
White	18.1445	20.3876	20.3686	21.1152	21.1168
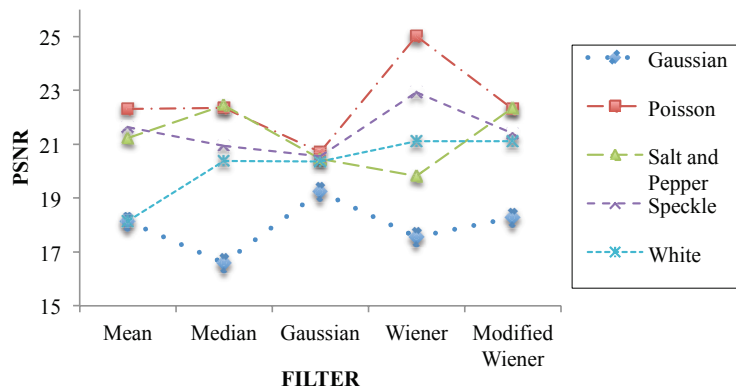
Figure 9. Performance of various filters based on PSNR

Linear filters tend to blur edges and other fine details of an image whereas non-linear filters can preserve edges and they are very effective at removing impulsive (salt and pepper) noise. Wiener filter is linear spatial filter, which is adaptive to the local variance. When the variance is small, its performance is smoother and vice versa, i.e. variance is indirectly proportional to smoothness. A single pixel with an unrepresentative value can influence the mean value of all the pixels in its neighbourhood significantly. The wiener filter especially for large windows causes dilation of edges and small peaks due to the mean computation, whereas the modified wiener preserves unaltered edges by computing the median value in the filtration process. The correlation coefficient between filtered image and the original image is computed and the results are shown in Table 2 and Figure 10.

Table 2
Correlation coefficient values between filtered and the original image

Noise	Mean	Median	Gaussian	Wiener	Modified Wiener
Gaussian	0.8248	0.7862	0.8796	0.8025	0.8404
Poisson	0.9367	0.937	0.9076	0.9668	0.9367
Salt and Pepper	0.9178	0.9385	0.9035	0.8854	0.9373
Speckle	0.9254	0.9132	0.9045	0.9452	0.9222
White	0.9083	0.901	0.9014	0.9151	0.9155
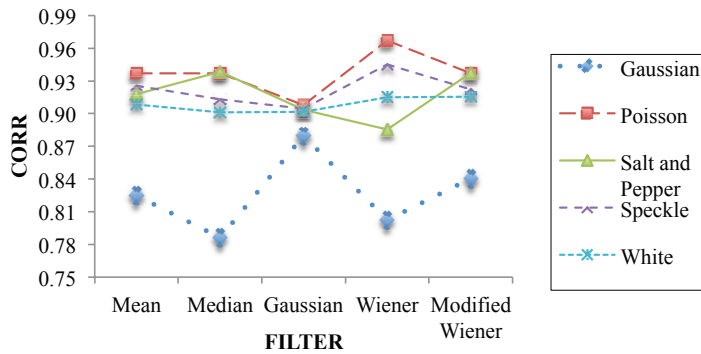
Figure 10. Performance of various filters based on CORR

Based on the results, it is found that the Mean and Gaussian filters are good at removing the Poisson noise but these techniques blur the image affecting the feature localisation. From the observations of PSNR value and the Correlation coefficient, it is clearly understood that a single filter cannot be used for various noise models in remote sensing images. The best filter for each noise model is thus found based on performance metrics, such as Peak signal to Noise Ratio and Correlation Coefficient. It is shown in Table 3.

Table 3
Image noise and suitable filter

Noise Model	Best Filter
Gaussian	Gaussian, Modified Wiener Filter
Poisson	Wiener, Median Filter, Modified wiener Filter
Salt and Pepper	Median, Modified Wiener Filter
Speckle	Wiener, Mean Filter
White	Modified Wiener, Wiener Filter

The Landsat images of Lake Turkana, Huang he Delta (Yellow River), Capecod USA obtained from the US Geological Survey and images from Earth Observation remote pixel dataset are used as test images for analysing the performance of various filters. It clearly shows the proposed modified wiener filter performs invariably better in most of the noise models.

CONCLUSION

The performance of various noise models and filtering techniques has been analysed for image enhancement and restoration. The similarity metrics such as peak signal to noise ratio and correlation coefficient are computed for all the filtered images on various noise models. The results show that a filter has to be chosen based on the application and type of data required for further processing in image analysis. The proposed modified wiener filter yields better performance compared with other linear and non-linear filters on various satellite images. The

edges are well preserved by the use of median operator in modified wiener filter. However, the computational complexity of the image reconstruction based on modified wiener filter need to be further studied. The resultant filtered images can be used to investigate the spatiotemporal changes that have occurred in recent decades using remote sensing technology.

REFERENCES

Ansari, R. A., & Budhhiraju, K. M. (2016). A comparative evaluation of denoising of remotely sensed images using wavelet, curvelet and contourlet transforms. *Journal of the Indian Society of Remote Sensing, 44*(6), 843–853.

Bovik, A. C. (2010). *Handbook of image and video processing.* USA: Academic press.

Chan, R. H., Ho, C. W., & Nikolova, M. (2005). Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *IEEE Transactions on Image Processing, 14*(10), 1479–1485. http://doi.org/10.1109/TIP.2005.852196

Chen, Y., Church, J., & Rice, S. (2008). A Spatial Median Filter for noise removal in digital images. In *Southeastcon, 2008, IEEE* (pp. 618-623). IEEE. http://doi.org/10.1109/SECON.2008.4494367

Darus, M. S., Sulaiman, S. N., Isa, I. S., Hussain, Z., Tahir, N. M., & Isa, N. A. M. (2017). Modified hybrid median filter for removal of low density random-valued impulse noise in images. In *Proceedings - 6th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2016*, (pp. 528–533). IEEE. http://doi.org/10.1109/ICCSCE.2016.7893633

Eimanifar, A., & Mohebbi, F. (2007). Urmia Lake (Northwest Iran): A brief review. *Saline Systems, 3*(1), 1–8. http://doi.org/10.1186/1746-1448-3-5

Ismail, M. H., Adnan, I., Malek, A., & Bebakar, S. (2008). Detecting and quantifying degraded forest land in Tanah Merah forest district, Kelantan using spot-5 image. *Pertanika Journal of Tropical Agricultural Science, 31*(1), 11–17.

Jain, P., & Tyagi, V. (2016). A survey of edge-preserving image denoising methods. *Information Systems Frontiers, 18*(1), 159-170.

Jiang, B., Yang, A., Wang, C., & Hou, Z. (2014). Comparison of motion-blurred image restoration using Wiener Filter and Spatial Difference technique. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 7*(2), 11–22.

Kalaivani, K., & Phamila, Y. A. V. (2016). Analysis of image fusion techniques based on quality assessment metrics. *Indian Journal of Science and Technology, 9*(31), 1-8. http://doi.org/10.17485/ijst/2016/v9i31/92553

Kumar, T. G., Murugan, D., & Rajalakshmi, K. (2015). Image enhancement and performance evaluation using various filters for IRS-P6 Satellite Liss IV remotely sensed data. *Geofizika, 32*(2), 179-189. http://doi.org/10.15233/gfz.2015.32.11

Plataniotis, K. N., & Venetsanopoulos, A. N. (2013). *Color image processing and applications.* New York, NY: Springer Science & Business Media.

Roche, A., Malandain, G., Pennec, X., & Ayache, N. (1998). The correlation ratio as a new similarity measure for multimodal image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 1115-1124). Springer, Berlin, Heidelberg.

Verma, R., & Jahid, A. (2013). A comparative study of various types of image noise and efficient noise removal techniques. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(10), 617–622.

Wang, Z., Ziou, D., Armenakis, C., Li, D., & Li, Q. (2005). A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing, 43*(6), 1391–1402. http://doi.org/10.1109/TGRS.2005.846874

Yoo, J. C., & Ahn, C. W. (2014). Image restoration by blind-Wiener filter. *IET Image Processing, 8*(12), 815–823.

GPU-based Optimization of Pilgrim Simulation for Hajj and Umrah Rituals

Abdur Rahman Muhammad Abdul Majid[1], Nor Asilah Wati Abdul Hamid[1]*, Amir Rizaan Rahiman[1] and Basim Zafar[2]

[1]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Custodian of the Two Holy Mosques Institute for Hajj Research, Umm Al-Qura University, Makkah, Saudi Arabia*

ABSTRACT

Tawaf ritual performed during Hajj and Umrah is one of the most unique, large-scale multi-cultural events in this modern day and age. Pilgrims from all over the world circumambulate around a stone cube structure called Ka'aba. Disasters at these types of events are inevitable due to erratic behaviours of pilgrims. This has prompted researchers to present several solutions to avoid such incidents. Agent-based simulations of a large number of pilgrims performing different the ritual can provide the solution to obviate such disasters that are either caused by mismanagement or because of irregular event plans. However, the problem arises due to limited parallelisation capabilities in existing models for concurrent execution of the agent-based simulation. This limitation decreases the efficiency by producing insufficient frames for simulating a large number of autonomous agents during Tawaf ritual. Therefore, it has become very necessary to provide a parallel simulation model that will improve the performance of pilgrims performing the crucial ritual of Tawaf in large numbers. To fill in this gap between large-scale agent-based simulation and navigational behaviours for pilgrim movement, an optimised parallel simulation software of agent-based crowd movement during the ritual of Tawaf is proposed here. The software comprises parallel behaviours for autonomous agents that utilise the inherent parallelism of Graphics Processing Units (GPU). In order to implement the simulation software, an optimized parallel model is proposed. This model is based on the agent-based architecture which comprises agents having a reactive design that responds to a fixed set of stimuli. An advantage of using agents is to provide artificial anomaly to generate heterogeneous movement of the crowd as opposed to a singular movement which is unrealistic. The purpose is to decrease the execution time of complex behaviour computation for each agent while simulating a large crowd of pilgrims at increased frames per second (fps). The implementation utilises CUDA

Abdur Rahman Muhammad Abdul Majid, Nor Asilah Wati Abdul Hamid, Amir Rizaan Rahiman and Basim Zafar

(Compute Unified Device Architecture) platform for general purpose computing over GPU. It exploits the underlying data parallel capability of an existing library for steering behaviours, called OpenSteer. It has simpler behaviours that when combined together, produces more complex realistic behaviours. The data-independent nature of these agent-based behaviours makes it a very suitable candidate to be parallelised. After an in-depth review of previous studies on the simulation of Tawaf ritual, two key behaviours associated with pilgrim movement are considered for the new model. The parallel simulation is executed on three different high-performance configurations to determine the variation in different performance metrics. The parallel implementation achieved a considerable speedup in comparison to its sequential counterpart running on a single-threaded CPU. With the use of parallel behaviours, 100,000 pilgrims at 10 fps were simulated.

INTRODUCTION

Many scientific applications require high-performance computing systems to perform their computational tasks expeditiously. This can be done using Graphics Processing Unit (GPU), as it can solve compute-intensive tasks on thousands of highly parallel, multi-threaded processing cores. The GPU utilises different levels of memory and high data throughput to accelerate the computational process. The GPUs has given rise to a programmable ecosystem that started leveraging on its highly parallel architecture. NVIDIA® standardised the customisation with the release of a toolkit called *Compute Unified Device Architecture* (CUDA). This API is an extension of C programming language. This has provided researchers with the opportunity to port several general-purpose applications onto the GPU. This notion is referred to as *General-purpose Computing on Graphics Processing Unit* (GPGPU).

Among the range of applications being ported to GPU, agent-based crowd simulation remains the most computer-intensive application. This is due to the additional computational power required with the increase in the number of agents present in a virtual environment. Researchers have proposed several models to simulate crowd movement including social force models (Helbing & Farkas, 2002), cellular automata models (Klüpfel, 2007), gas kinetic models (Hoogendoorn & Bovy, 2001) and agent-based models (Cherif & Chighoub, 2010). Among these, agent-based modelling (ABM) (Jennings, 2000), is the more frequently used methodology to implement autonomous agents. In the case of religious rituals such as Hajj and Umrah, simulating such a large crowd adds to computational demand of the system. Since each agent in the simulation possesses complex behaviours, this provides a real-life crowd which depicts the unique crowd phenomenon in these rituals.

Hajj and Umrah are performed by thousands of pilgrims all over the world. In 2015, a stampede killed more than 2000 pilgrims (Salamati & Rahimi-Movaghar, 2016). Such disasters can be prevented if there was a simulation earlier. Earlier studies were not able to simulate a large crowd to possible outcomes (Kim et al., 2015; A. N. Shuaibu, Faye, Malik, & Talal, 2014; Sakellariou et al., 2014; Sarmady, Haron, & Talib, 2011).

Therefore, there is a dire need for an agent-based simulation that can replicate dense crowd scenarios and depict the natural movement of pilgrims. This research reports the behaviours on the GPU for concurrent execution. This was implemented by Reynolds (2006) which is an open source steering library called *OpenSteer*. The current work extends previous research Rahman, Hamid, Rahiman, & Zafar, (2015).

LITERATURE REVIEW

In the field of crowd simulation and modelling, some researchers have focused on Hajj and Umrah. AlGadhi et al. was among the first to predict the throughput of pilgrims from the ``Jamarat'' site (AlGadhi & Mahmassani, 1990). His work provided the basis of pilgrim movement and their flow for future research. Abdelghany et al. proposed a micro-simulation based on the model provided by AlGhadi (Abdelghany, Abdelghany, Mahmassani, & Al-gadhi, 2006). They assessed the structure of Masjid Al-Haram by using three different levels of congestions. Zainuddin et al. simulated up to 1,000 pilgrims, implemented using proprietary software called ``SimWalk'' (Zainuddin, Thinakaran, & Abu-Sulyman, 2009). Similarly, Mulyana et al. developed a 2-dimensional software that was able to represent 500 agents for Tawaf and Sa'yee rituals performed during Hajj (Mulyana & Gunawan, 2010). Rahim et al. (2011) developed one of the first t 3-dimensional simulations for Tawaf with visualisation capability of 500 agents. Using *Cellular Automata* model, Sarmady et al. (2011) were able to mimic the circular movements of 15,000 pilgrims performing Tawaf. *Finite State Machine* (FSM) was used to simulate 35,000 agents in 2D (Curtis, Guy, Zafar, & Manocha, 2013). In a more recent research, the spiral model was used by Shuaibu et al. (2013) to simulate 1000 agents . Using X-machine model with NetLogo, Sakellariou et al. (2014) were able to represent results for 1500 primitive agents performing Sa'yee in Masjid Al-Haram. These studies faced certain limitations regarding numbers of virtual pilgrims because of limited computational power. In other words, the CPU was incapable of providing the required computational power to display the complex behaviour of pilgrims at interactive frame rates.

Therefore, there is a research gap between crowd simulation in Hajj and Umrah rituals, and the performance optimisation techniques used. Although many studies have examined Tawaf simulation, none of them addressed the performance of these simulations. The need to compute complex behaviours of a large virtual crowd is in line with the increasing number of pilgrims attending the religious event of Hajj each year (see Figure 1). Hence, to provide the required performance and computation for simulating large crowd in Hajj, the study proposes a Parallel *Agent-based crowd simulation*. It also shows some of the common behaviours observed among pilgrims during Hajj and Umrah rituals. This research will take advantage of the underlying capabilities of a GPU.
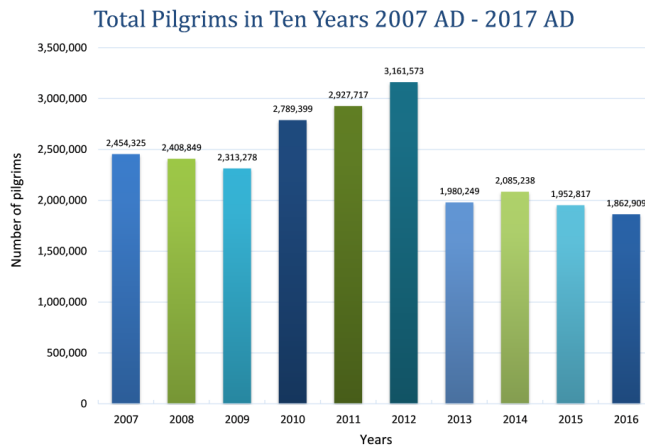
Figure 1. Number of pilgrims from 2007 – 2016 (Gen. Auth. for Stat. Saudi Arabia, 2016)

AGENT MODEL

In this research, it is necessary to design an agent model to comprehend the working of the proposed parallel agent-based simulation software. Each agent in the simulation acts as a building block for an agent-based system. Hence the understanding of their model is important before we proceed to the analysis of OpenSteer library.

The agent model is based on the reactive architecture in which is the response of the agent is based on fixed stimuli. The stimuli in this research include neighbouring agents, obstacles and the desired goal. The agent model for the proposed parallel implementation is shown in Figure 2.
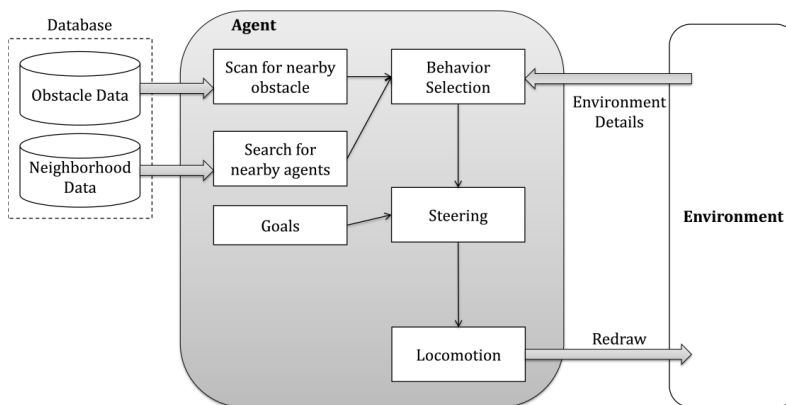
Figure 2. Agent model

In designing the agent model, the three main components of the agent includes behaviour selection, steering, and locomotion. The agent is intended to receive information about the environment. The behaviour selection component receives this information and decides on the behaviour available to use. This is also influenced by the obstacles and other agents in the vicinity. The selection of obstacle and agents to avoid will be performed on distance threshold value. This value provides the agent with ample time for determining an appropriate behaviour.

When the initial component selects the necessary behaviors, the steering part determines the underlying details. One of the most important detail that is selected is the steering force with the agent to proceed towards its goals. The information on the final goals is also provided which influences the force and truncate it accordingly.

The final module is referred to as locomotion. In this constituent, the steering force for the preceding component is utilised to calculate further properties of the agent. These properties include the velocity, orientation, and location. These properties will be used to redraw the agent in the next frame for representing the virtual environment. The following discussion provides a detailed analysis of OpenSteer library.

METHODOLOGY

The architecture of OpenSteer provides developers the ability to create complex behaviours by combining simpler behaviours within a scenario. These scenarios are referred as *plugins*. Before a plugin can starts simulating, each agent in the environment is first initialised. In this step, the objects for each agent is created, and other components are initialised (open ()). Then the created plugin is called. This call is placed in a loop, where each loop represents one simulation cycle. The first part of this step is known as the update () function. In this stage, calculations for each agent takes place and the second part is the redraw () function. The second function redraws each agent in at its new position. Ideally, this cycle should run 8 to 15 times in one second for the agents to move smoothly across the environment. At the end of the simulation, the next step (close ()) is activated which cleans up all the memory as shown in Figure 2.
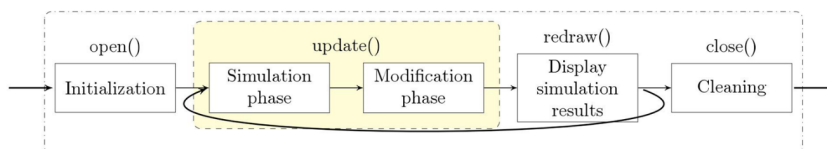
Figure 3. Simulation cycle (Rahman et al., 2015)

In OpenSteer library, the update function comprises two main parts, the *Simulation phase*, and the *Modification phase*. In the Simulation phase of the function, a *Steering Force* is calculated for each agent. If the agent abruptly moves towards the target, this will not produce realistic movement. To produce natural movement patterns, the steering force directs the agent towards its goal. In the Modification phase, the steering force from the previous phase is used to calculate the remaining properties (acceleration, velocity and future position) of the agents. The sequential implementation of library executes the update function for each agent one by

one. So, if the simulation contains a large number of agents, then the function will require a longer time to complete. After calculating properties for each agent, the results are then transferred to the redraw function. This creates a bottleneck, since the execution time for each simulation cycle increases.

The calculations in the Simulation phase do not depend on the Modification phase, so this type of data-independent tasks favours parallelisation onto GPU. The navigational behaviours of each agent determine the steering force in the first phase. For all the agents present in the environment the parallel implementation will launch n threads for n number of agents. Each thread will initially calculate the steering force and will synchronise before the launch of the section phase. This representation can be seen in Figure 3.
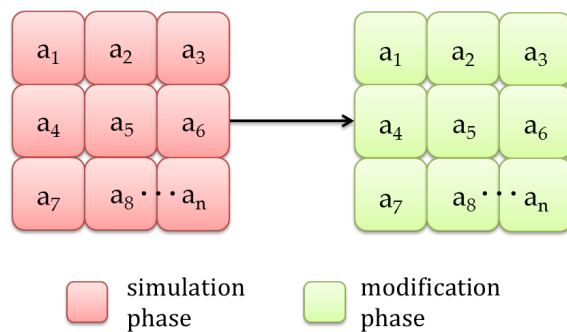
Figure 4. Traversing through simulation and modification phase in parallel

Two key navigational behaviours which are prominently seen in pilgrims performing rituals of Hajj and Umrah are selected. These behaviours include path following behavior and obstacle avoidance behavior. They allow agents to move freely in the virtual environment. The behaviours are very refined and are used to produce much of the natural motion of agents. The path following behaviour is essential for agents in the scenario of Hajj and Umrah. The complete path is considered as the goal of the character. Routes such as roads, streets, and footpaths are fixed. The character's movement on the track is not fixed; rather a character moves freely along the path like a human. This movement is done near to the pathway keeping it as a goal. By utilising the obstacle avoidance behaviour, the agents can avoid accidents initially with other agents and secondly, with obstacles within the environment. The next section discusses parallel algorithms for the selected behaviours.

PARALLEL NAVIGATIONAL BEHAVIOR ALGORITHMS

In the sequential implementation of OpenSteer library, each agent has a set of preliminary behaviours that is selected based on the required action and situation. The behaviour calculates the steering force that drives an agent. Each of these behaviours independently calculates and updates the value of steering force for each agent. In this section, the parallel algorithms of the selected navigational behaviours are discussed.

Since the computation for each agent is independent, therefore, the behaviours can be easily converted into GPU-executable kernels. In order to convert this behavior into a kernel, memory management routines must be performed before invoking them. The specific parameters related to the vehicle are initially copied to the global memory. A predefined variable such as *threadIdx, blockIdx, blockDim*, and *gridDim* are used to access the memory space from within the kernels. When the required data is retrieved, only then the process for computing steering force begins. Path following is one of the selected behavior that an agent poses to follow a predefined path. This type of behaviour is commonly observed in pilgrims performing different rituals during Hajj and Umrah (Zafar, 2011). Algorithm 1 provides the implementation for this behaviour.

Algorithm 3 Parallel Path Following Behavior

$Input : vehicleData, steeringVectors, direction, predictionTime$
$Output : steeringVectors$
$id \leftarrow (blockId \times blockSize + threadId)$
$Velocity(threadId) \leftarrow ((float*)(*vehicleData).forward)[Offset + threadId]$
$threadSyncronize()$
$Velocity(threadId) \leftarrow Velocity(threadId) \times Speed(threadId)$
$Position(threadId) \leftarrow ((float*)(*vehicleData).position)[Offset + threadId]$
$threadSyncronize()$
$pathDistanceOffset \leftarrow direction[id] * predictionTime * Speed(threadId)$
$futurePosition \leftarrow PredictFuturePosition$
$nowPathDistance \leftarrow mapPointToPathDistance(points, totalPoints,$
$\quad\quad Position(threadId))$
$futurePathDistance \leftarrow mapPointToPathDistance(points, totalPoints,$
$\quad\quad futurePosition)$
if $pathDistanceOffset > 0$ **then**
$\quad$**if** $nowPathDistance < futurePathDistance$ **then**
$\quad\quad rightway \leftarrow false$
$\quad$**else if** $nowPathDistance > futurePathDistance$ **then**
$\quad\quad rightway \leftarrow true$
$\quad$**end if**
end if
$onPath \leftarrow mapPointToPath(points, totalPoints, radius, futurePosition,$
$\quad\quad \&tangent, \&outside)$
if $(Position(threadId) - points[0]) < .radius$ **then**
$\quad direction[id] \leftarrow 1$
end if
if $(Position(threadId) - points[totalPoints - 1]) < radius$ **then**
$\quad direction[id] \leftarrow 1$
end if
if $(outside < 0)$ **and** $rightway$ **then**
$\quad target \leftarrow 0$
$\quad ignore \leftarrow 1$
else
$\quad targetPathDistance \leftarrow nowPathDistance + pathDistanceOffset$
$\quad target \leftarrow mapPathDistanceToPoint(points, totalPoints, isCyclic,$
$\quad\quad\quad targetPathDistance)$
$\quad ignore \leftarrow 0$
end if
$steerForSeekKernel(Position(threadId), Veocity(threadId), target,$
$\quad\quad steeringVectors, ignore)$

Figure 5. Parallel Path following Algorithm

In the previous algorithm, it was observed that the steering force was calculated using the function for seeking behaviour. The function was transformed into a GPU-executable kernel, calling a function residing on the host is not possible. So, the seek behaviour function was changed into a callable device kernel using a unique function type qualifier provided by CUDA. The new function can be executed on a device, and it is only callable by the device. It stores the newly computed steering behavior back to the global memory of the GPU.

Obstacle and Collision avoidance behaviour is utilised to avoid oncoming obstacles and other agents. It is a natural behaviour for virtual agents when navigating through the environment. Algorithm 2 provided the working of the obstacle avoidance behavior.

Algorithm 4 Parallel Obstacle Avoidance Behavior

$Input: vehicleData, steeringVectors$
$Output: steeringVectors$
$id \leftarrow (blockId \times blockSize + threadId)$
$minDistanceToCollision \leftarrow (*vehicleData).speed[id] \times 5.f$
for $i \leftarrow 0, numOfObstacles$ **do**
$\quad findNextIntersectionWithSphere()$
$\quad$ **if** $intersectionDistance < nearestIntersectionDistance$ **then**
$\quad\quad nearestIntersectionDistance \leftarrow intersectionDistance$
$\quad\quad nearestIntersectionID \leftarrow i$
$\quad$ **end if**
end for
if $(intersectionFound = 1)$ **and** $(nearestIntersectionDistance < minDistanceToCollision)$ **then**
$\quad offset = position[id] - obstacles[nearestIntersectionID].center)$
$\quad avoidance(threadIdx) = PerpendicularComponent(offset, (*vehicleData).forward[id])$
$\quad avoidance(threadIdx) = avoidance(threadId) \times (*vehicleConst).maxForce[id]$
$\quad avoidance(threadIdx) = avoidance(threadIdx) + ((*vehicleData).forward[id] \times (*vehicleConst).maxForce[id])$
end if
$threadSynchronize()$
if $(steeringVectors[id].x \neq 0.f$ **or** $steeringVectors[id].y \neq 0.f$ **or** $steeringVectors[id].z \neq 0.f)$ **then**
$\quad avoidance(threadIdx) = steeringVectors[id]$
else
$\quad avoidance(threadIdx) = avoidance(threadIdx) + steeringVectors[id]$
end if
$threadSynchronize()$
$steeringVectors[id] \leftarrow avoidance(threadIdx)$

Figure 6. Parallel Collision avoidance Algorithm

The behaviour starts by measuring the minimum distance to the collision. In the sequential implementation, the minimum time to the collision is provided as a function parameter. However, in the parallel implementation, that value is fixed. This value helps make the process simpler rather having a different value each time. After the calculation for the avoidance is completed, the thread synchronisation is called to ensure that all the thread has computed the avoidance vector for all agents. When all threads have finished their calculation, the measured vector is copied back to the steering force, again using the id to place the data back uniquely

into the allocated memory space. This data is then utilised by the redraw function to represent the agents in the virtual world.

Some necessary details that must be dealt with to ensure proper parallelisation. The first one was the change in phase traversal that is discussed earlier. The parallelised behaviours were blended together, each affecting the resultant steering force. This provides a more realistic movement for each agent. Each of these behaviours were invoked as CUDA kernels.

The second important thing to be considered is the change in data type that stores data for each agent. In the sequential implementation, the *Vec3* datatype was commonly used to store the three-dimensional coordinates of the agent. Since CUDA does not support this data type, the new datatype of *float3* is used. The *float3* data type is compatible with the CUDA runtime. It is of type struct and has three members *x, y,* and *z*. A built-in function make_*float3* converts the three components of the *Vec3* datatype to *float3 member variables*.

Another hurdle in the parallel conversion of the library is to carefully orchestrate the memory allocation and transfer sequences to execute the parallel algorithms properly. Attributes for all the agents needed to be copied to the GPU memory to execute the behaviour kernels developed in CUDA. Proper thread synchronisation and memory management techniques are used to ensure that the data is used for each agent.

With this, a parallel behavioral library is provided to developers, in which simpler behaviours can be combined to create complex behaviours.

KERNEL TYPES AND MANAGEMENT

During the transition from sequential to parallel implementation, two main kernels are introduced, namely, *Steering Kernel* and *Modification Kernel*. Steering kernel calculates the steering force for all agents, whereas the modification kernel applies the calculated steering force to the properties of every agent. In the parallel implementation, these kernels are used to determine the steering force of each pilgrim. In the update phase, the first *steering kernel* is responsible for describing how the pilgrim will navigate by providing an appropriate steering force, whereas, the second *modification kernel* uses the steering force value and measures the rest of the properties for the pilgrim. These include the *velocity, direction, speed* and other properties. For this purpose, an array of data type *float3* for the variable *steering Vectors* is allocated in the global memory of the GPU that stores the steering force for the total number pilgrims in each scenario. By transferring this data onto the GPU, each thread executing the kernel can read and update the data of pilgrims. So, each element of the array represents the steering force of one pilgrim. Figure 6 provides a visual representation of each steering kernel modifying the steering vector value. When the entire kernel has executed the final value, it is then acquired by the modification kernel to measure rest of the vehicle properties.
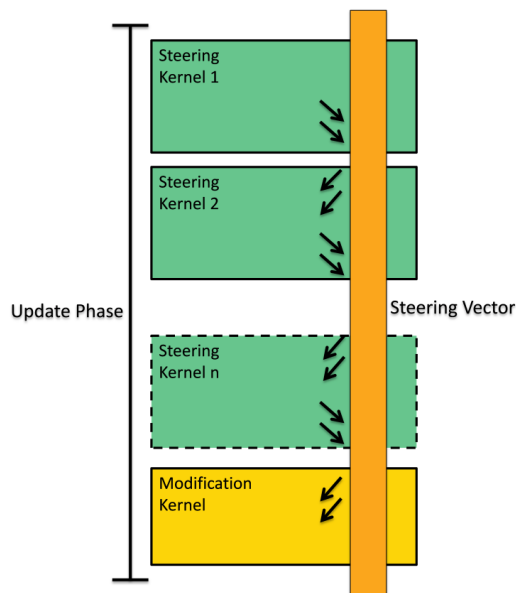
Figure 7. Steering vector accessed by different kernels

The modified data of the array containing the steering force needs to be cleared after every update cycle so the new steering forces can be stored in the next cycle. The navigational behaviours can be combined with each other to obtain more complex behaviours that can reflect more natural movement. There are two ways of combining these behaviours, namely, *Switching and Blending* (Pan, Han, Dauber, & Law, 2007).

Switching

Switching between different behaviours is one way of producing complex navigational behaviours. With this type of selection, a pilgrim can change its behavior from one to another depending upon the perception of the environment. For example, if a pilgrim is moving towards its goal, but suddenly it faces an obstacle, it will switch its behavior from goal seeking to obstacle avoidance.

Hence, in the parallelised version of the library, switching can be done manually, or it can be automated. Adding or removing the behaviours at the host, is the manual way and is to be determined at each simulation phase. The decision of this addition or removal is made by the first layer that is, the Action Selection layer. In order to automate this process, all the relevant behaviours will run and if there is no need for the kernel to apply steering force, null value will be saved in the *steering Vectors* array.

Blending

Blending is another method to achieve complex behaviours. In this type of selection, the resultant behavior is a blend of two or more behaviours. Each of the behaviours that are combined carries a weight. This weight reflects the movement of pilgrims while navigating in the virtual environment. Hence, to take the scenario from the previous example, the pilgrim will avoid the obstacle while moving towards its goal.

For utilising the blending technique in our parallel implementation, each kernel must be assigned a weight value that is used to determine its effect on the steering of the pilgrim. These forces are then added that gives a resultant force. Before this force is applied to the pilgrim, it is limited by the maximum force value *max_force*.

The next section discusses the conversion of path following and obstacle avoidance behaviours in Steering kernels.

CONVERSION OF NAVIGATIONAL BEHAVIOURS TO STEERING KERNELS

A GPU can achieve parallelism for a portion of code that can execute independently. There are *task parallelism* and *data parallelism*. In the case of agent-based crowd simulation, the task assigned to each agent is identical. However, data and properties of each agent are different. Thus, data parallelism is more suited for this type of application. This execution model is also referred to as *Single Instruction, Multiple Threads* (SIMT) where the set of instructions is same but several different threads execute it.

OpenSteer is an open-source software, and it is developed in an object-oriented manner. The source code can be acquired from its online repository that is distributed into different folders. The **"include"** folder contains the entire header file. The **"plugins"** folder contains the sample scenarios to provide initial developers with a head start. The **"src"** folder contains the source code which provides functionality for the functions used in the library. The package also provides solution folder for three operating systems, Windows, Mac OS x and Linux. For this research, we have used the Windows operating system

The initial implementation of OpenSteer provides scenarios in the form of *plugins*. Similarly, the parallel implementation also provided plugins, but these plugins contain parallelised behaviours that measured different properties of a pilgrim. Furthermore, each thread in the CUDA based kernels is assigned to a pilgrim. All participants in a particular situation are assigned a set of behaviours in the form of kernels.

Moreover, the study proposed a parallel implementation for the path following and collision avoidance behaviours. The behaviours are converted into kernels that can be executed by multiple threads concurrently. The kernels are named as *followPathKernel* and *avoidObstaclesKernel*. These kernels are called inside the normal code by providing the grid and block configuration. A sample kernel invocation is shown in Figure 7. Calling a kernel starts with its name followed by specific characters (<<< ... >>>) to inform the compiler that this is a kernel call. These characters are followed by the list of required parameters. The kernel also requires the number of required blocks and thread.

```
1   void  main()
2   {
3   int numThreads =  128;
4   int numBlocks = 8
5   followPathKernel<<<numBlocks, numThreads>>>(vehicleProp,
      steeringVector, direction, predictionTime);
6   }
```

Figure 8. Kernel invocation

Before executing the kernel, all essential data are allocated and transferred to the global memory of the GPU using c*udaMalloc* and *cudaMemcpy* functions. When the kernel is executed, the thread Id is calculated using the predefined device variable *threadIdx*. This thread id is unique and will be used to identify the agent in each scenario. This thread id is also used to indicate the properties related to the agent. This process is common for invoking both kernels from the CPU.

The kernel definition also includes some particular declaration specifier. This qualifier specifies whether the kernel is executed on the device or the host. For the kernels mentioned above, the study uses __*global*__ specifier (See Figure 9). It indicates the kernel is callable from the host and the device executes it.

```
1   __global__ void followPathKernel(VehicleProp *vehicleData, float3
      *steeringVectors, int *direction, float predictionTime)
2   {
3   ...
4   }
```

Figure 9. Kernel definition

The next section an abstraction discovered in this study to facilitate future development of the parallel library.

CUDA ABSTRACT LAYER

In existing steering library, developers who want to create plugins for simulating behaviours were able to utilize their well-structured object-oriented interface. After porting the sequential steering library using CUDA and, as a result, achieving a parallel library that can make use of

GPUs. For programmers to use this accelerated library for further development, they have to look out for the following details:

- All the parallelised behaviour kernels needed to be initialised in some ways to use them in a customised plugin. To initialise it, a programmer needs to handle all the details by hand.

- Researchers will need to have CUDA development experience to deal with the API specific details that include configurations for execution and multiple function call that are all exposed to them.

In this research, to facilitate with convenient development, a C++ abstract layer was created that encloses CUDA code. The main reason for this layer was to adopt and uphold similar features that the traditional OpenSteer provided. This layer enclosed kernel initialisation details that can be used by a programmer to produce new plugins. It comprised classes that are further described in the next sections.

PlugInCUDA Class

In the serial implementation, a new plugin inherits the *PlugIn* class that handles the initialisation. In the parallel implementation, since there are additional details that need to be taken care of which includes memory allocation, transfer, and deallocation, a new abstract layer is provided. The new plugins that will use the parallel behaviours will now inherit the *PlugInCUDA* class. This new class also inherits the existing PlugIn class and in addition, to that, it also handles the extra details related to the new vehicle properties and data transfers from CPU to GPU and vice versa.
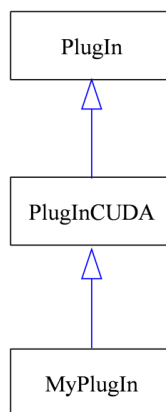
Figure 10. Class diagram for PlugInCUDA class

The next section discusses the performance results of the optimized library.

PERFORMANCE ANALYSIS AND RESULTS

Performance Metric

Frames Per Second (FPS). Frames per second or *fps* calibrates the performance of a visual application that contains a rendering of 2D or 3D environments. The OpenSteer library provides a 3D rendering of visual scenarios generated using OpenGL library. The new parallel implementation also uses the same library to render the 3D environment. This performance metrics provides the number of frames generated in one second in a simulation cycle. The range can vary greatly depending on the type of application. However, interactive frame rates are considered to be between 8-15 fps.

Execution Time of Simulation. Execution time is a key performance metric for evaluating the efficiency of an application. The execution time is inversely proportional to frames per second and expressed as:

$$execution\ time\ (ms) \propto \frac{1}{frames\ per\ second\ (fps)}$$

The equation shows that greater execution will result in fewer frames generated in one second at any given time. A high percentage of GPU utilisation will indicate a small execution time.

Speed-Up. Speedup refers to the speed of parallel application compared with the corresponding sequential application. More specifically, it shows improvement in terms of time required for execution of two different application frameworks. The term speedup was introduced in the context of Amdals's law. It can be expressed as follows:

$$S_p = \frac{T_1}{T_p}$$

Where S_p is the speedup for the parallel application, T_1 is the execution time of the sequential application, whereas T_p is the time consumed by the parallel application for completing the simulation cycle. It is necessary that while measuring the speedup, the workload must be equal. These performance metrics are necessary for measuring the efficiency of OpenSteer library. The following discussion presents the results achieved in parallelisation using the metrics mentioned previously.

An experiment on three high-performance (HPC) configuration was conducted. These HPC systems contained three different GPUs to compare their performance. The first two GPUs are from the NVIDIA® Tesla® category, C2050, and K40. The third one is from the GeForce® series, GTX 970. The results were collected by running a profiling tool provided by NVIDIA® called V*isual Profiler* (see Table 1). The primary performance metrics used for reporting the results of this research are the frames per second (fps) and execution time.

Table 1
Frames per second (FPS) for sequential and parallel implementation

No of Agents	Serial Implementation (CPU) frames per second (fps)	Parallel Implementation (GPU) frames per second (fps)		
		C2050	K40	GTX 970
500	27	83	91	101
1000	8	77	80	95
1920	6	63	70	80
2912	1	56	60	66
5920	0	36	37	45
7904	N/A	32	34	41
9920	N/A	28	29	36
19904	N/A	16	17	21
29920	N/A	12	13	18
59904	N/A	6	7	14
99904	N/A	3	5	10
199904	N/A	2	3	8
399904	N/A	1	2	3
499904	N/A	0	1	1
599904	N/A	0	1	1

The results in Table 1 shows the frames generated by the sequential and parallel implementation in a second (fps). The more the number of frames, the better the quality of the output visualisation. For a continuous video, without any noticeable delay in the frames, it should remain within the range of real-time frame rates. Eight frames per second are the minimum range of real-time frames rate. This study indicates that the fps generated by the sequential implementation cannot maintain the real-time fps for more than two thousand pilgrims. The parallel implementation shows a significant increase in performance. The high-performance configuration with NVIDIA® GTX 970 has more frames rates compared with the other two configurations for the respective number of pilgrims. The reason is that the GeForce® series is capable of handling computation as well as the visual aspect of such scientific applications. Table 2 compares different studies that have looked at the simulation of pilgrims.

Table 2

Comparison of previous studies with current implementation

Studies	Techniques	Focus	Dimensions	No of Agents	Frames Per Second (fps)
Narain et al. (2009) (Narain, Golas, Curtis, & Lin, 2009)	CFD	Tawaf	2-D	25,000	11 FPS
Zainuddin et al. (2009, 2010)	ABS	Tawaf	2-D	1000	N/A
Mulyana and Gunawan (2010)	ABS	Tawaf	2-D	150	10 FPS
Curtis et al. (2011)	ABS	Tawaf	2-D	35,000	11.5 FPS
Rahim et al. (2011)	N/A	Tawaf	3-D	500	4.3 FPS
Khan and McLeod (2012)	ABS, Cellular Automata	Tawaf	2-D	25,000	N/A
Shuaibu et al. (2013)	Mixed Mode	Tawaf	2-D	1000	N/A
Haghighati and Hassan (2013)	ABS	Tawaf	2-D	2000	N/A
Sakellariou et al. (2014)	ABS	Sa'yee	2-D	1500	N/A
Kim et al. (2015)	ABS	Tawaf	2-D	35,000	5.7 FPS
Abdur Rahman et al. (2017)	ABS	Hajj and Umrah	3-D	100,000	10 FPS

The results showed better implementation compared with previous sequential library. Crowd simulation for Tawaf and other rituals was better using this model. The parallel crowd simulation library presented in this research provided an add-on based framework through which the authorities can create multiple scenarios and event plans for different rituals of Hajj and Umrah.

CONCLUSION

This study presented a parallel implementation of navigational behaviours for agent-based crowd simulation in rituals of Hajj and Umrah which demonstrates to be a better performance. This study adds to the list of existing scientific applications that harness the high-performance capabilities of Graphics Processing Units. With a pluggable architecture of the applications, now equipped with the enhancement of high-performance computation, there is an improved program for specific event planning and management of Hajj and Umrah rituals. Experiments were conducted based on three different HPC system. The results showed robust improvements in performance for realistic crowd simulations. Encapsulation for the parallel code for further development was also provided. The research provided an abstract layer to the steering library that encapsulated the CUDA implementation. The abstraction will help developers to use improved steering behaviors.

This study was aimed at providing an optimized parallel system for large-scale agent-based crowd simulation during Tawaf ritual. There was a need for a system due to lack of computational requirements in existing sequential system to produce sufficient fps. A detailed overview of background knowledge and previous studies was followed by the design requirements of the

proposed system using existing OpenSteer library. This helped in providing a new parallel model of the optimised system. The implementation of the new model had challenges. Some of the main challenges were phase traversal and appropriate data type changes. However, these problems were overcome using GPUs and CUDA toolkit for parallel implementation. The result was an optimized high-performance application which generated large-scale crowd with realistic movement performing the Tawaf ritual. The new system was able to simulate 100,000 agents in a virtual environment at 10 fps. The results discussed earlier indicates that the new system provides a better solution as compared to the previous implementation. Therefore, the main objectives of this research have been fulfilled by implementing an optimised agent-based system to simulate crowd efficiently.

ACKNOWLEDGEMENTS

REFERENCES

Abdelghany, A., Abdelghany, K., Mahmassani, H. S., & Al-gadhi, S. A. (2006). Microsimulation assignment model for multidirectional pedestrian movement in congested facilities. *Transportation Research Record: Journal of the Transportation Research Board,* (1939), 123–132.

AlGadhi, S., & Mahmassani, H. (1990). Modelling crowd behavior and movement: Application to Makkah pilgrimage. In *Proceedings of the 11th International Symposium on Transportation and Traffic Theory* (pp. 59–78). Yokohama. Retrieved from http://www.academia.edu/2655470/Modelling_crowd_behavior_and_movement_application_to_Makkah_pilgrimage

Cherif, F., & Chighoub, R. (2010). Crowd simulation influenced by agent's socio-psychological state. *Journal of Computing, 2*(4), 48–54. Retrieved from http://arxiv.org/abs/1004.4454

Curtis, S., Guy, S. J., Zafar, B., & Manocha, D. (2011). Virtual Tawaf: A case study in simulating the behavior of dense, heterogeneous crowds. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 128–135). IEEE. http://doi.org/10.1109/ICCVW.2011.6130234

Curtis, S., Guy, S. J., Zafar, B., & Manocha, D. (2013). Virtual Tawaf: A velocity-space-based solution for simulating heterogeneous behavior in dense crowds. In *Modeling, Simulation and Visual Analysis of Crowds* (Vol. 11, pp. 181–209). Springer New York. http://doi.org/10.1007/978-1-4614-8483-7_8

Gen. Auth. for Stat. Saudi Arabia. (2016). *Hajj Statistics 1437H (2016)*. Riyadh, Saudi Arabia.

Haghighati, R., & Hassan, A. (2013). Modelling the flow of crowd during Tawaf at Masjid Al-Haram. *Jurnal Mekanikal, 36*(1), 2–18.

Helbing, D., & Farkas, I. (2002). Simulation of pedestrian crowds in normal and evacuation situations. *Pedestrian and Evacuation Dynamics, 21*(2), 21–58. Retrieved from http://www.researchgate.net/publication/224010870_Simulation_of_pedestrian_crowds_in_normal_and_evacuation_situations/file/d912f50eb0d9bb6224.pdf

Hoogendoorn, S. P., & Bovy, P. H. L. (2001). Generic gas-kinetic traffic: Systems modeling with applications to vehicular traffic flow. *Transportation Research Part B: Methodological, 35*(4), 317–336. http://doi.org/10.1016/S0191-2615(99)00053-3

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence, 117*(2), 277–296. http://doi.org/10.1016/S0004-3702(99)00107-1

Khan, I., & McLeod, R. (2012). Managing hajj crowd complexity: Superior throughput, satisfaction, health, and safety. *Kuwait Chapter of Arabian Journal of Business and Management Review, 2*(4), 45–59. Retrieved from http://www.arabianjbmr.com/pdfs/KD_VOL_2_4/4.pdf

Kim, S., Guy, S. J., Hillesland, K., Zafar, B., Gutub, A. A. A., & Manocha, D. (2015). Velocity-based modeling of physical interactions in dense crowds. *The Visual Computer, 31*(5), 541–555. http://doi.org/10.1007/s00371-014-0946-1

Klüpfel, H. (2007). The simulation of crowds at very large events. In *Traffic and Granular Flow'05* (pp. 341–346). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-47641-2_30

Mulyana, W. W., & Gunawan, T. S. (2010). Hajj crowd simulation based on intelligent agent. In *2010 International Conference onComputer and Communication Engineering (ICCCE)* (pp. 1-4). IEEE. http://doi.org/10.1109/ICCCE.2010.5556818

Narain, R., Golas, A., Curtis, S., & Lin, M. C. (2009). Aggregate dynamics for dense crowd simulation. *ACM Transactions on Graphics, 28*(5), 122. http://doi.org/10.1145/1618452.1618468

Pan, X., Han, C. S., Dauber, K., & Law, K. H. (2007). A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *Ai and Society, 22*(2), 113–132. http://doi.org/10.1007/s00146-007-0126-1

Rahim, M. S. M., Fata, A. Z. A., Basori, A. H., Rosman, A. S., Nizar, T. J., & Yusof, F. W. M. (2011). Development of 3D Tawaf simulation for hajj training application using virtual environment. In *Visual Informatics: Sustaining Research and Innovations: Second International Visual Informatics Conference, IVIC 2011, Selangor, Malaysia, November 9-11, 2011, Proceedings, Part I* (pp. 67–76). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-25191-7_8

Rahman, A., Hamid, N. A. W. A., Rahiman, A. R., & Zafar, B. (2015). Towards accelerated agent-based crowd simulation for Hajj and Umrah. In *2015 International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)* (pp. 65–70). IEEE. http://doi.org/10.1109/ISAMSR.2015.7379132

Reynolds, C. (2006). Big fast crowds on PS3. In *Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames - Sandbox '06* (pp. 113–121). New York, New York, USA: ACM Press. http://doi.org/10.1145/1183316.1183333

Sakellariou, I., Kurdi, O., Gheorghe, M., Romano, D., Kefalas, P., Ipate, F., & Niculescu, I. (2014). Crowd formal modelling and simulation: The Sa'yee ritual. In *2014 14th UK Workshop on Computational Intelligence (UKCI)* (pp. 1–8). IEEE. http://doi.org/10.1109/UKCI.2014.6930176

Salamati, P., & Rahimi-Movaghar, V. (2016). Hajj stampede in Mina, 2015: Need for intervention. *Archives of Trauma Research, 5*(5), 1. http://doi.org/10.5812/atr.36308

Sarmady, S., Haron, F., & Talib, A. Z. (2011). A cellular automata model for circular movements of pedestrians during Tawaf. *Simulation Modelling Practice and Theory, 19*(3), 969–985. http://doi.org/10.1016/j.simpat.2010.12.004

Shuaibu, A. N., Faye, I., Malik, A. S., & Talal, M. (2014). Collision avoidance path for pedestrian agent performing Tawaf. In T. Herawan, M. M. Deris, & J. Abawajy (Eds.), *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (Vol. 285, pp. 361–368). Singapore: Springer Singapore. http://doi.org/10.1007/978-981-4585-18-7

Shuaibu, N. A., Faye, I., Simsim, M. T., & Malik, A. S. (2013). Spiral path simulation of pedestrian flow during Tawaf. *IEEE ICSIPA 2013 - IEEE International Conference on Signal and Image Processing Applications* (pp. 241–245). IEEE. http://doi.org/10.1109/ICSIPA.2013.6708011

Zafar, B. (2011). *Analysis of the Mataf - Ramadan 1432 AH. Technical report.* Makkah, Saudi Arabia.

Zainuddin, Z., Thinakaran, K., & Abu-Sulyman, I. M. (2009). Simulating the circumambulation of the Ka'aba using SimWalk. *European Journal of Scientific Research, 38*(3), 454–464.

Zainuddin, Z., Thinakaran, K., & Shuaib, M. (2010). Simulation of the pedestrian flow in the Tawaf area using the social force model. *World Academy of Science, Engineering and Technology, 48,* 908–913.

Ultrasound-assisted Extraction of Natural Colourant from Husk of *Cocos nucifera*: A Comparison with Agitated-bed Extraction

Rodiah, M. H.*, Nur Asma Fhadhila, Z., Noor Asiah, H., Aziah, M. Y. and Kawasaki, N.

UNISEL Bestari Jaya Campus, Faculty of Engineering and Life Sciences,
Department of Science and Biotechnology, Jalan Timur Tambahan, 45600 Batang Berjuntai, Selangor,
Malaysia

ABSTRACT

This study investigated the use of ultrasound-assisted extraction (UAE) and agitated-bed extraction (ABE) to improve extraction efficiency of dyes from exocarp and mesocarp of *Cocos nucifera*. The UAE method was carried out at a frequency of 37 kHz and power of 150 W at 30°C for 1 hour and ABE technique was performed at 30°C for 24 hours at the agitation of 150 rpm. The effects of different solvent types (methanol, ethanol and acetone) and the different percentage of solvent (20%-70%) were determined. The results showed although the colourant yield increased with the increase of solvent concentration up to 60%, statistic value indicated no significant difference ($p > 0.05$) for the value of yield using solvent range from 20% to 70% for both UAE and ABE methods. Therefore, the solvent concentration of 20% was selected as the optimum concentration for each type of solvent used for both ABE and UAE methods. In comparison, higher extraction yields were achieved by UAE method where the optimal yield of the colourant of the mesocarp was 7.6% using acetone as solvent and the exocarp yielded about 6.4% using acetone or methanol. The recovery of natural colourant using ultrasound was found to be highly dependent on the type of solvents (acetone > methanol > ethanol) for both mesocarp and exocarp. The present study suggests that UAE method should be employed for hauling out colouring materials from exocarp and mesocarp due to its effectiveness in terms of time and economical usage of solvent.

INTRODUCTION

There is a growing demand for eco-friendly and non-toxic colourants or natural dyes, particularly for food colouration and textile as well as leather garments (Palanivel et

al., 2010). The demand for natural dyes all over the world is estimated to be 10,000 tonnes, which account for only 1% of the world synthetic dyes consumption but expected to rapidly grow in the near future (Sivakumar et al., 2011). *Cocos nucifera* (coconut) is a member of the family Arecacea (palm family), an ornamental tree growing in villages and towns in Malaysia (Abdulelah et al., 2011). The fruit contains three layers, namely exocarp (outer layer), mesocarp (fibrous husk) and endocarp (flesh). The exocarp and mesocarp make up the husk of coconut (Victor, 2013).

In Malaysia, there is an abundance of coconut husk which should be recycled. The researchers' previous study has successfully extracted natural dye using microwave-assisted extraction from these waste in order to make better use of this cheap and abundant agricultural waste. Five compounds were detected in dye extract of mesocarp, namely catechin conjugate, petunidin-3-glucoside, proanthocyanidin trimer, cyanidin-3-sambubioside and delphinidin-3-glucoside whereas four compounds were identified in dye extract exocarp which were catechin, epicatechin, petunidin-3-glucoside and cyanidin-3-(6'-feruloylsophoroside)-5-glucoside (Zulrushdi, 2018).

Previous studies have shown that agitated-bed extraction (ABE) is one of the methods to extract dyes due to its effortless and inexpensive protocol. The ABE is a type of mechanical agitation which is a simple and low-cost method that uses agitation or mixing action to extract certain phytochemical content from plant samples in a shake-flask which is placed onto a rotary shaker, or with a magnetic stirrer submerged into the flask directly (Lau et al., 2010). The ABE was also used in a study on the extraction of anthocyanins from skins of jabuticaba. In this study, ABE was compared with pressurised liquid extraction (PLE). However, the results showed PLE resulted in higher extraction efficiency compared with ABE method (Veggi, Santos, & Meireles, 2011).

The ABE method was also used in the super-critical fluid extraction (SFE) process for the recovery of phenolics using grape bagasse from Pisco residues. An agitated bed extraction was performed by placing the grape bagasse into a 250 ml of Erlenmeyer flask containing ethanol (96%). The extractions were performed in a shaker with agitation (168 rpm) for 6 hours. The findings indicated higher extraction yields were achieved by agitated bed extraction than those obtained with the supercritical fluid methods (Farías-campomanes, Rostagno, & Meireles, 2013). In another study, ABE method was compared with microwave assisted extraction (MAE) of flavonoids from cultivated *E. sagittatum* and the result showed microwave technique resulted a significant increase in extraction yield of flavonoids (Zhang et al., 2013).

Sivakumar et al. (2011) reported another improved method called ultrasound-assisted extraction (UAE) which could extract higher yield of dyes. The UAE has the capacity to improve extraction efficiency by promoting mass transfer and possible rupture of cell wall due to its effect in acoustic cavitation (Shirsath et al., 2012). Apart from that, UAE can produce cavitations in the investigated sample, and the collapse of these cavitations bubbles induces a mechanical stress on the cells. These resulted in cell disruption and penetration of the solvent into the cells (Leonelli & Mason, 2010). The UAE technology can possibly improve extraction of components, such as anthocyanins, polyphenolics, polysaccharides, aromatic compounds, and oils (Shirsath et al., 2012).

The utilisation of ultrasonic has been investigated recently in the extraction of the bioactive compound from plant materials. Examples are the extraction of total phenolics from coconut shell powder (Rodrigues & Pinto, 2007), rice bran (Tabaraki & Nateghi, 2011), the extraction of anthocyanins from Garcinia indica Choisy (Nayak & Rastogi, 2011), *Delonix regia* tree flowers (Adjé et al., 2010), carotenoids (Sun et al., 2011) and betalains (Sivakumar et al., 2009). The efficacy of ultrasound-assisted extraction is also proven to be more predominant in getting higher yield with shorter extraction time when contrasted with other extraction methods, such as the conventional reflux in the extraction of phenolic acid from *Citrus unshiu Marc* peels (Ma et al., 2009) and the extraction of salvianolic acid B from *Salvia miltiorrhiza* root (Dong et al., 2010). The use of ultrasound was found to have a significant enhancement in the extraction efficiency of colourant obtained from beet root. Around 80% improvement in the yield of the colourant was accomplished with ultrasound compared with the magnetic stirring process using 1:1 ethanol–water (Sivakumar et al., 2009). Zou et al. (2010) demonstrated that natural melanin from dried fruit bodies of *Auricularia auricula* was successfully extracted using ultrasound with significant yields of the melanin.

The nature of the bioactive compound present in a plant is varied and depends on the plant materials. Hence, in general, it is extremely difficult to suggest an appropriate extracting solvent for individual plant materials. The main objective of the present study is to evaluate the effect of various solvents (methanol, ethanol and acetone) and different types of extraction methods (agitated bed extraction and ultrasound assisted extraction) on the extractability of natural dye.

MATERIALS AND METHODS

Preparation of Sample

The exocarp (outer layer) and mesocarp (fibrous husk) of *Cocus nucifera* (coconut) were obtained from Tanjung Karang, Selangor (Malaysia). In this study, only mature (brown colored) husks were used. Exocarp was separated from the mesocarp before both parts were cut and dried in an oven at 50°C for 2 days. The samples were finely ground using a grinder, sieved with a 0.5 mm sieve and kept in a clean plastic container, away from heat and moisture prior to conducting the experiments.

Methods of Extraction

Ultrasound-assisted extraction (UAE) and Agitated-bed extraction (ABE). Approximately 3g of ground coconut exocarp and mesocarp were weighed separately before each of the samples was mixed with 60 mL of ethanol. In this study, the concentration of solvents varied ranging from 20% to 70%. The samples were sonicated at a frequency of 37 kHz and power of 150 W in a sonicator bath (Elmasonic, Germany) at 30°C for an hour based on the previous method adopted by Asma et al. (2015). The extraction process could not be prolonged beyond one hour as the temperature of water in the sonicator bath was increased to 65 to 70°C. The increased temperature caused dryness to the sample due to solvent evaporation. After ultrasound

extraction, the samples were filtered using filter paper (CHM, Germany) in order to remove plant materials before the solvent was evaporated using a rotary evaporator at room temperature for 5 minutes. The samples were then centrifuged at 10,000 rpm for 10 minutes. The supernatant was stored at 4°C in the dark prior to analysis. The same procedures were employed using different solvents (methanol and acetone).

The ABE was performed using an incubator shaker (Stuart, United Kingdom) at 30°C for 24 hours at an agitated speed of 150 rpm. The parameters for ABE extraction were examined in previous studies which indicated that maximum yield was obtained at 24 hours. A shorter period of extraction resulted in non-significant yield (unpublish data). Similar procedures of sample preparation, filtration, and removal of solvent and collection of the sample were the same as those used in the UAE technique.

Analytical Methods

Gravimetric Analysis. The yield of the natural colourant of each sample was determined by a gravimetric method which was performed according to the method described by Sivakumar et al. (2011). Sample extracts were dried in a hot air oven at 80°C, overnight until all the water evaporated. The samples were then cooled in a desiccator and weighed until a constant weight was achieved. The yield for total extract was calculated using the following formula:

% yield of natural colourant = $\underline{\text{Natural dye extract obtained (g)}}$ x 100

Weight of sample used (g)

% improvement of UAE compared to ABE =

$\underline{\text{\% yield of (UAE process–ABE process)}}$ x 100

% yield of ABE process

Colour Measurement Analysis. The colour was evaluated using a colorimeter (Konika, Minolta, Japan). The colour of the sample was quantified in terms of L*, a*, b*, c* and h* values which are the variables in the CIELAB color space and described as follows: A negative value of L* denotes darker shade while a e positive value of L* indicates a a lighter shade. Negative value of a* is indicated by greener color and positive value of a* is indicated by red. Negative value of b* is indicated by blue color whie positive value of b* is indicated by yellow color. c* signifies chroma or purity of colour while h* represents hue (shade) of color.

Statistical Analysis. All the experiments were carried out in triplicate, and the results were expressed as means ± SD (standard deviation). Statistical analysis was done using SPSS software version 16 using the Tukey test. A value of $p < 0.05$ was considered statistically significant. Meanwhile, the statistical significance for comparison between UAE and ABE was evaluated using Student's t-test and set at <0.05.

RESULTS AND DISCUSSION

Colourant yield between UAE and ABE

Mesocarp Extract. The percentage yield of natural dye of mesocarp extracted using agitated bed and ultrasound are shown in Table 1. In order to identify the most effective solvent which contributes the highest yield of colourant, varying the types and compositions of solvent were done. This is crucial as Shirsath et al. (2012) also mentioned that the amount of solvent utilised in the operation is essential and in some cases, it is beneficial to add some amount of water to the solvents. In this study, the concentration of solvent was chosen from 20% to 70% because the preliminary study showed that solvent at the concentration less than 20% did not give much effect on dye extraction within 24 hours. The concentration of the solvent that is more than 70% will evaporate very quickly within 30 to 60 minutes when employed, causing dryness to the sample. Based on the results, the highest yields were obtained when mesocarp was extracted with acetone at the concentration range between 50% and 60% using UAE method. This indicated that extending the concentration of solvent beyond 60% was unnecessary and also could result in reducing the amount of colourant extracted, which is supported previous findings. Shirsath et al. (2012) reported that to enhance the extraction of isoflavones from soybeans it was important to include a specific amount of water (40–60%) to the extracting solvent. This was probably due to the relative polarity of the compounds, and the increased propagation of ultrasonic waves in aqueous solvents. When large volumes of water (higher than 60%) were added, the effectiveness of extraction was reduced possibly attributed to increased production of radicals from the ultrasound induced dissociation of water (Shirsath et al., 2012).

Table 1

Percentage yield of natural dye from coconut husk (Mesocarp) extracts in different concentration of solvent

Extraction methods	Concentration (%)	Type of Solvent		
		Ethanol (%)	Methanol (%)	Acetone (%)
Ultrasound-assisted extraction	20	$7.2^a \pm 1.13$	$5.2^a \pm 1.69$	$7.6^a \pm 1.69$
	30	$7.2^a \pm 1.13$	$6.8^a \pm 0.56$	$8.4^a \pm 0.56$
	40	$7.6^a \pm 0.56$	$7.2^a \pm 2.26$	$8.0^a \pm 2.26$
	50	$6.8^a \pm 1.69$	$6.8^a \pm 0.56$	$9.2^a \pm 0.56$
	60	$6.0^a \pm 0.56$	$6.8^a \pm 0.56$	$9.2^a \pm 0.56$
	70	$5.2^a \pm 0.56$	$6.0^a \pm 0.56$	$6.8^a \pm 0.56$
Agitated-bed extraction Agitated-bed	20	$4.0^a \pm 1.13$	$5.2^a \pm 0.56$	$6.0^a \pm 2.82$
	30	$5.6^a \pm 1.13$	$4.4^a \pm 1.69$	$5.6^a \pm 0.00$
	40	$6.0^a \pm 1.69$	$8.0^a \pm 1.13$	$5.2^a \pm 2.82$
	50	$6.0^a \pm 0.56$	$8.0^a \pm 2.26$	$7.2^a \pm 0.00$
	60	$6.4^a \pm 1.13$	$6.8^a \pm 0.56$	$6.8^a \pm 2.82$
	70	$3.6^a \pm 1.34$	$5.6^a \pm 1.13$	$3.1^a \pm 0.35$

Results as means from triplicates. [a–b] Different superscript letters between the concentration of solvent for each extraction method denote significant differences ($p < 0.05$)

On the other hand, in ABE method, methanol extract at the concentration of 40%-50% revealed the highest percentage of colourant yield (8.0%) compared with other solvents. Apparently, for both methods, the yield of colourant increased markedly with the increase of the concentration of each type of solvent and reached a peak value with 50% of concentration. However, there was no significant difference ($p > 0.05$) among the concentration of each solvent used (20%-70%). Therefore, the solvent concentration of 20% was selected as the optimum concentration for each type of solvent used for both ABE and UAE methods.

Exocarp Extract. The percentage yield of the natural dye of exocarp for extraction using agitated bed and ultrasound is shown in Table 2. Similarly, an effective solvent for extracting natural dye from exocarp of *Cocos nucifera* was identified by varying the types and composition of the solvent. Based on the results, the highest yield (11.6%) was obtained when exocarp was extracted with acetone at the concentration of 50% using UAE method. Whereas, in ABE method, methanol extract at the concentration of 50%-60% gave the highest percentage of colourant (7.6%). However, there was no significant difference ($p > 0.05$) among the concentration of each solvent used (20%-70%). Therefore, the solvent concentration of 20% was selected as the optimum concentration for each type of solvent used in the exocarp sample.

Table 2

Percentage yield of natural dye from outer layer (Exocarp) extracts in different concentration of solvent

Extraction methods	Concentration (%)	Type of Solvent		
		Ethanol (%)	Methanol (%)	Acetone (%)
Ultrasound-assisted extraction	20	6.0[a] ± 0.56	6.4[a] ± 1.13	6.4[a] ± 1.13
	30	5.2[a] ± 0.56	5.6[a] ± 1.13	7.2[a] ± 2.26
	40	9.2[a] ± 1.69	6.0[a] ± 1.69	10.4[a] ± 1.13
	50	6.8[a] ± 0.56	7.6[a] ± 0.56	11.6[a] ± 2.28
	60	8.4[a] ± 2.82	6.8[a] ± 1.69	8.8[a] ± 1.13
	70	5.2[a] ± 0.56	6.8[a] ± 1.69	7.2[a] ± 1.13
Agitated-bed extraction	20	5.2[a] ± 0.56	2.8[a] ± 0.56	2.1[a] ± 0.14
	30	6.8[a] ± 0.56	5.6[ab] ± 1.13	5.6[a] ± 0.00
	40	6.4[a] ±1.13	3.6[a] ± 0.56	6.0[a] ± 1.69
	50	7.2[a] ± 2.26	7.[6b] ± 1.69	6.0[a] ± 1.69
	60	6.8[a] ± 0.56	7.6[b] ± 0.56	4.4[a] ± 0.56
	70	5.2[a] ± .69	5.2[ab] ± 1.69	4.0[a] ± 1.13

Results as means from triplicates. Different superscript letters between the concentration of solvent for each extraction method denote significant differences ($p < 0.05$).

Based on the observation, the increase of solvent concentration (ethanol, methanol and acetone) up to 70% resulted in decreasing pattern of the extraction yields for both UAE and ABE. The alcohol vaporised easily during the extraction process and resulted in slight dryness to the medium of extraction. The finding is in agreement with Cacace and Mazza (2003) who

reported that the increase in ethanol concentration reduces the dielectric constant of the solution and consequently reduces the interaction energy between the solute and solvent (Cacace & Mazza, 2003).

Comparison between UAE and ABE based on Selected Optimum Condition. In order to validate the efficiency of extraction method, a comparison was made between UAE and ABE methods based on the optimum concentration (20%) for each type of solvent used (Table 3). Under the optimised conditions, the greatest colourant yield was extracted by ethanol for the mesocarp and there was 80% of improvement in the percentage yield of coloring matter extract obtained due to the use of ultrasound compared with ABE method. This may be due to the intensity of ultrasonic cavitation in the ethanol mix with water increase as a result of the increase in surface tension and the decrease in viscosity as stated by Boonkird et al. (2008). Meanwhile, a three-fold increase in the percentage yield of the colourant was seen using the UAE method when the exocarp was extracted using acetone. This enhancement could be attributed to cavitation and thermal effects of the ultrasound technique, which cause disruption of the cell wall and intensification of mass transfer (Esclapez et al., 2011; Vajic et al., 2015).

Table 3
The effect of ultrasound on the yields of colourant extracted from mesocarp and exocarp using ethanol, methanol and acetone at 20% concentration

Plant sample	Type of solvent	% yield of UAE (A)	% yield of ABE (B)	% improvement due to ultrasound = ((A-B)/B) × 100
Mesocarp	Ethanol	$7.2^a \pm 1.13$	$4.0^b \pm 1.13$	80.0
	Methanol	$5.2^a \pm 1.69$	$5.2^a \pm 0.56$	0.0
	Acetone	$7.6^a \pm 1.69$	$6.0^a \pm 2.82$	26.6
Exocarp	Ethanol	$6.0^a \pm 0.56$	$5.2^a \pm 0.56$	15.4
	Methanol	$6.4^a \pm 1.13$	$2.8^b \pm 0.56$	128.6
	Acetone	$6.4^a \pm 1.13$	$2.1^b \pm 0.14$	204.8

Results as means from triplicates. [a–b] Different superscript letters between the percentage of yield between UAE and ABE for each solvent denote significant differences (independent t-test, $p<0.05$)

In this study, the mesocarp and exocarp samples were sonicated at a frequency of 37 kHz and power of 150 W in a sonicator bath at 30°C for an hour. In a similar manner, Khan et al (2010) discovered that the UAE method was faster and more efficient than industrial hot maceration process. The results revealed that UAE at ultrasonic bath operating at 35 kHz for 3 h gave 87.4% yield compared with 79.4% of the maceration process for 15 h. Rouhani et al. (2009) studied the effect of UAE on the extraction of curcuminoids from turmeric plant rhizomes using ultrasonic bath at 35 kHz for 15 min and discovered that the yield of UAE was three times higher than the traditional method (solvent extraction, soxhlet extraction and maceration).

The efficacy of ultrasonic-assisted extraction not only improved natural colourant yield recovery but also reduced extraction time. Despite the long period of extraction up to 24 h, the low yield of colourant values was observed for both mesocarp and exocarp produced using the

agitated-bed. In this study, the extraction time is shortened to 1 h from 24 h if the ultrasound is chosen as the preferable method. Thus, the UAE process with 20% ethanol and 20% acetone as extracting solvent seems to be the best combination for the extraction method of natural dye recovery for both mesocarp and exocarp.

However, it is crucial that the sonication time is carefully optimised, since exposure to ultrasonic irradiations may damage the quality of the solute in certain heat sensitive materials. (Romdhane & Gourdon, 2002). Indeed the findings are in agreement with Rouhani et al. (2009) who reported that extraction of curcuminoids from turmeric plant rhizomes using ultrasonic bath increased but if the extraction time was longer than 15 min, the extraction yield declined. The finding showed that the induced cavitation can degrade compounds when exposed for a longer time.

Generally, agitated bed extraction mainly depended upon continuous permeation and solubilisation processes to leach target constituents. Extraction of colouring matter is a solid-liquid leaching process involving mass transfer problem (Sivakumar et al., 2011). Since the colouring matter is tightly bound to plant cell membranes, extraction could be improved by methods such as ultrasound which use shorter time (less than an hour). This finding is supported by Sivakumar et al. (2011) who concluded that the improvement of yield of extraction was due to better leaching of natural dye material from plant cell membranes and mass transfer to solvent assisted by acoustic cavitation from the ultrasound.

The percentage of recovery of natural colourant using ultrasound was found to be highly dependent on the type of solvents (acetone > methanol > ethanol) for both mesocarp and exocarp. The determination of the best solvent for ultrasound assisted extraction typically relies on the physical properties (surface tension, vapor pressure and viscosity) of the solvent because these properties influence the cavitational intensity in a liquid phase. Solvents vary in their extraction abilities depending on their own and the solute's chemical structure. Once the functional group of the solute is identified, possible solvents can be chosen. Common solvents used are acetone, alcohols, and ether to extract bio-active substances from natural products due to their broad solubility propensity (Shirsath et al., 2012).

In the present study, acetone exhibited the highest yield which may be attributed to its semi-polar characteristic which might attract both polar and nonpolar compounds. In fact, it was verified that most natural antioxidants from plants, similar to phenolic compounds, flavonoids and anthocyanins, are easily dissolved by low-polar organic solvents, such as ethanol, methanol, and acetone, with different levels of water (Xu et al., 2017). Indeed, these findings indicated that organic solvents diffuse into solid material during extraction procedure and then solubilise the compound with similar polarity. The nature of the solvent used decides the sort of chemicals likely extracted from plant materials (Tiwari et al., 2011). Polarity is the capacity of a molecule to engage in strong interactions with other polar molecules (Barwick, 1997). Most of the bioactive components of plant matrices are medium-sized molecules. The presence of aromatic delocalised µ-electrons may cause the molecules to be highly polarisable (Kolar et al., 2002).

Jadhav et al., (2009) reported that various solvents for extraction of vanillin demonstrated that the utilisation of polar solvents resulted in a maximum degree of extraction. The use of

polar and non-polar solvents combined with ultrasonic energy has been reported to result in quick and effective extraction of total lipids from solid matrices (Metherel et al., 2009). Furthermore, Shirsath et al. (2012) described that solvents, such as methanol, ethanol and hexane are frequently used for UAE instead of water.

Colour Strength

The results of different types and solvent proportions on the color of the natural dye of mesocarp and exocarp using ultrasound-assisted extraction (UAE) and agitated bed extraction (ABE) are shown in Table 4 and Table 5. The color values of the samples were measured in terms of L*, a*, b*, c* and h. This indicated that the method employed for the extraction slightly affected the colour values. All mesocarp and exocarp extracted from UAE and ABE had almost similar color, with lightness (L*) values ranging from 23.2 to 27.7 (mesocarp) and 23.3 to 31.8 (exocarp). However, the colour was lighter using UAE method (compared with ABE method). L* represents lightness value, the lower the lightness value, the higher the colour yield. In addition, the results obtained were correlated with the result from the percentage yield of colourant extracted using both methods of UAE and ABE (Table 1).

Table 4
Color coordinate of mesocarp extracts using ultrasound-assisted extraction (UAE) and agitated-bed extraction (ABE) technique

Method of extraction	Solvent	Concentration	Color Coordinates				
		(%)	L*	a*	b*	c*	h*
UAE	Ethanol	20	24.4±0.8	1.4±0.4	3.9±0.7	4.1±0.8	70.6±2.9
		30	25.8±0.2	0.3±0.1	1.9±1.3	2.0±1.4	73.2±7.5
		40	24.0±0.1	1.3±0.2	1.2±0.2	1.8±0.1	47.4±0.5
		50	24.5±0.5	0.7±0.4	2.2±0.8	2.3±0.5	68.0±18.1
		60	23.5±0.5	0.4±0.1	0.9±0.3	0.9±0.2	68.2±8.1
		70	24.1±1.8	4.4±2.4	3.5±1.2	5.7±2.3	40.1±12.9
	Methanol	20	25.0±0.1	1.4±0.0	3.5±0.6	3.8±0.5	67.9±3.1
		30	24.9±0.3	0.3±0.0	5.2±0.5	5.2±0.5	87.1±0.4
		40	25.1±0.0	1.6±0.3	3.6±0.4	4.0±0.6	66.2±2.9
		50	25.2±0.7	2.0±0.9	3.6±0.1	4.2±0.4	61.5±11.2
		60	24.9±0.2	0.5±0.2	3.7±0.8	3.8±0.7	81.5±4.4
		70	24.9±1.5	2.2±1.4	5.3±1.3	5.8±1.6	69.3±10.2
	Acetone	20	25.1±0.1	0.4±0.2	4.1±0.1	4.1±0.1	84.8±2.2
		30	25.1±0.4	0.7±0.1	3.0±0.2	3.1±0.2	74.1±2.7
		40	24.7±0.4	1.2±0.4	3.7±0.2	3.8±0.2	72.6±4.7
		50	24.0±0.4	1.5±1.3	3.7±1.0	4.2±0.4	66.9±21.7
		60	23.3±0.6	2.9±0.8	2.8±0.1	4.0±0.7	45.0±6.9
		70	24.5±0.3	3.1±0.4	3.6±0.4	4.8±0.5	49.3±0.3

Table 4 *(continue)*

Method of extraction	Solvent	Concentration	Color Coordinates				
		(%)	L*	a*	b*	c*	h*
ABE	Ethanol	20	24.5±1.1	1.7±0.2	10.6±1.3	10.7±1.3	80.9±0.1
		30	25.3±1.3	0.5±0.3	5.1±2.8	5.2±2.8	84.5±1.5
		40	25.8±0.6	1.7±0.2	5.8±2.5	5.9±2.3	79.8±15.6
		50	27.4±1.0	0.6±0.3	4.4±1.3	4.5±1.2	80.2±8.7
		60	26.3±1.2	1.7±0.9	6.2±3.1	6.5±2.7	71.2±18.4
		70	27.5±0.6	0.1±0.2	4.9±0.2	4.9±0.2	88.3±2.7
	Methanol	20	31.8±0.1	0.6±0.2	1.3±0.1	1.5±0.0	67.2±7.8
		30	31.3±1.0	0.7±0.4	1.5±0.9	1.7±1.0	64.2±4.2
		40	29.5±0.0	0.5±0.4	2.2±0.4	2.3±0.3	75.7±11.8
		50	30.2±0.4	0.1±0.3	0.4±0.5	0.5±0.5	74.3±2.2
		60	31.4±0.6	0.9±0.5	3.4±0.4	3.5±0.1	75.2±9.5
		70	30.3±0.0	0.6±0.5	3.0±0.2	3.1±0.1	78.6±9.8
	Acetone	20	30.9±0.1	0.3±0.2	2.4±0.2	2.5±0.2	81.4±5.0
		30	31.5±0.1	0.4±0.3	1.6±0.1	1.7±0.0	77.8±9.0
		40	31.5±0.1	0.5±0.1	1.6±0.1	1.7±0.1	73.5±3.0
		50	31.8±0.1	0.6±0.2	1.3±0.1	1.5±0.0	67.2±7.8
		60	31.2±0.1	5.5±0.2	11.4±0.1	12.6±0.0	64.4±7.8
		70	30.4±0.4	1.5±0.1	3.4±0.5	3.8±0.5	67.6±2.5

Table 5
Color coordinate of exocarp extracts using ultrasound-assisted extraction (UAE) and agitated-bed extraction (ABE) technique

Method of extraction	Solvent	Concentration	Color Coordinates				
		(%)	L*	a*	b*	c*	h*
UAE	Ethanol	20	24.1±0.2	1.8±0.2	1.7±0.2	2.4±0.1	43.8±7.9
		30	23.3±0.6	2.9±0.8	2.8±0.1	4.0±0.7	45.0±6.9
		40	23.9±0.1	1.9±0.3	2.9±0.3	3.5±0.4	56.7±1.8
		50	23.7±0.6	1.9±1.1	2.5±0.6	3.2±0.8	53.5±17.1
		60	23.2±1.5	3.2±2.9	2.0±0.5	4.0±2.3	40.8±28.5
		70	23.4±1.6	3.3±2.8	2.5±0.9	4.6±1.8	45.1±29.2
	Methanol	20	23.3±0.6	2.9±0.8	2.8±0.1	4.0±0.7	45.0±6.9
		30	24.2±0.5	3.7±0.9	3.2±1.0	5.2±0.7	48.1±1.9
		40	25.8±0.5	2.7±0.6	3.7±0.8	4.6±0.8	53.5±5.8
		50	23.4±1.6	3.3±2.8	2.5±0.9	4.6±1.8	45.1±29.2
		60	24.2±1.1	1.9±0.6	2.6±0.4	3.2±0.5	54.4±9.5
		70	24.1±0.3	2.4±0.9	2.4±0.3	3.4±0.8	45.9±11.1

Table 5 *(continue)*

Method of extraction	Solvent	Concentration	Color Coordinates				
		(%)	L*	a*	b*	c*	h*
	Acetone	20	24.4±0.3	2.4±0.4	4.3±0.3	4.9±0.4	61.2±2.2
		30	24.5±1.2	2.8±0.2	5.3±0.2	6.0±0.2	62.3±1.1
		40	23.9±0.1	1.9±0.3	2.9±0.3	3.5±0.4	56.7±1.8
		50	24.5±1.0	3.0±0.6	3.6±0.9	4.6±0.7	49.7±9.9
		60	23.6±0.3	2.0±0.3	2.0±0.3	2.8±0.2	45.8±7.9
		70	23.4±0.7	2.7±0.7	1.8±0.9	3.4±0.2	33.8±19.1
ABE	Ethanol	20	26.2±1.0	0.7±1.2	3.1±0.5	3.3±0.8	80.4±15.5
		30	27.4±1.0	0.6±0.3	4.4±1.3	4.5±1.2	80.2±8.70
		40	27.2±0.3	0.5±0.5	5.8±1.3	5.8±1.2	84.7±6.40
		50	27.2±1.5	0.6±0.3	4.8±3.3	5.5±4.4	72.8±15.3
		60	26.1±1.2	0.8±0.0	5.8±1.4	12.1±4.0	85.8±1.80
		70	26.8±0.6	0.6±0.2	5.5±2.5	7.8±2.3	81.1±15.6
	Methanol	20	27.0±0.0	0.2±0.0	3.1±0.1	3.1±0.1	85.7±0.2
		30	27.0±0.1	0.3±0.1	3.4±0.5	3.4±0.6	85.1±0.8
		40	27.4±1.0	0.6±0.3	4.4±1.3	4.5±1.2	80.2±8.7
		50	27.2±1.5	0.6±0.3	4.8±3.3	5.5±4.4	72.8±15.3
		60	27.4±1.0	0.6±0.3	4.4±1.3	4.5±1.2	80.2±8.7
		70	27.7±0.4	1.5±0.2	4.2±1.3	4.4±1.2	69.3±7.3
	Acetone	20	26.0±2.1	1.2±0.6	3.0±1.3	3.3±1.4	69.5±3.6
		30	26.8±1.1	1.0±0.6	5.0±4.0	5.1±4.0	77.4±6.0
		40	26.4±1.7	0.7±1.0	3.9±1.9	4.0±2.0	82.8±9.2
		50	27.1±0.8	0.3±0.2	6.3±1.7	6.3±1.7	86.7±2.0
		60	27.6±1.2	1.8±0.5	15.3±3.6	15.4±3.6	83.4±0.3
		70	27.1±0.4	1.0±1.3	3.9±0.6	4.1±1.0	77.1±13.2

Based on Table 4 and Table 5, all samples (mesocarp and exocarp) indicated that natural colourant extracted had lighter hue (yellow and red) due to low a* and b* values. In comparison, natural colourant extracted from UAE and ABE did not have a remarkable increase in a* and b* value throughout the experiment. The emergence of plant pigments and phenolic compounds (colourants) was closely related to colorful substances consisting of biochromes, which absorb or reflect light of varying wavelengths (Boonsong et al., 2012).

For all mesocarp samples extracted from ABE, hue angle (h*) was slightly lower (33.8 till 62.8) than UAE (69.3 till 86.7) which indicated that the longer the extraction time, the darker the color of extract. As previously reported, ABE method took about 24 hours to complete the extraction process compared with UAE which required only one hour. Overall, the color difference (mesocarp and exocarp) using UAE method was smaller than ABE method. The color of the mesocarp and exocarp from both UAE and ABE methods was dark (low L* value) but with low intensity (less vivid).

CONCLUSION

Two methods were performed and compared known as UAE and ABE method in maximising the extraction of natural dyes from the husk of *Cocos nucifera*. The result indicated that UAE performed better in 20% of ethanol and 20% of acetone (80% and 204.8%, respectively; in the improvement of percentage yield of colorant due to ultrasound) in mesocarp and exocarp respectively. In addition, UAE proved to be better technique due to acoustic cavitations provided by ultrasound. Moreover, UAE is a more promising method as the time taken to complete the process was only one hour compared with the ABE method which required 24 hours. Finally, the positive results on color analysis highlighted the potential of utilising *Cocos nucifera* husk as a source of natural colourant.

ACKNOWLEDGEMENT

REFERENCES

Abdulelah, H. A., Zurainee M. N., Hesham M. A. M., Adel, A. A., & Rohela, M. (2011). Evaluation of the use of *Cocos nucifera* as antimalarial remedy in Malaysian folk medicine. *Journal of Ethnopharmacology, 134*(3), 988–991.

Adjé, F., Lozano, Y. F., Lozano, P., Adima, A., Chemat, F., & Gaydou, E. M., (2010). Optimization of anthocyanin, flavonol and phenolic acid extractions from *Delonix regia* tree flowers using ultrasound-assisted water extraction. *Industrial Crops and Products, 32*(3), 439-444.

Asma, F. Z., Rodiah, M. H., Aziah, M. Y., Norakma, M. N., & Nurhafizah, I. (2015). Ultrasound-assisted extraction of natural dye from exocarp and mesocarp of *Cocos nucifera*. *Advanced Materials Research, 1113*, 477-480.

Barwick, V. J. (1997). Strategies for solvent selection – A literature review. *Trends in Analytical Chemistry, 16*(6), 293–309.

Boonkird, S., Phisalaphong, C., & Phisalaphong, M. (2008). Ultrasound-assisted extraction of capsaicinoids from *Capsicum frutescens* on a lab- and pilot -plant scale. *Ultrason. Sonochem, 15*(6), 1075–1079.

Boonsong, P., Laohakunjit, N., & Kerdchoechuen, O. (2012). Natural pigments from six species of Thai plants extracted by water for hair dyeing product application. *Journal of Cleaner Production, 37*, 93-106.

Cacace, J. E., & Mazza, G. (2003). Mass transfer process during extraction of phenolic compounds from milled berries. *Journal of Food Engineering, 59*(4), 379–389.

Dong, J., Liu, Y., Liang, Z., & Wang, W. (2010). Investigation on ultrasound-assisted extraction of salvianolic acid B from *Salvia miltiorrhiza* root. *Ultrason. Sonochem, 17*(1), 61–65.

Esclapez, M. D., Garcia-Perez, J. V., Mulet, A., & Carcel, J. A. (2011). Ultrasound-assisted extraction of natural products. *Food Engineering Reviews, 3*(2), 108–120.

Farías-campomanes, A. M., Rostagno, M. A., & Meireles, M. A. A. (2013). Production of polyphenol extracts from grape bagasse using supercritical fluids: Yield, extract composition and economic evaluation. *The Journal of Supercritical Fluids, 77*, 70–78.

Jadhav, D., Rekha, B. N., Gogate, P. R., & Rathod, V. K. (2009). Extraction of vanillin from vanilla pods: A comparison study of conventional soxhlet and ultrasound assisted extraction. *Journal of Food Engineering, 93*(4), 421-426.

Khan, M. K., Vian, M. A., Tixier, A.S.F., Dangles, O., & Chemat, F. (2010). Ultrasound-assisted extraction of polyphenols (*flavanone glycosides*) from orange (*Citrus sinensis L*.) peel. *Food Chemistry, 119*(2), 851-858.

Kolar, P., Shen, J. W., Tsuboi, A., & Ishikawa, T. (2002). Solvent selection for pharmaceuticals. *Fluid Phase Equilibria, 194*(7), 771–82.

Lau, E. V., Gan, S., & Ng, H. K. (2010). Extraction techniques for polycyclic aromatic hydrocarbons in soils. *International Journal of Analytical Chemistry, 2010*, 1-9.

Leonelli, C., & Mason, T. (2010). Microwave and ultrasonic processing: Now a realistic option for industry. *Chemical Engineering Proccess, 49*(9), 885-900.

Metherel, A. H., Taha, A. Y., Izadi, H., & Stark, K. D. (2009). The application of ultrasound energy to increase lipid extraction throughput of solid matrix samples (flaxseed). *Prostaglandins, Leukotrienes and Essential Fatty Acids, 81*(5-6), 417-423.

Nayak, C. A., & Rastogi, N. K. (2011). Optimization of solid–liquid extraction of phyto-chemicals from *Garcinia indica Choisy* by response surface methodology. *Food Research International, 50*(2), 550-556.

Palanivel, V., Seralathan, K., Vellingiri, B., Perumalsamy, L., Chae, J. C., & Oh, B. T. (2010). Natural pigment extraction from five filamentous fungi for industrial applications and dyeing of leather. *Journal of Carbohydrate Polymers, 79*(2), 262-268.

Rodrigues, S., & Pinto, G. A. S. (2007). Ultrasound extraction of phenolic compounds from coconut (*Cocos nucifera*) shell powder. *Journal of Food Engineering, 80*(3), 869-872.

Romdhane, M., & Gourdon, C. (2002). Investigation in solid–liquid extraction: influence of ultrasound. *Chemical Engineering Journal, 87*(1), 11-19.

Rouhani, S., Alizadeh, N., Salimi, S., & Haji-Ghasemi, T. (2009) Ultrasound assisted extraction of natural pigments from rhizomes of *Curcuma Longa L. Progress in Color, Colorants and Coatings, 2*, 103–113.

Shirsath, S. R., Sonawane, S. H., & Gogate, P. R. (2012). Intensification of extraction of natural products using ultrasonic irradiations—A review of current status. *Chemical Engineering and Processing, 53*, 10–23.

Sivakumar, V., Anna, J. L., Vijayeeswarri, J., & Swaminathan, G. (2009). Ultrasound assisted enhancement in natural dye extraction from beet root for industrial applications and natural dyeing of leather. *Ultrasonics Sonochemistry, 16*(6), 782-789.

Sivakumar, V., Vijaeeswarri, J., & Anna, J. L. (2011). Effective natural dye extraction from different plant materials using ultrasound. *Industrial Crops and Products, 33*(1), 116-122.

Sun, Y., Liu, D., Chen, J., Ye, X., & Yu, D. (2011). Effects of different factors of ultrasound treatment on the extraction yield of the all-trans-b-carotene from citrus peels. *Ultrasonics Sonochemistry, 18*(1), 243–249.

Tabaraki, R., & Nateghi, A. (2011). Optimization of ultrasonic-assisted extraction of natural antioxidants from rice bran using response surface methodology. *Ultrasonics Sonochemistry, 18*(6), 1279-1286.

Tiwari, P., Kumar, B., Kaur, M., Kaur, G., & Kaur, H. (2011). Phytochemical screening and extraction: A review. *Internationale Pharmaceutica Sciencia, 1*(1), 98-106.

Vajic, U. J., Grujic-Milanovic, J., Zivkovic, J., Savikin, K., Godevac, D., Miloradovic, Z., … & Mihailovic-Stanojevic, N. (2015). Optimization of extraction of stinging nettle leaf phenolic compounds using response surface methodology. *Industrial Crops and Products, 74*, 912–917.

Veggi, P. C., Santos, D. T., & Meireles, M. A. A. (2011). Anthocyanin extraction from Jabuticaba (*Myrciaria cauliflora*) skins by different techniques: Economic evaluation. *Procedia Food Science, 1*, 1725-1731.

Victor, E. (2013). *Cocus nucifera* (coconut) fruit: A review of its medical properties. *Journal of Advance in Agriculture, Sciences and Engineering, 3*, 718–723.

Xu, D., Zheng, J., Zhou, Y., Li, Y., Li, S., & Li, H. (2017). Ultrasound-assisted extraction of natural antioxidants from the flower of *Limonium sinuatum*: Optimization and comparison with conventional methods. *Food Chemistry, 217*, 552-559.

Zhang, H., Zhang, X., Yang, X., Qiu, N., Wang, Y., & Wang, Z. (2013). Microwave assisted extraction of flavonoids from cultivated *Epimedium sagittatum*: Extraction yield and mechanism, antioxidant activity and chemical composition. *Industrial Crops and Products, 50*, 857–865.

Zou, Y., Xie, C., Fan, G., Gu, Z., & Han, Y. (2010). Optimization of ultrasound-assisted extraction of melanin from *Auricularia auricula* fruit bodies. *Innovative Food Science and Emerging Technologies, 11*(4), 611-615.

Zulrushdi, N. A. F. (2018). *Extraction and characterization of natural dye from mesocarp and exocarp of cocos nucifera using microwave assisted extraction* (Master's thesis). Universiti Selangor, Malaysia.

Distinct pH-dependent Aggregation of Citrate-Capped Colloidal Gold in Presence of Citrate Competitors

Fatemeh Javadi-Zarnaghi[1]*, Fahimeh Hosseini[2] and Dorsa Mohammadrezaei[3]

[1]*Department of Biology, Faculty of Sciences, University of Isfahan, Isfahan, Iran*
[2]*Mechanical and Industrial Engineering Department, University of Toronto, Toronto, Canada*
[3]*Department of Chemical Engineering, Isfahan University of Technology, Isfahan, Iran*

ABSTRACT

Many colorimetric biosensors utilise citrate-capped gold nanoparticles ($AuNP_{cit}$) in combination with functional macromolecules. Proper function of the utilised macromolecules is strongly dependent on the buffer systems. However, it is well known that solvents and buffers might cause aggregation of nanoparticles. A comprehensive and systematic investigation on the effect of buffer composition, concentration and pH on the aggregation of $AuNP_{cit}$ is reported in this study. Distinct aggregation behaviours were observed in acidic and basic pH. In acidic pH, the increase in pH, caused stabilisation of $AuNP_{cit}$, while in basic pH, the stability was dependent on the ionisation degree of the applied buffer. Theoretical analyses revealed that ionic buffer species act as citrate competitors and control aggregation of $AuNP_{cit}$. Understanding the fundamental principles of competition between citrate and buffer components allows scientists to choose orthogonal conditions for development of gold nanoparticle-based biosensors which guarantee stability of gold nanoparticles and proper folding of macromolecules simultaneously.

Keywords: Aggregation, citrate buffer, gold nanoparticles, Good's buffers, Henderson-Hasselbach equation

INTRODUCTION

Detecting threatening substances and microorganisms is an essential requirement for human health and environmental care. Nowadays, biosensors play an important role on detection of such molecular hazards. As defined by the international union of pure and applied chemistry (IUPAC), biosensors are "chemical sensors in which the recognition system utilizes a biochemical mechanism". The recognition elements of biosensors are

usually composed of peptides (Tang et al., 2012), enzymes, antibodies (Lazcka et al., 2007) and single stranded DNAs such aptamers and (deoxy)ribozymes (Endo et al., 2005; Zhao et al., 2008). Each of these macromolecules requires specific and distinct buffer with defined pH, concentration and composition for proper fold and function.

Biosensors benefit from several methods of data acquisition including colorimetric readouts (Chen et al., 2015). Most colorimetric methods depend on special characteristics of gold nanoparticles (AuNPs, colloidal gold). Gold nanoparticles have an intensive and distinctive colour which is due to the localised surface plasmon resonance (LSPR) effect (Liu et al., 2006; Petryayeva et al., 2011). The position of the LSPR peak in the visible spectrum depends on the shape, size and dispersity of the gold nanoparticles in the colloid. A well-dispersed AuNP colloid bears an LSPR peak around 510-530 nm and is red (Petryayeva et al., 2011). The peak intensities are reduced and are shifted to 620-670 nm upon aggregation. As a result of such red-shift, the colour of the colloid turns blue (Niu et al., 2014).

Since many colorimetric AuNP-based biosensors harness the colour of the colloid as the readout of the sensor, the aggregation of gold nanoparticles and thus the LSPR red-shift must be monitored strictly. There are two strategies to implement the colour-change of the AuNP-based biosensors. In one strategy, the gold nanoparticles are designed to aggregate upon presence of the target molecules. In the others, the target stabilises the gold nanoparticles and inhibits the aggregation (Lonne et al., 2014). In either of the strategies, it must be guaranteed that aggregation of gold nanoparticles does not occur spontaneously.

The gold nanoparticles used in many of reported biosensors are synthesised by a well-known method of reduction of hydrogen tetrachlorocuprate (III) with citrate. The colloids of citrate-capped gold nanoparticles ($AuNP_{cit}$) are prone to aggregate when subjected to extra additives. It is well known that $AuNP_{cit}$ aggregates in presence of solutions of chlorides and other halides (Zhang et al., 2014). Aggregation of gold nanoparticles was also reported in presence of organic molecules such as urea, thiourea, glutathione and ethanolamine (Chegel et al., 2012).

Susceptibility of $AuNP_{cit}$ to aggregation in the presence of additives complicates choosing a proper buffer system. In one hand, it must be assured that undesired aggregation of gold nanoparticles does not ensue with non-target species of the buffer. On the other hand, the buffer system must accommodate functional folding of the macromolecules of the biosensor. Hence, applying orthogonal conditions with defined buffer component, concentration and pH is crucial for development of colorimetric biosensors. The orthogonal condition must support proper folding and activity of the macromolecule and simultaneously, must not induce gold nanoparticle aggregation.

In this respect, there is a demand for comprehensive and systematic investigations of the effects of the common buffer systems on the aggregation of $AuNP_{cit}$, which has not been studied, to the best of our knowledge.

A handful of buffers over the range of pH 3.0-9.0 were selected to be analysed in this study. The buffer components have been either citrate or members of Good's buffer that are of valuable interest in pharmacy (March, 2004), industry (Zhou et al., 2010) and biosensor development (Jiang et al., 2008). Citric acid buffer systems are one of the common buffers for development of biosensors (Ang et al., 2015). Citrate buffer is a carboxylic acid buffer with three acidic protons that compose a continues range of buffering system from ca. 2.0 to

7.0. Good's buffers (Good et al., 1972) are N-substituted taurine or glycine molecules and have specific properties which prioritize them as biological buffers. Solubility, non-toxicity, little absorption within UV and visible spectrum and less interaction with metal ions are some advantageous properties of Good's buffers. Moreover, the pKa values of Good's buffers are beneficially less dependent on the concentration of the buffer and temperature.

Here, the effect of citrate and several Good's buffers with different pH is presented on the aggregation of citrate-capped colloidal gold (Figure 1). An aggregation constant is reported for each condition that help molecular biologists to decide most suitable buffer for individual biosensors. It is demonstrated that the aggregation of citrate-capped gold nanoparticles follows distinct mechanisms in acidic and basic pH. Theoretical explanations are included to understand the particular response of the gold nanoparticles to various buffer systems.
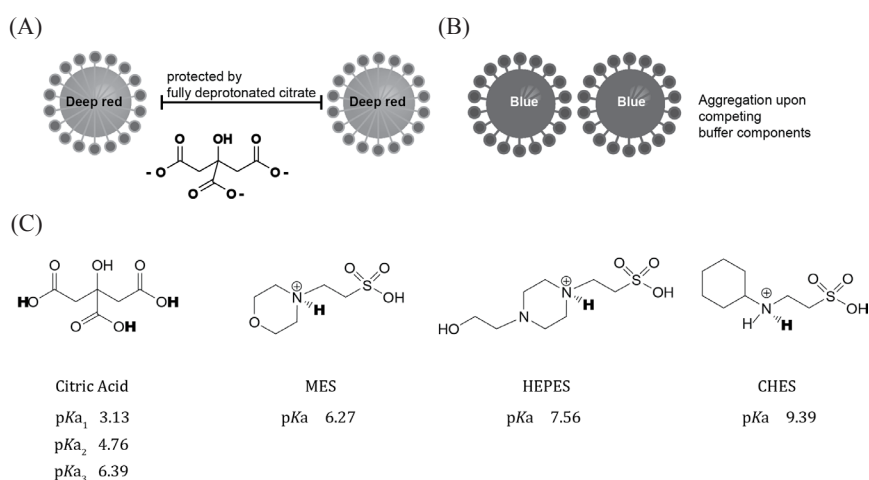

Figure 1. (A) Schematic illustrations of the coordination of the positive charges at the surface of gold nanoparticles with citrate species. (B) Shcematic representation of the aggregated form of gold nanoparticles. (C) Illustration of the buffers used in this study in their fully protonated form. The protons for which the pKa values are marked with bold letters. MES: 2-(N-Morpholino)ethanesulfonic acid, CEHS: 2-(Cyclohexylamino) ethanesulfonic acid, HEPES: 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

MATERIALS AND METHODS

Chemicals

Good's buffers i.e. 2-(N-Morpholino)ethanesulfonic acid hydrate (MES), 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES) and 2-(Cyclohexylamino)ethanesulfonic acid (CHES), citric acid monohydrate, trisodium citrate dehydrate, sodium hydroxide and hydrogen tetrachlorocuprate (III) (HAuCl4) were purchased from Sigma-Aldrich.

Synthesis of Gold Nanoparticles

All glassware were washed with aqua regia (3:1 concentrated HC$_1$:HNO$_3$) for 30 min and subsequently rinsed with generous amount of deionised water and then ultrapure water (18.2

$M\Omega \cdot cm^{-1}$). Synthesis of gold nanoparticles was based on earlier studies (Liu et al., 2006). To begin the synthesis, 100 ml of 1 mM $HAuCl_4$ was brought to reflux on a heater with a magnetic stirrer, consequently 10 ml 38.8 mM sodium citrate was added. The change from pale yellow to deep red was observed within one minute. The system refluxed for further 20 minutes before it cooled down to room temperature while still stirring.

Dynamic Light Scattering (DLS) Measurement

DLS was measured on a DynaPro (Wyatt Technology) with diluted synthesised gold nanoparticles. The dilution factor was 1:100. The laser intensity of the instrument was set to 80% and measurements were performed in 2 ml cuvettes.

Buffer Preparation

Citrate buffer pH 3.0 and 5.0, MES buffer pH 6.0, HEPES buffers pH7.0, 7.5 and 8.0 were prepared by dissolving appropriate amount of buffer in 100 ml of ultrapure water. After pH adjustment with 2 N NaOH, the volume adjusted to 200 ml to achieve 1 M buffer solutions. CHES buffers pH 9.0, 9.5 and 10.0 were prepared in 0.5 M stock concentration based on the same procedure.

Titration Procedure

Measurements were performed in a total volume of 1 ml. To 250 microlitre of citrate-capped AuNP, appropriate amount of ultrapure water and appropriate amount of the concentrated buffer were added with the stated order. Final concentrations of buffers were 250, 200, 150, 100, 80, 50, 20, 10 and 5 mM. The mixtures gently pipetted up and down and incubated for 30 minutes at room temperature. The spectra of AuNPs incubated with buffers were measured in the visible range of 450 nm to 750 nm in a quartz cuvette. Titration series for each buffer composition and pH were repeated three times with three stock buffer solutions which were prepared independently.

Data Analysis

The ratio of absorbance at the peak of aggregated forms (620 nm) to the absorbance of the peak of non-aggregated form (520 nm) is usually considered as a measurement for aggregation (Mei et al., 2013; Song et al., 2012; Song et al., 2011). For each individual spectrum, the absorbance at 620 nm was divided to the absorbance at 520 nm to reach the A_{620} nm/A_{520} nm

ratio (Absorbance ratio, AR) as a measure of aggregation. The outcome was plotted as bar graphs. To eliminate the background, progress of aggregation (PAG) was defined as equation 1:

$$\text{PAG} = \left[A_{620\,nm}/A_{520\,nm}\right] - \left[A_{620\,nm,\ no\ buf.}/A_{520\,nm,\ no\ buf.}\right] \tag{1}$$

The PAG Values were plotted against the concentration of the buffer system as scatter. The Hill equation was fitted to the data (equation 2).

$$\text{PAG} = \frac{\text{PAG}_{max}[buf.]^n}{K_{Agg.} + [buf.]^n} \tag{2}$$

In equation 2, PAG is a measure of an increase in the absorbance ratio (AR) and shows progress of aggregation. $KA_{gg.}$ is aggregation constant. The concentration of buffer is denoted as [*buf.*] and the letter n, indicates the Hill constant which is a measure of cooperativity.

RESULTS

Dynamic light scattering (DLS) or photon correlation spectroscopy (PCS) is a physical method that is usually utilised for measuring the size distribution of nano-scale particles in a suspension or macrocmolecules in solutions (Stetefeld et al., 2016). The size distribution of the synthesised gold nanoparticle was characterised by DLS. The population mode of particles in the synthesised gold colloid was 11 nm and 90% of the particles had the size distribution between 6-14 nm.(Figure 2A). The spectrum of the citrate-capped synthesised AuNP had a localised surface plasmon resonance peack at 520 nm (Figure 2B). The conditions at which buffers cause aggregation were observed. Such aggregations are non-target and must be avoided in development of biosensors. As an example, presence of 250 mM citrate buffer pH 5.0 caused a red-shift in LSPR peak (Figure 2C). The LSPR band shift is usually due to a reflection of aggregaion of the gold nanoparticles (Mei et al., 2013). The-red shift in the absorption spectra was also visible with the change in the colloidal gold from red to blue/purple. Such buffer-induced red-shifts that is independent of the presence of specific targets must be avoided during development of biosensors. In this study, comprehensive studies were performed by titration experiments. Titration of the citrate-capped gold nanoparticles with individual buffers resulted in a different degree of aggregation. The aggregation level depended on the composition and final concentration of buffers. The LSPR band shifted from 520 nm to 620 nm when high concentrations of buffers were present.
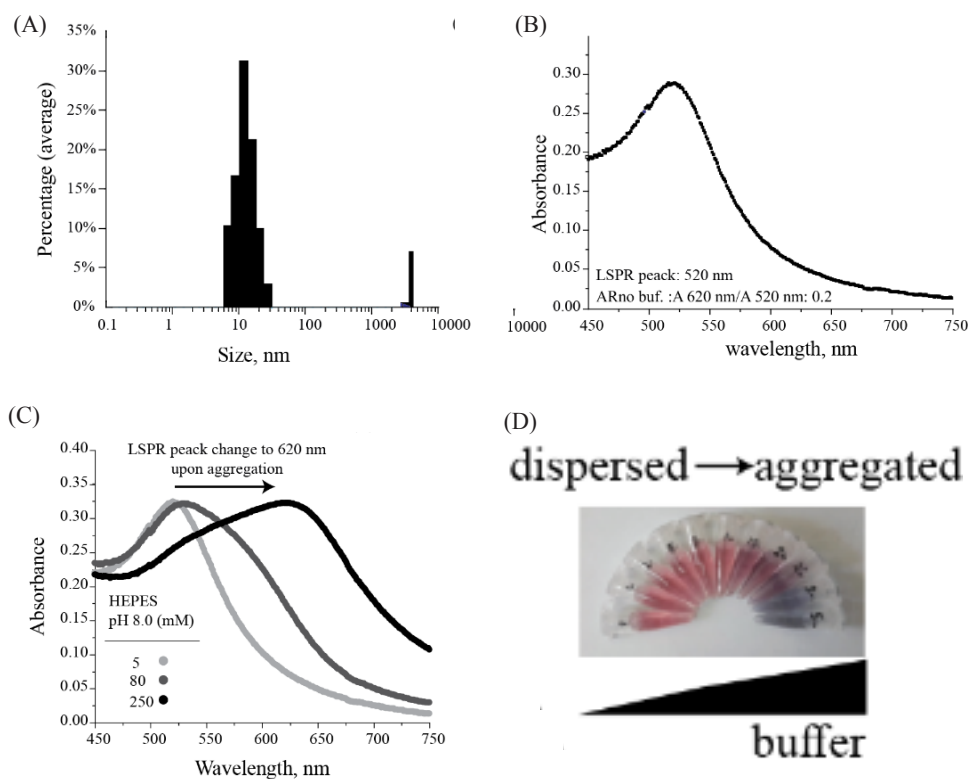
Figure 2. (A) Dynamic Light Scattering (DLS) data for the synthesized citrate-capped gold nanoparticles in this study. (B) The visible spectrum of the synthesised nanoparticle (4 times diluted sample). (C and D) Representative spectra and colloid color change upon presence of the buffers. The LSPR peak shifts toward longer wavelength upon aggregation and the colour of the colloid changes to blue

Increase in the ratio of the spectral absorption at 620 nm to the spectral absorption at 520 nm (A_{620} nm/A_{520} nm, absorbance ratio, AR) is a measure of aggregation (Mei et al., 2013; Song et al., 2012; Song et al., 2011). The AR had a background amount which was the AR for the monodisperesed citrate-capped gold nanoparticles in the absence of any additive ($AR_{no\ buf.}$). The synthesised gold nanoparticles in this study had an $AR_{no\ buf.}$ of 0.20 (Figure 2B). Figure 3 illustrates the AR for each condition as a measure of aggregation level. As AR increases, species that are transformed from monodispersed to aggregate form get higher percentage of the population. At AR above 1.0 the population of aggregated forms are predominated. Thus, conditions at which AR is above 1.0, should be avoided in the development of biosensors.

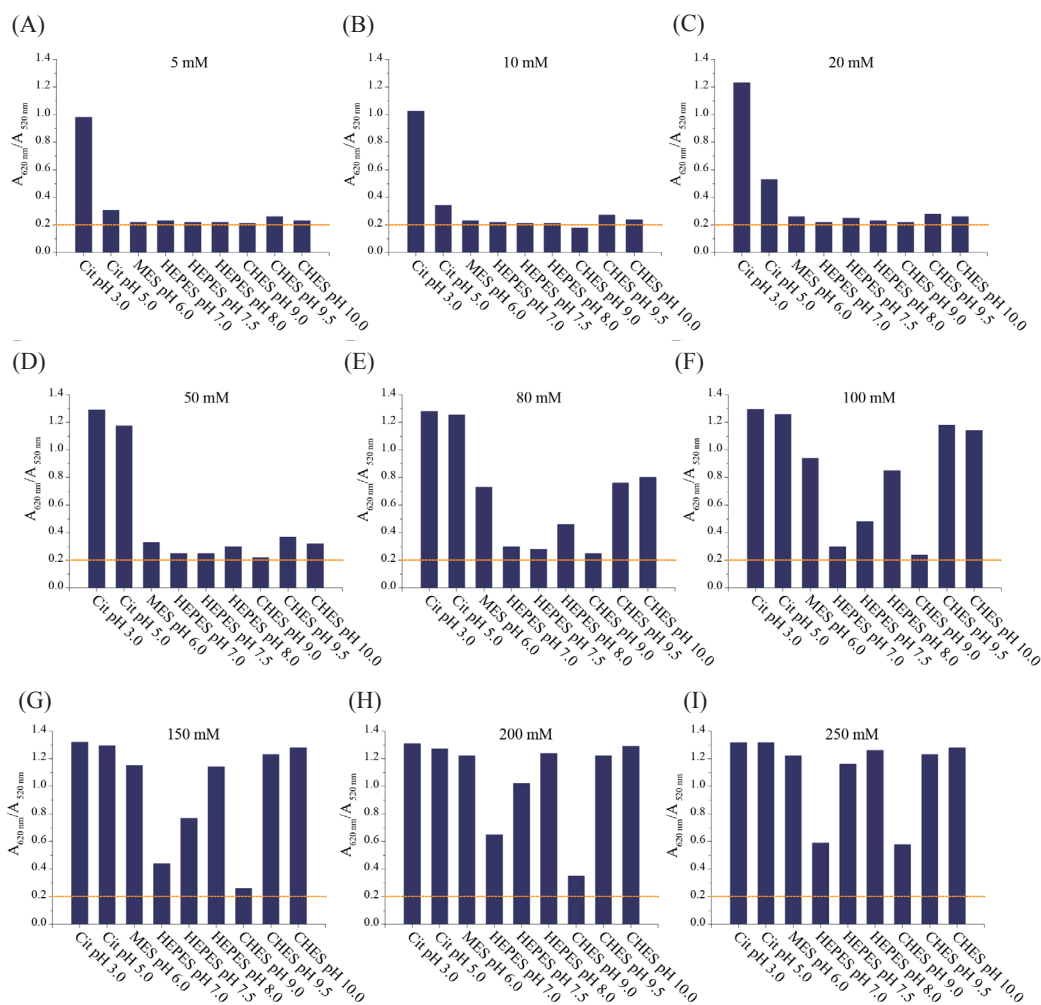*Figure 3.* The A_{620 nm}/A_{520 nm} (AR) as a measure of aggregation level of citrate-capped gold nanoparticles in different pH from 3.0 to 10.0. Concentration of the applied buffer is increased from 5 mM to 250 mM from pannels a to i. The background AR is shown with the dashed orange lines

Addition of the citrate buffer pH 3.0 to the AuNP colloid caused a severe aggregation even at low concentrations. At 5 mM pH 3.0 citrate buffer, the AR was 0.98 which was ca. 5 times higher than synthesised gold nanoparticle in the absence of any additive. Increasing the concentration of citrate buffer pH 3.0 to 250 mM increased AR ratio to 1.31. Citrate buffer pH 5.0 showed less harsh effects. Presence of 5 mM pH 5.0 citrate buffer had almost no effect on AR. Increasing the concentration of citrate buffer pH 5.0 to 20 mM resulted in AR of 0.52 which was ca. 2.5 times higher than synthesised gold nanoparticle. At 250 mM this ratio reached 1.31.

2-(N-Morpholino)ethanesulfonic acid (MES) buffer provided the pH 6.0 for this study. In the presence of MES buffer pH 6.0, the colloidal gold resisted for aggregation up to 80 mM. Increasing the concentration of MES buffer pH 6.0 to 250 mM resulted in AR ratio of 1.22 which was slightly less severe change compared with what was observed with citrate buffers pH 3.0 and 5.0.

Presence of 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer pH 7.0 was best tolerated even at relatively high concentrations. The AR reached only 0.54 when the concentration of HEPES buffer pH 7.0 was raised to 250 mM. Increasing the pH of HEPES buffer to 7.5 and 8.0 returned sensitivity of the gold nanoparticles. The AR reached 1.16 and 1.26 at 250 mM buffer for pH 7.5 and 8.0 respectively. However, the effect was only visible in concentrations higher than 80 mM. Notably, although at 250 mM the HEPES pH 7.5 and 8.0 almost showed the same effect as citrate pH 3.0 and 5.0 and MES pH 6.0, the instability of gold nanoparticle colloids was less noticeable for lower concentrations of HEPES buffers.

Changing the buffer composition to 2-(Cyclohexylamino)ethanesulfonic acid (CHES) had a stabilising effect. CHES buffer pH 9.0 showed very similar effect as HEPES buffer pH 7.0. The AR at 250 mM CHES buffer was 0.58. In other words, pH 9.0 was better tolerated than pH 8.0, keeping in mind that the buffer composition was changed from HEPES to CHES. In presence of CHES buffer pH 9.5 and 10 the sensitivity returned and the AR reached 1.03 and 1.08 at 250 mM respectively (Figure 4).
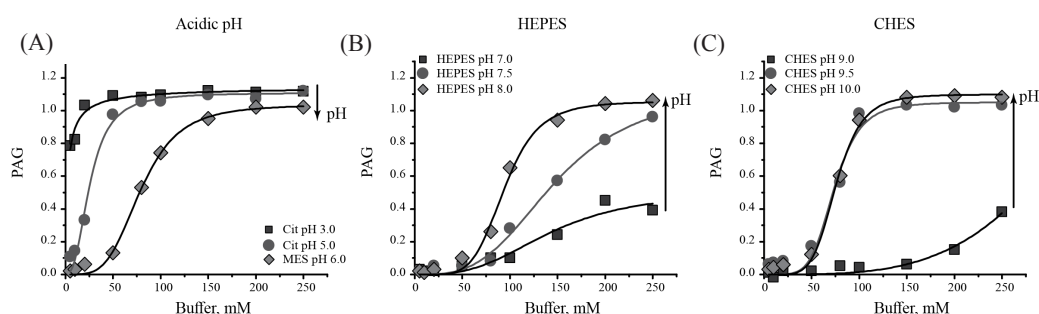

Figure 4. Progress of aggregation (PAG) of citrate-capped gold nanoparticle colloids in the presence of increasing concentrations of buffers. The aggregation phenomena was distinct in acidic and basic pH. Additionally, changing the buffer component from HEPES to CHES has a stabilising effect

In the next step, progress of aggregation (PAG) was defined as an increase in the AR. Aggregation constant ($K_{agg.}$) as the concentration of buffer at which PAG reaches half of its maximum value. Figure 4 illustrates the titration curves. The $K_{agg.}$ and Hill coefficient for each buffer composition and pH were achieved by fitting the Hill equation and are shown in Table 1.

Table 1
Fit values of the Hill curve for the PAG

Buffer	Citrate		MES	HEPES				CHES	
pH	3.0	5.0	6.0	7.0	7.5	8.0	9.0*	9.5	10.0
PAG_{max}	1.15	1.11	1.04	0.51	1.13	1.06	n.d.	1.05	1.10
$K_{Agg.}$, mM	2.24	25.88	79.33	142.18	147.92	94.26	n.d.	73.77	75.92
Hill constant	0.84	2.54	3.99	3.05	3.28	5.31	n.d.	5.49	5.68

* Fit values of CHES pH 9.0 are not defined since the data did not reach to a plateau of aggregation level within the concentration range

Distinct observations were achieved with acidic and basic pH (Figure 4A vs. B and C). In addition, exchanging the buffer system from HEPES to CHES caused a clear difference (Figure 4B and C). From pH 3.0 to 6.0, stability of gold nanoparticles was increased with an increase in pH. From pH 7.0 to 8.0, with HEPES buffer stability was decreased with an increase in pH. Changing the buffer from HEPES 8.0 to CHES pH 9.0, had a stabilising effect. From pH 9.0 to 10.0, again the stability was decreased with an increase in the pH.

In the acidic buffers, increasing pH (from pH 3.0 to 6.0, Figure 4A), had almost no effect on PAG_{max}, while K_{agg} was increased from ca. 2 mM to 80 mM. Constant PAG_{max}, with increased K_{agg}, was attributed as increased stability of the gold nanoparticles. The cooperativity of aggregation phenomenon (n) was also increased. Aggregation at near-neutral pH was more dependent on the concentration of the buffer.

Despite acidic buffers, increasing pH with HEPES buffer (from pH 7.0 to 8.0, Figure 4B), was destabilising. PAG_{max} was increased from ca. 0.5 to ca. 1.1, while K_{agg} was decreased from ca. 140 mM to ca. 94 mM. The cooperativity of aggregation was increased upon increase in the pH. Hence, aggregation at higher pH was shown to be dependent on the concentration of the buffer.

Changing HEPES to CHES buffer caused stability of the colloidal gold. However, increase in pH with the CHES buffer had the same effects as with the HEPES buffer i.e. stability decreased upon increase in pH (Figure 4C).

In summary, distinct effects on the stability of colloidal gold upon increase in pH was observed. In acidic pH, the citrate-capped gold nanoparticles were stabilised when the pH increased. The effect was independent of the used buffer. In contrast, in basic pH, increased pH with a unique buffer resulted in increased instability. Finally, the stability of the colloidal gold was strongly dependent on the buffer type and concentration.

DISCUSSION

In this study, citrate-capped gold nanoparticles were subjected to increasing concentrations of various buffers. Citrate buffers were chosen for highly acidic pH (3.0-5.0). Good buffers were chosen for pH 6.0 and above. A different pattern of aggregation was observed for acidic and basic pH. It can be hypothesised that "accumulation of Cit^{3-} and its competitors" control such distinctive observations. Citrate-capped gold nanoparticles are best stabilised when the concentration of fully deprotonated form of citrate (Cit^{3-}) is at maximum. At very acidic pH, protonated forms of citrate are present which cannot coordinate at the surface of the gold nanoparticle. Hence, the colloid is sensitive to buffer moieties, even at very low concentrations. In pH above the pKa_3 of citrate (basic pH), all citrates are in the fully deprotonated form (Cit^{3-}) which supports maximum stability for gold nanoparticles. However, within this pH range, the deprotonated forms of the applied buffers compete with the Cit^{3-}. At pH above pKa of the applied buffer, the competition is in favour of buffer competitors since more deprotonated species of buffer are available to compete with Cit^{3-}.

Such competitions are explained below. Citric acid has three acidic functional groups. Calculating exact percentage of citric acid ionised species with Henderson-Hasselbach equation (H-H) supports the hypothesis of this study. H-H equation defines the relationship between pH,

pKa and the logarithmic ratio of deprotonated form of an acid to its protonated form. Based on H-H equation, three constant values (Q_1-Q_3) are defined as the ratios of deprotonated to protonated forms of citrate. In addition, Q_4 is defined as total concentration (100%) of citrate species (equations 3-6).

$$Q_1 = CitH_2^- / CitH_3 = 10^{(pH-pKa_1)} \tag{3}$$

$$Q_2 = CitH^{2-} / CitH_2^- = 10^{(pH-pKa_2)} \tag{4}$$

$$Q_3 = Cit^{3-} / CitH^{2-} = 10^{(pH-pKa_3)} \tag{5}$$

$$Q_4 = CitH_3 + CitH_2^- + CitH^{2-} + Cit^{3-} = 100\% \tag{6}$$

The four equations above are solved with a linear algebraic method to achieve exact percentage of each of the deprotonated form of citrate at any desired pH (equations 7-10). The theoretical percentage of each citrate species at given pH in this study is calculated using these equations (Table 2).

Table 2
Calculated percentage of each citrate species at pH used in this study

	pH	$CitH_3\%$	$CitH_2^-\%$	$CitH^{2-}\%$	$Cit^{3-}\%$
acidic	3.0	57.006	42.259	0.734	0.0003
	5.0	0.478	35.434	61.578	2.508
	6.0	0.005	3.927	68.259	27.807
neutral	7.0	1.53E-05	0.113	19.686	80.199
	7.5	5.59E-07	0.013	7.202	92.784
	8.0	1.86E-08	0.001	2.395	97.602
basic	9.0	1.90E-11	1.41E-05	0.244	99.755
	9.5	6.02E-13	1.41E-06	0.077	99.922
	10.0	1.90E-14	1.41E-07	0.024	99.975

$$CitH_3 = Q_4 / (1 + Q_1 + Q_1Q_2 + Q_1Q_2Q_3) \tag{7}$$

$$CitH_2^- = Q_1Q_4 / (1 + Q_1 + Q_1Q_2 + Q_1Q_2Q_3) \tag{8}$$

$$CitH^{2-} = Q_1Q_2Q_4 / (1 + Q_1 + Q_1Q_2 + Q_1Q_2Q_3) \tag{9}$$

$$Cit^{3-} = Q_1Q_2Q_3Q_4 / (1 + Q_1 + Q_1Q_2 + Q_1Q_2Q_3) \tag{10}$$

As shown in Table 2, theoretically, at pH 3.0, 57% of citrate is in the fully protonated form ($CitH_3$), 42% is di-protonated ($CitH_2^-$), 0.73% mono-protonated ($CitH^{2-}$) and only 0.0003% is in the fully deprotonated format (Cit^{3-}). In contrast to pH 3.0, at pH 7.0, less than 0.00002

% is fully protonated, 0.11 % is di-protonated, 19.7% is mono-protonated and 80% of the citrate is fully deprotonated. As it is clear, the species of Cit^{3-} accumulates upon increase in pH as the concentration of deprotonated forms increases from pH 3.0 to 7.0. Notably, at basic pH, the concentrations of protonated species of citrate turn to be negligible.

The fully protonated form of citric acid cannot serve as the stabilising agent for colloidal gold. In contrast the citrate species with more negative charges better stabilises the colloidal gold. Therefore, accumulation of Cit^{3-} species proportionally, increases the stability of the colloidal gold and decreases aggregation level (Figure 5A and Figure 5B). For instance, at pH
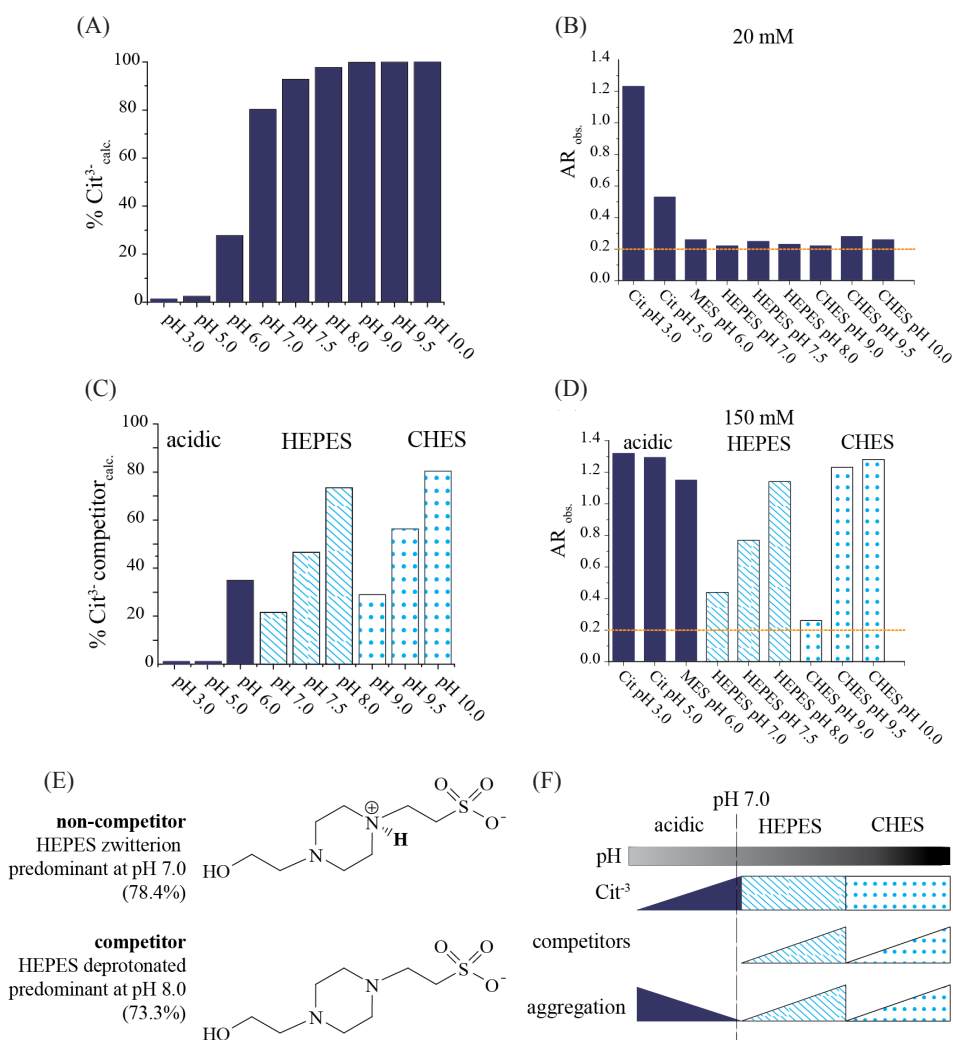

Figure 5. (A) Calculated percentage of cit3- species at given pH values. (B) Observed absorption ratio (AR obs.) of the citrate-capped gold nanoparticles in presence of 20 mM buffer of the indicated buffers. (C) Calculated percentage of the deprotonated forms of the MES, HEPES and CHES which act as competitors of the cit3-. (D) Observed absorption ratio (AR obs.) of citrate-capped gold nanoparticles in presence of 150 mM buffer of the indicated buffers. (E) The zwitterion and deprotonated forms of HEPES buffer. (F) A summary of the accumulation of cit3- and its competitors and their effects on the aggregation level of the citrate-capped gold nanoparticles. Acidic pH: solid fill; basic pH, HEPES: dashed fill; basic pH, CHES: dotted fill

3.0, most citric acid molecules are in the fully protonated form (57%) and cannot protect gold nanoparticles from aggregation and cause severe instability within the titration range (figure 3, first column in all panels).

At pH above 7.0, the concentration of Cit^{3-} is theoretically at its maximum and thus, a good stability is expected. However, new players come into play. As the pH increases the deprotonated form of the present buffer (HEPES or CHES) accumulates and competes with the Cit^{3-} (Figure 5E). Henderson-Hasselbach equation is used to calculate exact percentage of Good's buffers component at all pH that have been studied here (Table 3).

Table 3
Calculated percentage of deprotonated forms of the Good buffer that are used in this study

Used buffer	pKa	pH	HA	A-
MES	6.27	6	65.1	34.9
HEPES	7.56	7	78.4	21.6
		7.5	53.4	46.6
		8	26.6	73.4
CHES	9.39	9	71.1	28.9
		9.5	43.7	56.3
		10	19.7	80.3

Interestingly, the observed trend in the percentage of A⁻ (competitor form of the buffer) is proportional to the observed trend in aggregation of gold nanoparticles (Figure 5C and Figure 5D). As the pH increases from 7.0 to 8.0, the percentage of competitor form of HEPES (Figure 5E) is increased from 22% to 73%, which was accompanied by increase in aggregation (AR). Changing the buffer system to CHES, caused a drop in the percentage of the present competitor (since the pKa value of CHES is higher than HEPES), which was interestingly in concert with a stabilising effect on the gold nanoparticle, even though the pH is increased. As the pH increased from 9.0 to 10.0 the percentage of the competitor form of CHES increased from 28% to 80% and again, and this increase is reflected in the instability of gold nanoparticles. Competitor dependent displacement of citrate with halide ions was proposed to be the reason of aggregation by Zhang et al. (2014).

As expected, the effect of the competitors depends on the concentration of the used buffer. The higher concentration would increase competitor species to compete with the 3.9 mM Cit^{3-} which was present in the synthesised gold nanoparticles. The dependence of the competitor effect on the concentration of the used buffer was also well reflected in the Hill constants of the fit curves in titration experiments.

Altogether, titration of gold nanoparticles with different buffer composition, pH and concentration, revealed that the stability of the citrate-capped gold nanoparticles is dependent on all three items. Buffer type, the pH of the buffer and its concentration all can affect aggregation. In acidic pH, the accumulation of Cit^{3-} is the major player, since no competitor is present. Therefore, stability is increased upon increases in pH. In basic pH, stability us a

dependent on the competitor concentration. Accumulation of competitors causes aggregation of gold nanoparticles.

CONCLUSION

This is a comprehensive and systematic study on the aggregation of the sensitive citrate-capped gold nanoparticles in presence of generally used buffers in molecular biology and biochemistry. It is shown the stabilisation of gold nanoparticles that are capped with citrate is dependent on both the pH and the component of the buffer. In acidic conditions, the higher pH, forms higher ratio of Cit^{3-} in comparison to other citrate ionisation states. Hence, as the pH approaches to 7.0, the gold nanoparticles are better stabilised. In basic buffers, however, pH is not the sole ruler of the system for aggregation and at pH above 7.0, the ionisation degree of the applied buffer has a major impact on the spontaneous aggregation of the gold nanoparticles. This study promotes development of colorimetric biosensors that are based on gold-nanoparticles and allows scientists to better decide for the concentration and pH of the buffer, which prevents non-desired aggregation of the gold-nanoparticle and simultaneously, is compatible for the folding of their biomolecules.

LIST OF ABBREVIATIONS

Localised surface plasmon resonance (LSPR)
Gold nanoparticles (AuNP)
Progress of aggregation (*PAG*)
A620 nm/A520 nm, absorbance ratio (AR)

ACKNOWLEDGMENTS

REFERENCES

Ang, L. F., Por, L. Y., & Yam, M. F. (2015). Development of an amperometric-based glucose biosensor to measure the glucose content of fruit. *PLoS One, 10*(3), e0111859. doi: 10.1371/journal.pone.0111859

Chegel, V., Rachkov, O., Lopatynskyi, A., Ishihara, S., Yanchuk, I., Nemoto, Y., … & Ariga, K. (2012). Gold nanoparticles aggregation: Drastic effect of cooperative functionalities in a single molecular conjugate. *The Journal of Physical Chemistry C, 116*(4), 2683-2690. doi: 10.1021/jp209251y

Chen, G., Guo, Z., Zeng, G., & Tang, L. (2015). Fluorescent and colorimetric sensors for environmental mercury detection. *Analyst, 140*(16), 5400-5443. doi: 10.1039/c5an00389j

Endo, T., Kerman, K., Nagatani, N., Takamura, Y., & Tamiya, E. (2005). Label-free detection of peptide nucleic acid-DNA hybridization using localized surface plasmon resonance based optical biosensor. *Analytical Chemistry, 77*(21), 6976-6984. doi: 10.1021/ac0513459

Good, N. E., & Izawa, S. (1972). Hydrogen ion buffers. [Comparative Study]. *Methods in Enzymology, 24*, 53-68.

Jiang, Z., Liang, A., Li, Y., & Wei, X. (2008). Immunonanogold-catalytic Cu_2O-enhanced assay for trace penicillin G with resonance scattering spectrometry. *IEEE Trans Nanobioscience, 7*(4), 276-283. doi: 10.1109/tnb.2008.2011860

Lazcka, O., Del Campo, F. J., & Munoz, F. X. (2007). Pathogen detection: A perspective of traditional methods and biosensors. *Biosensors and Bioelectronics, 22*(7), 1205-1217. doi: 10.1016/j.bios.2006.06.036

Liu, J., & Lu, Y. (2006). Preparation of aptamer-linked gold nanoparticle purple aggregates for colorimetric sensing of analytes. *Nature Protocols, 1*(1), 246-252. doi: 10.1038/nprot.2006.38

Lonne, M., Zhu, G., Stahl, F., & Walter, J. G. (2014). Aptamer-modified nanoparticles as biosensors. *Advances in Biochemical Engineering/Biotechnology, 140*, 121-154. doi: 10.1007/10_2013_231

March, J. B. (2004). Improved formulations for existing CBPP vaccines recommendations for change. *Vaccine, 22*(31-32), 4358-4364. doi: 10.1016/j.vaccine.2004.03.066

Mei, Z., Chu, H., Chen, W., Xue, F., Liu, J., Xu, H., … & Zheng, L. (2013). Ultrasensitive one-step rapid visual detection of bisphenol A in water samples by label-free aptasensor. *Biosensors and Bioelectronics, 39*(1), 26-30. doi: 10.1016/j.bios.2012.06.027

Niu, S., Lv, Z., Liu, J., Bai, W., Yang, S., & Chen, A. (2014). Colorimetric aptasensor using unmodified gold nanoparticles for homogeneous multiplex detection. *PLoS One, 9*(10), e109263. doi: 10.1371/journal.pone.0109263

Petryayeva, E., & Krull, U. J. (2011). Localized surface plasmon resonance: Nanostructures, bioassays and biosensing--A review. *Analytica Chimica Acta, 706*(1), 8-24. doi: 10.1016/j.aca.2011.08.020

Song, K. M., Cho, M., Jo, H., Min, K., Jeon, S. H., Kim, T., … & Ban, C. (2011). Gold nanoparticle-based colorimetric detection of kanamycin using a DNA aptamer. *Analytical biochemistry, 415*(2), 175-181. doi: 10.1016/j.ab.2011.04.007

Song, K. M., Jeong, E., Jeon, W., Cho, M., & Ban, C. (2012). Aptasensor for ampicillin using gold nanoparticle based dual fluorescence-colorimetric methods. *Analytical and Bioanalytical Chemistry, 402*(6), 2153-2161. doi: 10.1007/s00216-011-5662-3

Stetefeld, J., McKenna, S. A., & Patel, T. R. (2016). Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews, 8*(4), 409-427. doi: 10.1007/s12551-016-0218-6

Tang, J., Wang, J. Y., & Parker, L. L. (2012). Detection of early Abl kinase activation after ionizing radiation by using a peptide biosensor. *Chembiochem, 13*(5), 665-673. doi: 10.1002/cbic.201100763

Zhang, Z., Li, H., Zhang, F., Wu, Y., Guo, Z., Zhou, L., & Li, J. (2014). Investigation of halide-induced aggregation of au nanoparticles into spongelike gold. *Langmuir, 30*(10), 2648-2659. doi: 10.1021/la4046447

Zhao, W., Chiuman, W., Lam, J. C., McManus, S. A., Chen, W., Cui, Y., … & Li, Y. (2008). DNA aptamer folding on gold nanoparticles: from colloid chemistry to biosensors. *Journal of the American Chemical Society, 130*(11), 3610-3618. doi: 10.1021/ja710241b

Zhou, H., Purdie, J., Wang, T., & Ouyang, A. (2010). pH measurement and a rational and practical pH control strategy for high throughput cell culture system. *Biotechnology Progress, 26*(3), 872-880. doi: 10.1002/btpr.369

PERTANIKA
JOURNALS

Experimental Evaluation of Jatropha Oil Methyl Ester (JOME) and Fish Oil Methyl Ester (FOME) in a Compression Ignition Engine with Exhaust Gas Recirculation

K. Bhaskar[1] and S. Sendilvelan[2]*

[1]*Department of Automobile Engineering, Rajalakshmi Engineering College, Chennai, India*
[2]*Department of Mechanical Engineering, Dr M.G.R. Educational and Research Institute,*
University, Chennai, India

ABSTRACT

Simultaneous reduction of soot and oxides of nitrogen (NO_x) is a prime requirement for modern day diesel engine to meet the increasingly stringent emission standards. Exhaust gas recirculation (EGR) is one of the most effective techniques for reducing oxides of nitrogen emissions in diesel engine. This study is an attempt to analyse experimentally the performance and emission characteristics of methyl esters of jatropha oil methyl ester (JOME) and fish oil methyl ester (FOME) blends with diesel with and without exhaust gas recirculation on a stationary single cylinder diesel engine. Compared with the diesel fuel, the performance of 20% methyl ester blends and 20% EGR shows a considerable reduction of oxides of nitrogen 6.1 g/kWh for JOME blends and 6.3 g/kWh for FOME blends compared with 7.3 g/kWh for Diesel. Adverse effects are a reduction of brake thermal efficiency 25.6% for FOME blends and 26% for JOME blends compared with 28.4% for diesel, an increase of unburnt hydrocarbons 0.8 g/kWh for JOME and 0.9 g/kWh for FOME compared to 0.7 g/kWh for diesel and carbon monoxide 23.0 g/kWh for JOME and 25.5 g/kWh for FOME compared to 16.8 g/kWh for diesel. Considering both NO_x and soot emissions, 20% EGR is observed to be optimum for both 20% JOME and 20% FOME.

Keywords: Combustion analysis, Emission control, Exhaust gas recirculation, FOME, JOME

INTRODUCTION

Reducing air pollutants and increasing fuel efficiency of Internal Combustion engines are a primary concern for all developing nations. Fast depleting fossil fuel and its impact on the environmental add to the concern. Bio-diesel taken from edible, non-edible oils and animal fats can be used as a fuel in a compression ignition engine with little or no modifications

(El-Kasaby & Nemit-Allah, 2013; Raj et al., 2010). Hydrocarbons are present in the exhaust gas runs with alternate biofuel blends with diesel (Payri et al., 2009). Researchers conclude oxygenated fuels reduce emissions. Studies also report that nearly 20% of the blends show favourable performance and combustion characteristics when using diesel engine. Higher fatty acid content helps to reduce hydrocarbon emissions. Demirbas analysed the sources of biodiesel and their global projections (Demirbas, 2008).

Bio-fuels are easily available and benefit the environment, economy and consumers. Even though biodiesel offers a reduction in smoke, unburnt hydrocarbon (UBHC) and carbon monoxide (CO) emissions, it produces more nitrogen oxides (NO_x) than the diesel fuel, but it can be reduced by using recirculated exhaust gas (Agarwal et al., 2006). The factors that help in reducing the NO_x emissions are engines which recirculate exhaust gas along with the incoming fuel vapour which decreases oxygen available during combustion which slows down the reaction of the combustion product, so that temperature of the product decreases, which in turn reduces the NO_x (Abd-Alla, 2002) and also the peak combustion temperature is reduced as the heat capacity of vapour and carbon dioxide is higher (Pradhan el al., 2014).

The premixed combustion in an optical diesel engine helps to reduce emissions and improves engine performance (Fang et al., 2009). The results show heat release rate is dominated by this combustion pattern and the injection timing is retarded. There is a delay in the peak heat with increased biodiesel percentage. The fuel particles stick on the engine cylinder walls which is found during the all the test conditions. Recirculated exhaust gas reduces the NO_x emission considerably with biodiesel blends with diesel in a direct injection diesel engine (Ruijun et al., 2009). The percentage increase in recirculated exhaust gas results in longer ignition delay and the combustion is delayed e while the duration of the combustion is reduced. The NO_x emissions are reduced with recirculated exhaust gas but increases with other omissions. The soot emissions increase sharply while using recirculated exhaust gas but NO_x emission decreases considerably (Bhaskar et al., 2013) .

The air-fuel ratio affects exhaust emission due to oxygen availability which affects the exhaust emissions (Murali Manohar et al., 2012; Heffel, 2003). The recirculated exhaust gases increase specific heat and decrease the flame temperature thus, reducing rate of NO_x formation (Abd-Alla, 2002). The recirculated exhaust engines emit a lower quantity of exhaust gases compared with non-EGR engines (Hountalas et al., 2008). At low loads, EGR reduces hydrocarbon emission and at the same time improves brake thermal efficiency (Talibi et al., 2017). The EGR is also used to increase fuel efficiency (Bai et al., 2010; Galloni et al., 2014). However, the use of EGR leads to an increase in soot emissions and emits unburned hydrocarbons (20-30%) compared with conventional engines (Ladommatos, et al., 2000). Therefore, in this work 20% EGR is used to study the performance with jatropha oil methyl ester (JOME) blend with diesel.

MATERIALS AND METHODS

Jatropha Oil Methyl Ester (JOME)

Jatropha oil methyl esters are well-proven alternatives to petroleum diesel. The main reasons for choosing jatropha oil in this work are its environmental friendly nature, cost-effective, the plants are easy to cultivate and need very little water, grow in all climatic conditions and soils, has high yields, and they have 50-year lifespan (Axelsson et al., 2012). Additionally, Jatropha oil can be used as a substitute for petroleum diesel to reduce the cost of importing petroleum products. The jatropha oil is extracted and refined through transesterification process.

The transesterification process is based on the chemical reaction of a triglyceride with alcohol in the presence of a catalyst potassium hydroxide, producing biodiesel and glycerine. Castor, palm, sunflower, pea nut and soybean oils can be used as biodiesel sources, but all these are used for cooking purposes. Hence, instead of using edible oils to produce biodiesel, non-edible oils can be used for the same. Low cost renewable raw material is a very important requirement for economical production of biodiesel.

Among the vegetable oils, jatropha oil has very good properties. It is non-edible oil and has comparatively higher calorific value and cetane rate. Based on literature review, it is the best substitute for fossil fuel.

Fish Oil Methyl Ester (FOME)

Biodiesel produced from fish oil is very effective and a cheaper alternative to diesel. Fish oil is produced from large quantities of tissue waste, such as the viscera, fins, eyes, and tails. These discarded tissue wastes, and by-products can be converted to bio-diesel at a low cost. India has a long coastline with excellent potential for marine fishing. Biodiesel-based fish oil is easy to produce and provides cleaner-burning fuel. Biodiesel and blended diesel, (Petroleum-based diesel mixed with biodiesel), could potentially replace or reduce petroleum-based diesel fuel requirement of the country. Even though biodiesel produces less smoke, UBHC and CO emission and NO_x emissions are higher which can be reduced by using exhaust gas recirculation (EGR).

Fish oil methyl ester has long carbon chain fatty acids to optimise the combustion process and reduces emissions efficiently. The by-product obtained from this transesterification process is glycerine which is used for pharmaceutical and cosmetic purposes. Biodiesel obtained from the trans-esterification of fats and oils is a possible fuel for diesel engines. Transesterification is a chemical process of transforming large triglyceride molecules into smaller, straight chain molecules which are similar to the molecules present in diesel fuel (Leung et al., 2010). The process takes place by reacting to the vegetable oil with an alcohol in the presence of a catalyst. Most modern diesel engines have direct injection fuel systems which are more sensitive to fuel spray quality than indirect injection engines and they require a fuel with properties that are closer to diesel fuel. The fuel properties of biodiesel from FOME, JOME and Diesel fuel are shown in Table 1. The composition of fatty acids in JOME and FOME is given in Table 2.

Table 1
Fuel properties of biodiesel from FOME, JOME and diesel fuel

Fuel Property	Unit	Source	Diesel	Limits as per IS 15607-2005 ASTM D6751	JOME	FOME
Density at 15°C	kg/m^3	Measured	830	860-900	882	890
Kinematic Viscosity at 40°C	cSt	Measured	3.52	1.9-6.0	4.5	5.2
Flash Point	°C	Measured	54	120 min	160	157
Calculated Cetane Index	-	Measured	50	-	54	52.5
Calorific Value	MJ/kg	Measured	43.5	-	39.64	38.65
Element O	wt. %	Given by Supplier	-	-	10.8	8.1

Table 2
Comparison of fatty acid composition (wt %) of fish oil methyl esters and jatropha oil methyl esters

Types of Fatty Acids	Chemical Structure	Type	JOME*	FOME*
Myristic Acid	C14:0	S	0.70	4.98
Palmitic Acid	C16:0	S	15.30	19.42
Palmitoleic Acid	C16:1	US	-	6.43
Heptadecanoic Acid	C17:0	S	-	1.74
Stearic Acid	C18:0	S	9.60	3.80
Oleic Acid	C18:1	US	40.60	20.22
Linoleic Acid	C18:2	US	33.40	3.20
Linolenic Acid	C18:3	US	0.30	1.20
Arachidic Acid	C20:0	S	-	3.56
Eicosadienoic Acid	C20:2	US	-	0.45
Eicosatetraenoic Acid	C20:4	US	-	2.20
Eicosapentaenoic Acid	C20:5	US	-	7.80
Behenic Acid	C22:0	S	-	1.25
Docosapentaenoic Acid	C22:5	US	-	3.25
Docosahexa-Enoic Acid	C22:6	US	-	18.25
Saturated Fatty Acids (S)	C14-C18		25.60	33.37
Unsaturated Fatty Acids (US)	C18:1,2,3		74.30	24.62
Long Carbon-chain Fatty Acid	C20-C22		-	36.76

* provided by the supplier

Fatty acid composition of FOME used in the present work is comparable with that of FOME used by other researchers Wisniewski et al., 2010; (Lin and Li, 2009; Preto et al., 2008). Table 2 shows that fatty acids in FOME are composed of long-chain hydrocarbons ranging from C:20 to C:22. Such long-chain hydrocarbons are generally not found (or found only in traces) in biodiesel derived from edible and non-edible oils. There are no long-chain hydrocarbons ranging from C:20 to C:22 in JOME.

It is also observed that 25.6% of JOME and 33.37% of FOME comprises saturated fatty acid methyl esters. Another key point is the presence of high content of poly-unsaturated fatty acid in FOME with more than three double bonds which are not present in JOME. The presence of high amount of unsaturated fatty acid in FOME leads to low oxidation stability. Therefore, after preparation, FOME must be utilised as quickly as possible to avoid precipitation in fuel injector and fuel injection pump. Present experimental work is confined to analysing performance, emission, and combustion characteristics with Jatropha Oil Methyl Esters blend with diesel in a compression ignition direct injection (CIDI) engine. The effects on performance, emission and combustion are analysed in depth. The tests are conducted on a single cylinder diesel engine coupled with an electrical dynamometer. Standard smoke meter and the gas analyser is used to measure HC, CO, and NO_x values.

RESULTS AND DISCUSSION

An optimum percentage of methyl esters in the blend is necessary as simultaneous reduction of soot and NO_x is desirable. The variation of soot and NO_x values normalised with respect to baseline diesel operation at rated power output for various percentages of JOME and FOME in the blends shows that 20% blends are observed to be optimum considering both NO_x and soot emissions (Reitz et al., 2015; Bhaskar et al., 2017). The brake thermal efficiency of 20% JOME and 20% FOME is marginally lower than that of diesel (Elsanusi et al., 2017). It can be observed that CO, HC and soot emissions are lower while NO_x emissions are higher for 20% methyl ester blends compared with diesel (Sassykova et al., 2017). From the previous experimental results and from literature survey, it is concluded that 20% JOME and 20% FOME can be successfully used in existing diesel engines without any modifications (Dias et al., 2013). 20% JOME and 20% FOME are observed to be optimum for CI engines from the results of investigations carried out but they exhibit higher NO_x emissions compared with diesel fuel. Simultaneous reduction of soot and NO_x is a prime requirement for modern day diesel engines to meet the increasingly stringent emission standards. Three important factors which lead to the formation of NO_x in diesel engines are high temperature, availability of oxygen and residence time for the reaction to complete. As observed from the literature, recirculated exhaust gas is the most effective method to reduce the NO_x emissions (Labecki & Ganippa, 2012), as it reduces the in-cylinder temperatures and availability of oxygen. The recirculated exhaust gas reduces the oxygen concentration of the intake mixture (Verhelst et al., 2009). It also lowers the peak combustion temperature and reduces the NO_x emission. In the present work, part of the exhaust gases from the engine exhaust is cooled down to 30°C and admitted along with the intake air in the manifold. Cooled EGR was used throughout the experimental investigation. At high percentages of EGR, high levels of UBHC, CO and soot emissions are

observed in the exhaust. The EGR rate is calculated based on carbon dioxide (CO_2) in the intake charge and exhaust gas and the ratio is limited to 0.3. The results are compared with those of diesel without EGR.

Experiments are conducted with 10%, 20% and 30% EGR for 20% blends of JOME and FOME. The optimum EGR rate is decided considering the variation of NO_x and soot emissions at various EGR rates for both the methyl ester blends. Figure 1 and 2 show the trade-off between NO_x and soot emissions with various percentages of EGR for 20% JOME and 20% FOME at rated power output. Since the units of NO_x and soot density are different, normalised values of NO_x and smoke are indicated. It can be observed that approximately 20% EGR gives the optimum NO_x and soot emission for both 20% JOME and 20% FOME. Hence, further results are presented only for 20% EGR for both the blends of JOME and FOME.
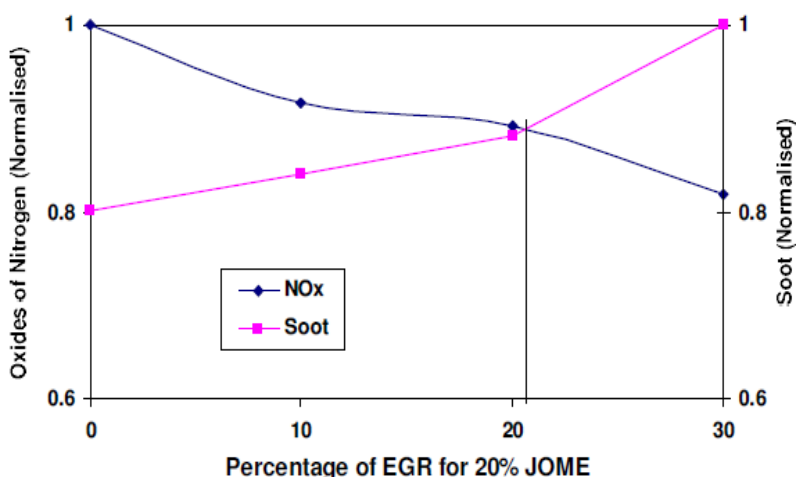

Figure 1. Trade-off between oxides of nitrogen and soot emissions for 20% JOME for various EGR percentages
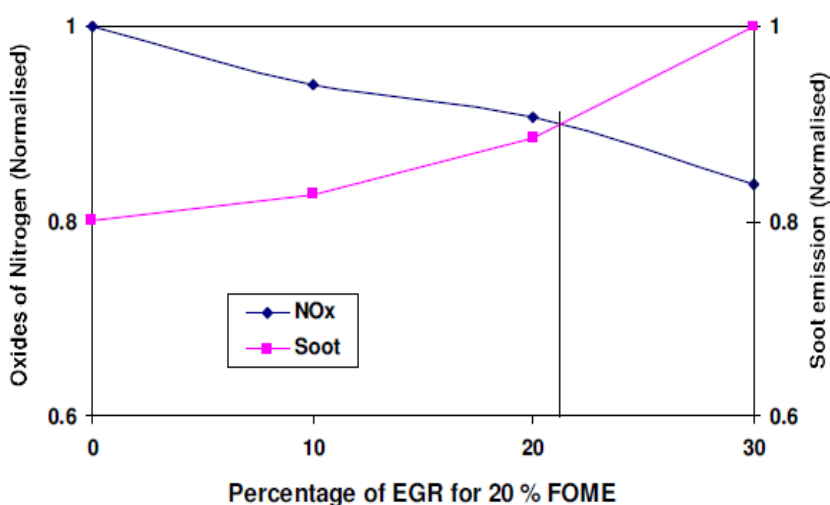

Figure 2. Trade-off between oxides of nitrogen and soot emissions for 20% FOME for various EGR percentages

Heat Release Rate

Figure 3 and Figure 4 show the variation of heat release rate with crank angle at rated power for 20% methyl esters with and without EGR. With EGR, the oxygen available for combustion is reduced which retards the start of combustion and decreases the peak heat release rate. This may lower the peak combustion temperature and decrease NO_x emission with EGR. The figure shows that at rated power output, the peak heat release rate decreases from 66.5 J/°CA for 20% JOME without EGR to 58.7 J/°CA with 20% EGR while it is 77.5 J/°CA in diesel. It is observed that the heat release rate after TDC is higher when the exhaust gas is re-circulated. In the case of 20% FOME, the peak heat release decreases from 70.5 J/°CA for 20% FOME without EGR to 65.8 J/°CA with 20% EGR at rated power output.
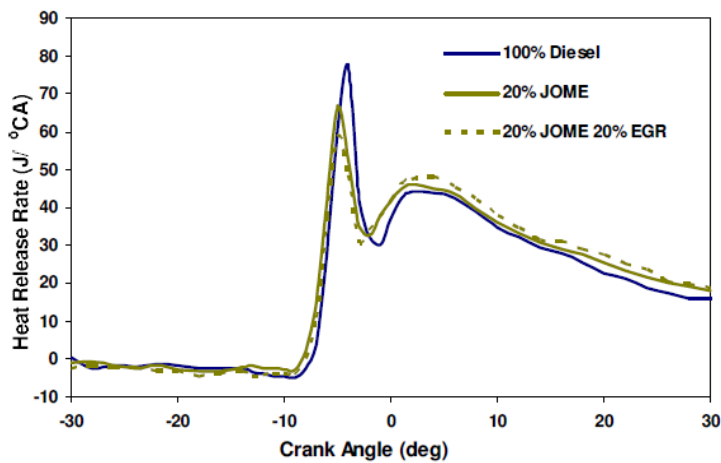

Figure 3. Variation of heat release rate with crank angle at rated power output for JOME with and without 20% EGR
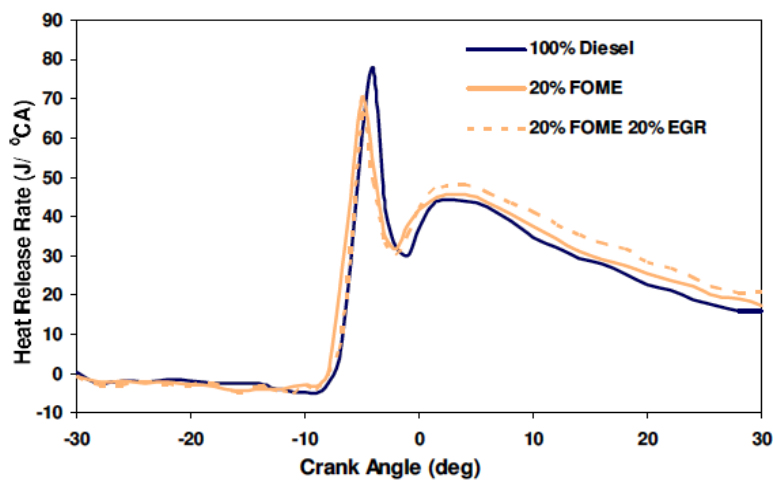

Figure 4. Variation of heat release rate with crank angle at rated power output for FOME with and without 20% EGR

Pressure-Crank Angle Diagram

Figure 5 and Figure 6 show the variation of in-cylinder pressure with crank angle at rated power output for 20% methyl esters respectively with 20% EGR and without EGR. With EGR, the oxygen availability is reduced, and the start of combustion is delayed. It is observed that with EGR, the in-cylinder pressure is marginally low before the occurrence of peak pressure compared to that without EGR for both the methyl esters. The exhaust gas reused increases the heat capacity of the charge in the cylinder and reduces pressure and temperature. It is also observed that in-cylinder pressure with EGR is higher over the entire range of operation.
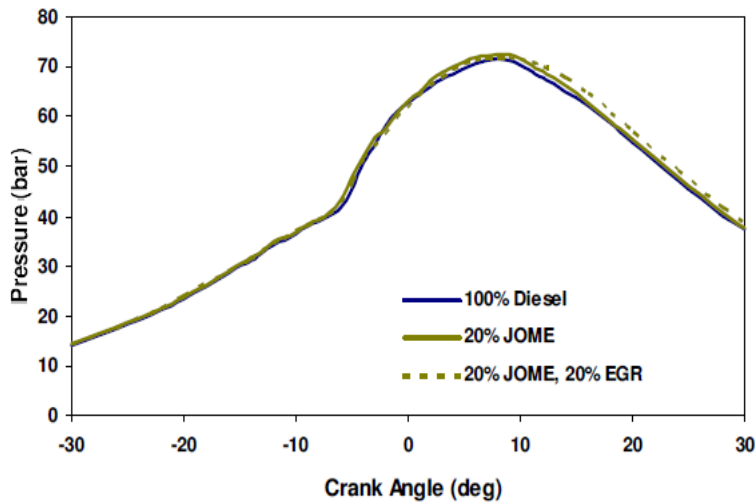

Figure 5. Variation of In-cylinder pressure with crank angle at rated power output for JOME and diesel blend with and without 20% EGR
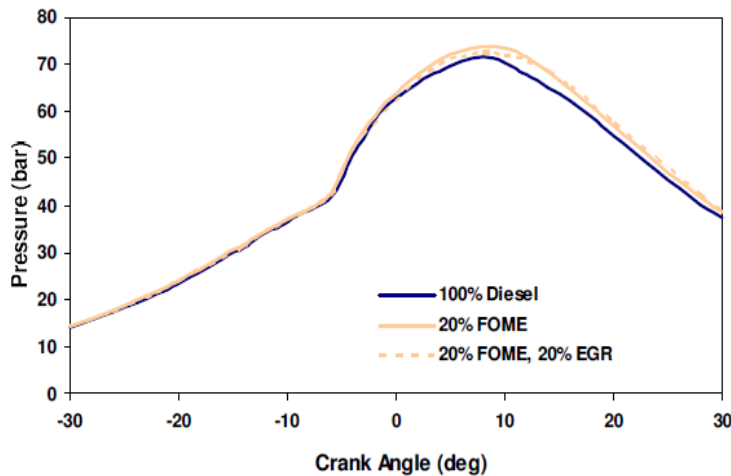

Figure 6. Variation of In-cylinder pressure with crank angle at rated power output for FOME and diesel blend with and without 20% EGR

The variation of performance and combustion characteristics of 20% methyl esters compared with diesel without EGR and with 10, 20 and 30% EGR at rated power output is shown in Figures 7 and Figure 8. As EGR increases, the brake thermal efficiency and exhaust gas temperature decrease. The effect is more pronounced at higher EGR rates. It is observed that ignition delay increases, and peak pressure decrease with increase in EGR for 20% methyl esters. The figure shows the variation of emission of 20% methyl esters with and without EGR compared with diesel at rated power output. NO_x emissions are higher for 20% JOME and 20% FOME blend without EGR. As methyl esters are oxygenated fuels, they have higher combustion temperature and NO_x emission. Experiments were conducted with 10, 20 and 30% EGR and the investigations show that with EGR, NO_x emission is reduced but UBHC, CO and soot emissions increase. For both the methyl esters up to 20% EGR, UBHC and CO emissions are lower compared with diesel. Further increase in EGR increases the UBHC and CO emissions
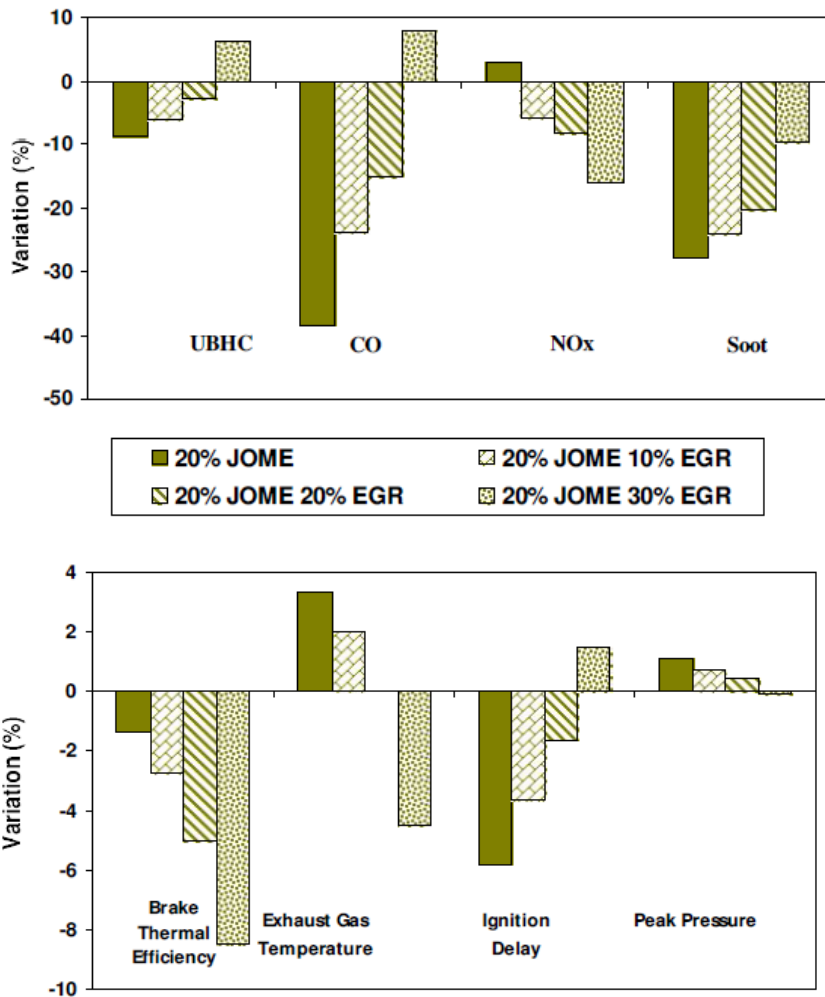

Figure 7. Variation of performance and combustion characteristics of 20% JOME compared to diesel without EGR and with 10,20 and 30% EGR at rated power output

significantly. Thus, it can be concluded that 20% EGR is optimum considering the emissions from both the methyl esters. The following are the summary of investigations carried out to study the effect of EGR with 20% JOME and 20% FOME blends in CIDI combustion mode:
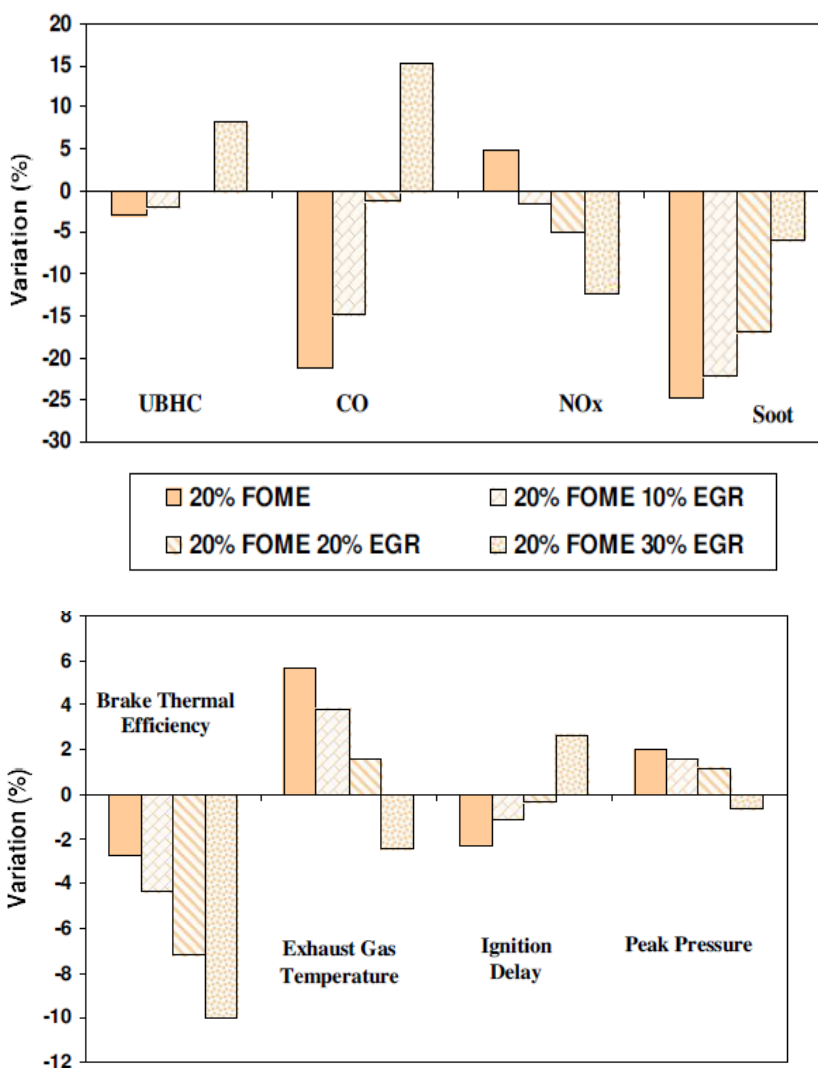

Figure 8. Variation of performance and combustion characteristics of 20% FOME compared to diesel without EGR and with 10,20 and 30% EGR at rated power output

Brake Thermal Efficiency

At rated power output, brake thermal efficiency decreased from 28.0% to 26.0% for 20% JOME when EGR flow rate varied between 0 and 30% while the variation was from 27.7% to 25.6% for 20% FOME compared with 28.4% for diesel.

Exhaust Gas Temperature

At rated power output, the exhaust gas temperature without EGR is 460°C for 20% JOME and 470°C for 20% FOME. With 20% EGR, the exhaust gas temperature is 445°C for 20% JOME and 450°C for 20% FOME whereas in the case of diesel, exhaust gas temperature is 445°C. The exhaust gas temperature in the case of 20% JOME with 20% EGR is almost equal to that of diesel. The percentage variation in exhaust gas temperature is zero hence, it is not shown in the figure.

Unburnt Hydrocarbon Emissions

At rated power output, the UBHC emission varied between 0.7 and 0.8 g/kWh for 20% JOME when EGR is varied between 0 and 30% while the variation is from 0.7 to 0.9 g/kWh for 20% FOME compared with 0.7 g/kWh for diesel. 20% FOME with EGR shows higher UBHC emissions at all the power outputs compared with 20% JOME with EGR due to its lower percentage of oxygen and higher percentage of longer chain fatty acid components present in it.

Carbon Monoxide Emissions

CO emissions at rated power output for various EGR flow rates varied from 10.3 to 23.0 g/kWh for 20% JOME while they vary from 13.2 to 25.5 g/kWh for 20% FOME compared to 16.8 g/kWh for diesel. Higher CO in the case of FOME may be due to lower intrinsic oxygen available for combustion compared with JOME.

Oxides of Nitrogen Emissions

At rated power output, the NO_x emissions without EGR and with 10,20 and 30% EGR vary from 7.5 to 6.1 g/kWh for JOME while it varied between 7.6 and 6.4 g/kWh for 20 FOME compared with 7.3 g/kWh for diesel. Higher heat capacity of the mixture requires more energy, and lowers the flame temperature which reduces the NO_x emissions at rated power output.

Soot Emissions

At rated power output, the soot emissions with no EGR, 10, 20 and 30% EGR varied between 120 and 150 mg/m^3 for 20% JOME while the soot emissions varied between 125 and 156 mg/m^3 for 20% FOME compared with 166 mg/m^3 for diesel.

Ignition Delay and Peak Pressure

At rated power output, the ignition delay period increases as the percentage of EGR increases. Peak pressure is marginally lower with EGR for both the methyl esters at rated power output compared to without EGR.

The UBHC, CO and soot emissions increase as the percentage of EGR increases. With EGR, NO_x emissions are significantly lowered for both the methyl esters, Increasing the percentage of EGR decreases NO_x emissions significantly. Considering both NO_x and soot emissions, 20% EGR is observed to be optimum for both 20% JOME and 20% FOME.

CONCLUSION

Based on the results obtained from the above, the following conclusion may be drawn.

The brake thermal efficiency decreases marginally with 20% methyl esters with EGR. As the percentage of EGR increases, the brake thermal efficiency decreases further. The exhaust gas temperature decreases marginally with 20% methyl esters with EGR. The exhaust gas temperature with 20% JOME and 20% EGR equals to that of Diesel. The ignition delay increases as the percentage of EGR increases compared to that without EGR in diesel. Peak pressure is marginally lower with EGR for both the methyl esters at rated power output compared with without EGR. The UBHC, CO and Soot emissions increase as the percentage of EGR increases. With EGR, NO_x emissions are significantly lowered for both the methyl ester. Increasing the percentage of EGR decreases NO_x emissions significantly. Considering both NO_x and soot emissions, 20% EGR is observed to be optimum for both 20% JOME and 20% FOME.

RECOMMENDATIONS

The cost per hour of operation at rated power output for Diesel, 20% JOME with 20% EGR and 20% FOME with 20% EGR are US$1.30, US$1.60 and US$1.50 respectively. For 20% FOME, the cost is marginally lower than that of 20% JOME. Use of 20% FOME is recommended for use in coastal areas where it is easily available and 20% JOME is recommended for rural areas where it can be cultivated.

ACKNOWLEDGMENT

REFERENCES

Abd-Alla, G. H. (2002). Using exhaust gas recirculation in internal combustion engines: A review. *Energy Conversion and Management, 43*(8), 1027–1042.

Agarwal, D., Sinha, S., & Agarwal, A. K. (2006). Experimental investigation of control of NO_x emissions in biodiesel-fueled compression ignition engine. *Renewable Energy, 31*(14), 2356–2369.

Axelsson, L., Franzén, M., Ostwald, M., Berndes, G., Lakshmi, G., & Ravindranath, N. H. (2012). Perspective: Jatropha cultivation in southern India: Assessing farmers' experiences. *Biofuels, Bioproducts and Biorefining, 6*(3), 246–256.

Bai, Y. L., Wang, Z., & Wang, J. X. (2010). Part-load characteristics of direct injection spark ignition engine using exhaust gas trap. *Applied Energy, 87*(8), 2640–2646.

Bhaskar, K., Sendilvelan, S., Nagarajan, G., & Sampath, S. (2017). Emission characteristics of biodiesel obtained from jatropha seeds and fish wastes in a diesel engine. *Sustainable Environment Research, 27*(6), 283-290.

Das, L. M., & Mathur, R. (1993). Exhaust gas recirculation for No_x control in a multicylinder hydrogen-supplemented S.I. engine. *International Journal of Hydrogen Energy, 18*(12), 1013–1018.

Demirbas, A. (2008). Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. *Energy Conversion and Management, 49*(8), 2106–2116.

Dias, J. M., Araújo, J. M., Costa, J. F., Alvim-Ferraz, M. C. M., & Almeida, M. F. (2013). Biodiesel production from raw castor oil. *Energy, 53*, 58-66.

El-Kasaby, M., & Nemit-Allah, M. A. (2013). Experimental investigations of ignition delay period and performance of a diesel engine operated with Jatropha oil biodiesel. *Alexandria Engineering Journal, 52*(2), 141–149.

Elsanusi, O. A., Mohon R. M., & Sidhu, M. S. (2017). Experimental investigation on a diesel engine fueled by diesel-biodiesel blends and their emulsions at various engine operating conditions. *Applied Energy, 203*, 582-593.

Fang, T., Lin, Y. C., Foong, T. M., & Lee, C. Fon. (2009). Biodiesel combustion in an optical HSDI diesel engine under low load premixed combustion conditions. *Fuel, 88*(11), 2154–2162.

Galloni, E., Fontana, G., & Staccone, S. (2014). Numerical and experimental characterization of knock occurrence in a turbo-charged spark-ignition engine. *Energy Conversion and Management, 85*, 417–424.

Heffel, J. W. (2003). NO_x emission reduction in a hydrogen fueled internal combustion engine at 3000 rpm using exhaust gas recirculation. *International Journal of Hydrogen Energy, 28*(11), 1285–1292.

Hountalas, D. T., Mavropoulos, G. C., & Binder, K. B. (2008). Effect of exhaust gas recirculation (EGR) temperature for various EGR rates on heavy duty DI diesel engine performance and emissions. *Energy, 33*(2), 272–283.

Labecki, L., & Ganippa, L. C. (2012). Effects of injection parameters and EGR on combustion and emission characteristics of rapeseed oil and its blends in diesel engines. *Fuel, 98*, 15–28.

Ladommatos, N., Abdelbalim, S., & Zhao, H. (2000). The effects of exhaust gas recirculation on diesel combustion and emissions. *International Journal of Engine Research, 1*(1), 107-126.

Leung, D. Y. C., Wu, X., & Leung, M. K. H. (2010). A review on biodiesel production using catalyzed transesterification. *Applied Energy, 87*(4), 1083-1095.

Lin, C. Y., & Li, R. J. (2009). Engine performance and emission characteristics of marine fish-oil biodiesel produced from the discarded parts of marine fish. *Fuel Processing Technology, 90*(7–8), 883–888.

Murali Manohar, R., Prabhahar, M., & Sendilvelan, S. (2012). Experimental investigation of combustion and emission characteristics of engine is fueled with diesel and UVOME blends of B20K and B80K. *European Journal of Scientific Research, 76*(3), 327-334.

Payri, F., Bermdez, V. R., Tormos, B., & Linares, W. G. (2009). Hydrocarbon emissions speciation in diesel and biodiesel exhausts. *Atmospheric Environment, 43*(6), 1273–1279.

Pradhan, P., Raheman, H., & Padhee, D. (2014). Combustion and performance of a diesel engine with preheated *Jatropha curcas* oil using waste heat from exhaust gas. *Fuel, 115*, 527–533.

Preto, F., Zhang, F., & Wang, J. (2008). A study on using fish oil as an alternative fuel for conventional combustors. *Fuel, 87*(10–11), 2258–2268.

Raj, C. S., Arul, S., Sendilvelan, S., & Saravanan, C. G. (2010). A comparative assessment on performance and emissions characteristics of a diesel engine fumigating with methanol, methyl ethyl ketone, and liquefied petroleum gas. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 32*(17), 1603-1613.

Reitz, R. D., & Duraisamy, G. (2015). Review of high efficiency and clean reactivity controlled compression ignition (RCCI) combustion in internal combustion engines. *Progress in Energy and Combustion Science, 46*, 12-71.

Ruijun, Z., Xibin, W., Haiyan, M., Zuohua, H., Jing, G., & Deming, J. (2009). Performance and emission characteristics of diesel engines fueled with diesel-dimethoxymethane (DMM) blends. *Energy and Fuels, 23*(1), 286–293.

Sassykova, L., Gil'mundinov, Sh., Nalibayeva, A., & Bogdanova, I. (2017). Catalytic systems on metal block carriers for neutralization of exhaust gases of motor transport. *Revue Roumaine de Chimie, 62*(2), 107-114.

Talibi, M., Hellier, P., & Ladommatos, N. (2017). The effect of varying EGR and intake air boost on hydrogen-diesel co-combustion in CI engines. *International Journal of Hydrogen Energy, 42*(9), 6369-6383.

Verhelst, S., Maesschalck, P., Rombaut, N., & Sierens, R. (2009). Increasing the power output of hydrogen internal combustion engines by means of supercharging and exhaust gas recirculation. *International Journal of Hydrogen Energy, 34*(10), 4406-4412.

Wisniewski, A., Wiggers, V. R., Simionatto, E. L., Meier, H. F., Barros, A. A. C., & Madureira, L. A. S. (2010). Biofuels from waste fish oil pyrolysis: Chemical Composition. *Fuel, 89*(3), 563–568.

Benders Decomposition for Distribution Networks with Cross-docking Centre

Manpreet Singh, Divya Aggarwal* and Vijay Kumar

Computer Science and Engineering Department, Thapar University, Patiala, India

ABSTRACT

This paper proposes the Benders Decomposition approach for modelling distribution networks with cross-docking centres. The cross-docking centre eliminates the requirement for inventory stores. The mathematical formulation of the proposed model is also presented, and the latter consists of plants, cross-docking centres and distribution centres. The Bender Decomposition approach is utilised to solve the proposed model which is tested on 15 different characteristics of test instances. The effect of plants, and cross-docking centres are also investigated. The experimental results reveal the proposed formulation provides promising solutions with reasonable computation time.

Keywords: Benders Decomposition, Cross-docking, Dual Sub-Problem, Master Problem, Mixed Integer Programming

INTRODUCTION

There is an increasing need to develop an effective supply chain that provides better service to customers' demand and minimise the investment cost. Therefore, the selection of facility location is an important factor

in the supply chain. The cost of opening and setting a new facility can greatly affect profits. Therefore, an efficient warehousing strategy is required to reduce the distribution cost. Hence, the cross-docking concept is introduced in the distribution network. The main concept behind cross-docking is the delivery of goods from suppliers to customers through intermediate cross-docking centres where the materials are not stored for long time, generally less than 24 hours.

Cross-docking system facilitates the rapid movement of materials by eliminating storage. It improves organisations' response time and reduces transportation cost. Therefore, cross-docking has become a great distribution strategy as it can speed up the flow of goods

which would eventually lower the transportation cost. If locations of cross-dock centres are chosen carefully, then the transportation cost can be reduced to a great extent. The Cross-Dock technique can eventually reduce the distribution costs while satisfying customer demand (Kellar et al., 2015; Mousavi et al., 2014).

The main contribution of this paper is to incorporate the cross-docking concept in the supply chain. A novel model is proposed that utilise the concept of cross-docking. The mathematical formulation of proposed model has been introduced. The Bender Decomposition approach has been used to optimise transportation and fixed costs. The performance of proposed model has been tested on 15 different test instances. The results have been compared with IBM CPLEX.

The reminder of this paper is organised as follows. Section 2 presents the related work done in the field of facility location problem while Section 3 describes the proposed model. The proposed solution methodology is described in Section 4 and Section 5 presents the experimental setup, results and discussions. Finally, the paper is concluded in Section 6.

RELATED WORK

Over the last few decades, the issue of facility location had gained attention of researchers (Melo et al., 2009; Erengüç et al., 1999). The main focus of these researches are on determining which site location should be selected for establishing new facilities from the available set of potential sites while satisfying the constraints. Hindi and Basta (1994) proposed a structure of problem inspired by Geoffrion and Graves (1974) which is a branch and bound approach as solution. Uster et al. (2007) studied a distribution system design problem that has a fixed number of capacitated facilities and suggested various metaheuristics to solve this problem. Cintron et al. (2010) proposed a multi-criterion problem. They solved the problem using Mixed Integer Programming (MIP). Sun and Wang (2015) proposed strategic distribution network design problem of bulk materials considering different distribution modes before and after an intermediate unloading in the distribution centres. They solved mixed- integer model by using the Benders Decomposition (BD).

Cross-docking (CD) is a relatively new strategy in logistics (Van Belle et al., 2012). The various well-known companies, such as Toyota, UPS, Wal-Mart, Kodack Co, Good Year, etc. have successfully implemented Cross-docking in their distribution system. Ross and Jayaraman (2008) suggested a solution approach with the combination of Simulated Annealing and Tabu Search to solve linear programming model of supply chain. The supply chain includes potential retailers, cross-docking centres, and regional warehouses. Dondo and Cerda (2012) considered pickup and delivery problem in a cross-docking system known as Vehicle Routing Problem with Cross-docking (VRPCD). They worked on a monolithic formulation which determined the truck scheduling at the cross-docking centre and routes. It involved a solution strategy based on sweep heuristic which can solve big problems within reasonable CPU time.

Arabani et al. (2010) suggested a cross-docking scheduling problem in which product delivery should have a pre-determined time schedule. Moreover, penalties are considered for any late delivery. Three solution approaches namely Differential Evolution (DE), Genetic Algorithm (GA), and Particle Swarm Optimisation (PSO) were developed to solve this problem. Lee et al. (2006) considered a cross docking system in supply chain. They developed

a mathematical model for vehicles. Tabu search algorithm was used to optimise the developed mathematical model.

The above-mentioned researchers missed BD approach for modelling the distribution network with cross-docking centres tracking flow of materials. The concept of cross-docking centres is utilised in this paper and a model is proposed that involves single-sourcing constraints which ensures that each distribution centre is exclusively served by a single cross-docking or merge-in-transit centre. Here, single-period, single-product, and multi-echelon are considered in a deterministic situation to understand the fundamental concepts of cross-dock planning problem. This approach increases the utilisation of warehouse and prevents the construction of links and location of the warehouse with low utilisation.

Moreover, when the problem is considered with large instances, it does not provide optimal solutions. Hence, the BD approach is used to solve mathematical formulation of the given problem. The problem is decomposed into two sub-problems. Figure 1 shows the two-stage classical distribution network where supply is made between the distribution centres and facility location, and customer locations and distribution centre. Figure 2 describes the distribution network with intermediate Cross-dock.
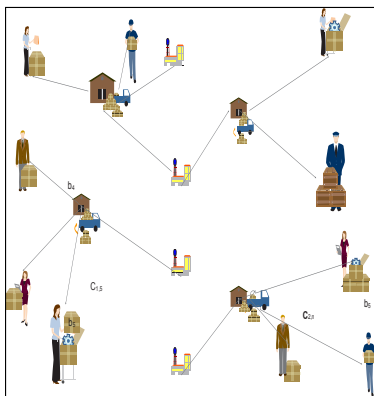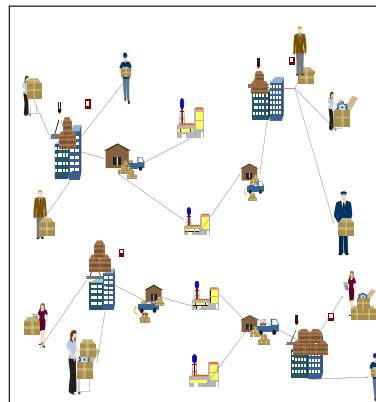

Figure 1. Two-stage distribution network

Figure 2. Cross-dock distribution network

MATHEMATICAL FORMULATION

Figure 3 shows the three-stage graph network that considers facility location problem with cross-docking centres. It consists of plants, warehouses and distribution centres which are represented by nodes and relationship between these nodes are represented by links. The model considers two binary variables for opening the cross-docking and allocating to distribution centres. The flow of products from manufacturing plants to operating cross-docks centres are considered in this model.

The following notations, parameters and decisions variables have been used in mathematical formulation of mixed-integer programming model.
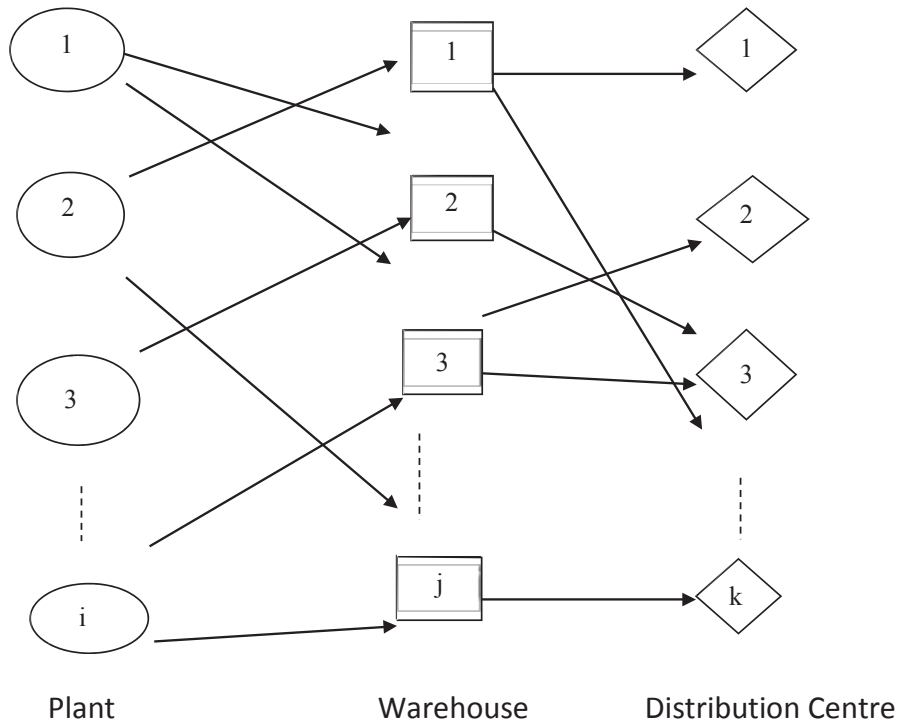
Figure 3. Distribution network incorporate Cross-docking centres

Notations

1. I : set of manufacturing Plants, indexed by i, $\forall i \in I$

2. J : set of Cross-docking warehouses, indexed by j, $\forall j \in J$

3. K : set of Distribution Centres, indexed by k, $\forall k \in K$

Constants

1. C_i : Capacity of Plant i.

2. W_j : Capacity of Cross-docking warehouse j.

3. F_j : Fixed opening cost of cross-docking warehouse in location w.

4. T_{ij} : Transportation cost per unit product from plant i to cross-docking warehouse j.

5. S_{ik} : Shipping cost of product from Cross-dock Warehouse j to distribution centre k.

6. R_k : Request for distribution centre k.

Binary Decision variables

$$X_{jk} = \begin{cases} 1, & \text{if warehouse } j \text{ fulfill demand of distribution center } k \\ 0, & \text{otherwise} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{if cross-dock warehouse is opend at location } j \\ 0, & \text{otherwise} \end{cases}$$

Q_{ij} : Amount of product sent from plant i to the cross-dock j

The mathematical formulation of the problem is as follows:

$$\min Z = \sum_{i \in I} \sum_{j \in J} T_{ij} * Q_{ij} + \sum_{j \in J} F_j * Y_j + \sum_{j \in J} \sum_{k \in K} S_{jk} * R_k * X_{jk} \tag{1}$$

subject to

$$\sum_{j \in J} Q_{ij} < C_i, \forall i \in I \tag{2}$$

$$\sum_{k \in K} R_k * X_{jk} = \sum_{j \in J} Q_{ij}, \forall j \in J \tag{3}$$

$$\sum_{j \in J} X_{jk} = 1, \forall k \in K \tag{4}$$

$$\sum_{k \in K} r_k * X_{jk} = W_j * Y_j, \forall j \in J \tag{5}$$

$$P * Y_j \le \sum_{k \in K} X_{jk}, \forall j \in J \tag{6}$$

$$P = \min\{R_k\}, \forall k \in K \tag{7}$$

$$Q_{ij} \ge 0, \forall j \in J, \forall i \in I \tag{8}$$

$$Y_j \in \{0,1\}, \forall j \in J \tag{9}$$

$$X_{ij} \in \{0,1\}, \forall i \in I, \forall j \in J \tag{10}$$

The objective function mentioned in Eq. (1) consists of three terms. The first term is the cost of transportation from manufacturing plant i to Cross-dock Warehouse j. The second term describes the fixed cost to open cross-dock warehouse j. The third term includes the amount for achieving demand raised by distribution centre k. The constraint (2) implies that the output of plant i does not violate the capacity of plant i. The constraint (3) confirms the quantity

of arrival products to be same as sent from plant i. The constraint (4) ensures the demand satisfaction of each distribution centre k by a single Cross-docking warehouse j. The number of products that can be sent to a distribution centre k from an open cross-docking warehouse j is confirmed by constraint (5). The constraint (6) guarantees that at least the minimum amount of demand is received by cross-dock warehouse. The constraint (7) ensures that the minimum demand of each distribution centre k is considered. The constraints (8), (9), and (10) ensure the non-negative and integrity conditions.

PROPOSED SOLUTION APPROACH

The BD approach is used to solve the mathematical model mentioned above. Figure 4 shows the flowchart of BD which consists of following steps:

- Compute the Sub-Problem (SP) from given problem by fixing integer variables.
- Compute Dual Sub-Problem (DSP) which provides lower-bound (LB) of the problem.
- Compute Master Problem (MP) which provides upper bound (UB) of the problem.
- Compute LB and UB repeatedly until $UB - LB < \varepsilon$. Here, ε is a very small constant.

Steps of Algorithm

1. Initialise the parameters of BD approach. The lower bound (LB) and upper bound (UB) are set to $-\infty$ $+\infty$ respectively. The values for X_{ij} and Y_j are initialised.

2. Compute DSP which gives LB.

    $Max_u \{(\beta^t \bar{y} + \Omega - b \bar{y})^T u \mid \alpha^T u \leq C, u \geq 0\}$

3. *If* LB is unbounded

    Get unbounded Ray $\bar{u}$.

    Add cut $(\Omega - by)\ \bar{u} \leq 0$ to MP

    Else

    Get extreme point $\bar{u}$

    Add cut $Z \geq \beta^T y + (\Omega - by)^T \bar{u}$ to MP

    Set UB to Min $\{UB, \beta^T \bar{y} + (\Omega - b \bar{y})^T \bar{u}\}$

    End If

4. Solve MP which gives upper bound.

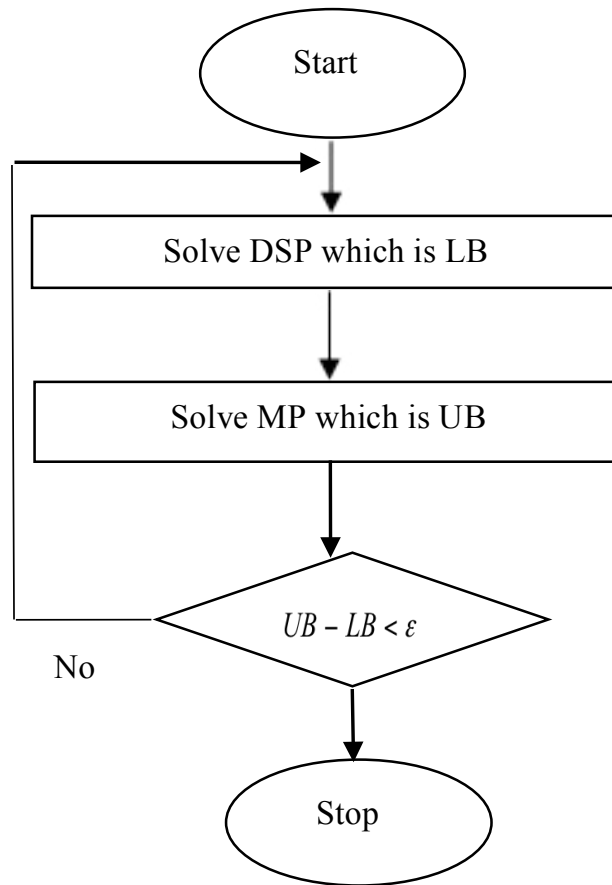5. Repeat Steps 2-6 until $UB - LB < \varepsilon$

Figure 4. Flow chart of Benders Decomposition

Solving the BD Problem

The given SP is decomposed into two sub-problems such as DSP and MP. The mathematical representation of the sub-problem is given below:

$$\min Z = \sum_{i \in I} \sum_{j \in J} T_{ij} * Q_{ij} + \sum_{j \in J} F_j * \overline{Y}_j + \sum_{j \in J} \sum_{k \in K} S_{jk} * R_k * \overline{X}_{jk} \tag{11}$$

subject to

$$- \sum_{j \in J} Q_{ij} \geq -C_i, \forall i \in I \tag{12}$$

$$\sum_{i \in I} Q_{ij} = \sum_{k \in K} R_k * \overline{X}_{jk}, \forall j \in J \tag{13}$$

$$\sum_{i \in I} Q_{ij} \geq P * \overline{Y}_i , \forall i \in I \tag{14}$$

For DSP, three new variables, namely U1, U2, U3 are introduced. The value of these variables is taken as positive. The value of these variables is fixed for MP. The mathematical representation of DSP is given below:

$$\max Z = - \sum_{i \in I} C_i U1_i + \sum_{j \in J} \sum_{k \in K} R_k \overline{X_{jk}} U2_{jk} + \sum_{j \in J} P \overline{Y_j} U3_j \tag{15}$$

subject to

$$-U1_i + U2_{jk} + U3_j \leq T_{ij}, \qquad\qquad \forall i \in I, \forall j \in J \tag{16}$$

The mathematical representation of MP is given below:

$$Z \geq \sum_{j \in J} F_j * Y_j + \sum_{j \in J} \sum_{k \in K} S_{jk} * R_k * X_{jk} + - \sum_{i \in I} C_i * \overline{U1}_i + \sum_{j \in J} \sum_{k \in K} R_k * \overline{X_{jk}} \, \overline{U2_{jk}} + \sum_{j \in J} P * \overline{Y_j} * \overline{U3_j} \tag{17}$$

subject to

$$- \sum_{i \in I} C_i * U1_i + \sum_{j \in J} \sum_{k \in K} R_k * \overline{X}_{jk} U2_{jk} + \sum_{j \in J} P * \overline{Y}_j * U3_j \leq 0 \tag{18}$$

$$X_{jk} \in \{0,1\} \tag{19}$$

$$Y_j \in \{0,1\} \tag{20}$$

Computation of Dual Sub-Problem (DSP)

DSP gives lower bound for fixed values of X_{jk} and Y_j. It is computed using the following procedure:

Step 1. Initialise the values of decision variables X_{jk} and Y_j.

Step 2. *If* problem is feasible

      Add new optimality cut

      Else

      Add new feasibility cut

      End If

Step 3. Repeat Step 2 until termination condition is not satisfied.

Computation of Master Problem (MP)

MP is used to obtain feasible and optimal solutions. It depends on the results generated by DSP. After solving DSP, the feasible values for X_{jk} and Y_j are obtained which may not be optimal solution for the original problem. It is used to generate feasible and optimal solution which provides an upper bound of given problem. If the new value of *UB* is better than the previous one, then the value of *UB* is updated. This procedure is repeated until the bound gap is reached. If both MP and DSP are not feasible, then the algorithm is terminated.

Termination Conditions

The algorithm will terminate if any of the following conditions are satisfied:

1. Bound Gap: It is reached when difference of UB and LB gets close to zero; i.e., *(UB−LB)* < *ε*, *ε* is very small constant value. Typically, pre-specified value of *ε* is 0.00001.

2. Iteration limit: When the iteration limit has reached a pre-determined value *maxItrLimt;* i.e., If *n>maxItrLimt*, then processing will terminate. *MaxItrLimt* is typically set to 1000.

3. Step size Limit: When the step size α becomes Negligible (approximately zero); i.e., If *α<minStepSize*, then processing will terminate. *MinStepSize* is typically set to 0.0001.

4. Infeasible Solution: If both DSP and MP give infeasible solution, then no results will be generated for this problem.

PERFORMANCE ANALYSIS

This section validates the performance of the proposed model over 15 test instances and compares it with IBM CPLEX solver.

IBM CPLEX Solver

The IBM CPLEX is one of the tools widely used to solve combinatorial optimisation problems. It has a concert technology that provides interfaces to C++, C# and Java languages. It is accessible through independent modelling systems, such as AMPL, and TOMLAB. It is also recognised as a constraint solving toolkit suitable for solving optimisation models. It uses inbuilt procedures to solve the mixed integer programming in short time. It can be used to solve a variety of different optimisation problems in a variety of computing environments. The IBM CPLEX is an exact solver that uses mixed integer programming to search the desired solutions.

Test Dataset

The three types of datasets are randomly generated in Table 1. Dataset *X* has large opening cost of cross-dock warehouse, while dataset *Y* has moderate opening cost of cross-dock warehouse with large transportation cost from plant to warehouse and from warehouse to distribution centres. Therefore, the shipping is costlier than opening a new warehouse in *Y* dataset. The instances generated in dataset *Z* are based on small to medium sized organisations. It has moderate transportation and opening cost of cross-dock warehouse. The datasets for problem are randomly generated by keeping realistic characteristics for small and medium sized organisation. Table 1 represents the range of possible values for cost, demand, supply and capacity in each dataset. Here, *U* indicates uniform distribution of numbers over the specified range. The average value of 10 independent simulation runs are reported. Table 2 shows the size and characteristics of test instances.

Table 1
Detail of dataset used

Datasets	X (Type 1)	Y (Type 2)	Z (Type 3)
Cross-Dock Warehouse Opening Cost (F_j)	U [1500, 4000]	U [500,1250]	U [400,2000]
Transportation cost from Plant to i warehouse j (T_{ij})	U [50, 110]	U [80, 130]	U [75, 130]
Shipping Cost from Cross-Dock warehouse to Distribution centres (S_{jk})	U [50, 110]	U [80, 120]	U [50, 120]
Demand (R_k)	U [10,25]	U [10, 25]	U [20, 45]
Plant's Capacity (M_i)	U [15, 45]	U [25, 40]	U [20, 50]
Warehouse Capacity (N_j)	U [10,30]	U [25, 60]	U [25,75]

Table 2
Characteristics of test instances

Instance	Plants (I)	Warehouse (W)	Distribution Centre (D)	Constraints	Continuous Variables	Binary Variables	Iterations
1	11	17	25	106	385	910	2755
2	15	25	30	115	525	1085	16448
3	28	43	71	185	1204	3096	2253815
4	26	41	53	161	1066	2214	511439
5	17	27	35	138	731	1548	17283
6	22	37	45	141	814	1702	3374
7	37	42	65	186	1554	2772	92978
8	19	63	91	236	1197	5796	2593380
9	22	36	45	139	792	1656	924374
10	21	34	39	128	714	1360	6231
11	32	61	95	249	1952	5856	2753898
12	47	79	127	332	3713	10112	1179933
13	23	61	91	236	1403	5612	1497516
14	21	35	47	138	735	1680	1773725
15	29	43	49	164	1247	2150	51670

Experimental Results

The test instances are checked by decomposing the problem into DSP and MP which give the lower bound and upper bound to the optimal solution respectively.

The optimality gap is calculated to measure performance and is defined as follows:

$$Optimality\ Gap = \frac{UB - LB}{UB} \times 100 \qquad (21)$$

Table 3 depicts optimality gap, computation time and mean objective for the various test instances. It is observed from table that the proposed model provides better computational time than CPLEX Solver which solves the problem by computing mixed integer programming. The

problem instances are solved in lesser time than CPLEX Solver. The optimality gap obtained from the proposed model is smaller than 0.25. This means that the solution produced by the proposed model is close to an optimal one. The computational test on the random instances with size up to 50 has maximum gap of 0.2%. This shows that the proposed approach is quite promising.

Table 3
Performance evaluation on test instances

CPLEX Solver					
Instance	Plants (I)	Warehouse (W)	Distribution Centre (D)	Gap%	CPU Time (in sec)	Mean Objective
1	11	17	25	0.00	1.89	6.849E+3
2	15	25	30	0.00	2.14	8.999E+3
3	28	43	71	0.20	34.13	2.110E+4
4	26	41	53	0.11	24.09	1.510E+4
5	17	27	35	0.12	1.84	8.199E+3
6	22	37	45	0.21	2.05	2.128E+4
7	37	42	65	0.28	10.30	2.153E+4
8	19	63	91	0.32	173.12	2.908E+4
9	22	36	45	0.09	44.71	1.899E+4
10	21	34	39	0.10	1.09	1.265E+4
11	32	61	95	0.20	171.85	3.600E+5
12	47	79	127	0.21	165.20	4.848E+5
13	23	61	91	0.23	126.47	2.935E+4
14	21	35	47	0.02	78.81	2.975E+5
15	29	43	49	0.04	4.11	1.954E+5
Benders Decomposition (Proposed Model)					
Instance	Plants (I)	Warehouse (W)	Distribution Centre (D)	Gap%	CPU Time (in sec)	Mean Objective
1	11	17	25	0.00	1.11	7.438E+03
2	15	25	30	0.00	1.53	1.012E+03
3	28	43	71	0.12	31.34	2.101E+04
4	26	41	53	0.08	20.87	1.635E+04
5	17	27	35	0.00	1.25	1.039E+03
6	22	37	45	0.07	1.34	1.502E+04
7	37	42	65	0.13	7.50	1.963E+04
8	19	63	91	0.15	138.39	2.649E+04
9	22	36	45	0.05	37.57	1.241E+04
10	21	34	39	0.07	0.87	1.182E+04
11	32	61	95	0.15	151.32	2.756E+04
12	47	79	127	0.20	153.87	3.688E+05
13	23	61	91	0.19	122.59	2.641E+04
14	21	35	47	0.00	61.54	1.415E+05
15	29	43	49	0.00	3.21	1.588E+05

Analysis of Parameters Involved

Figure 5 shows the impact of number of plants, cross-docking and distribution centres on the objective function. The horizontal axis gives the approximate objective value when multiplied by 10^3. The bigger the number of plants, warehouses and distribution centres, the greater the objective function. This figure also shows the value of plants, warehouses and distribution centres affect the value of objective function.
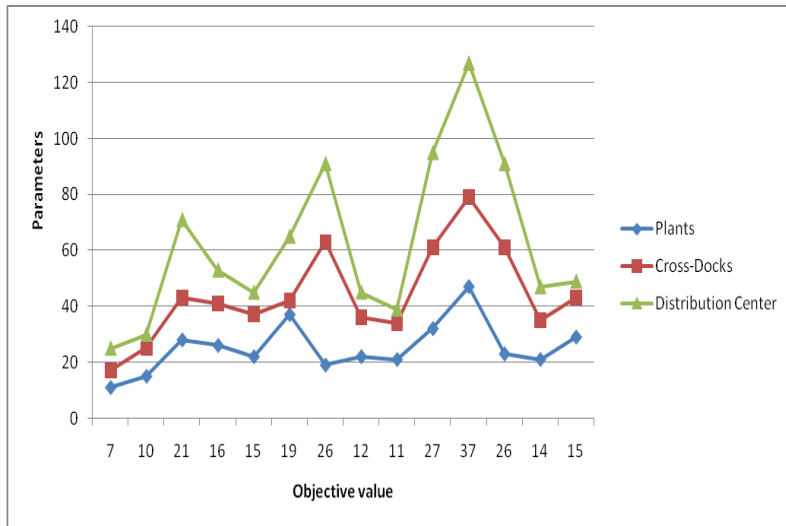

Figure 5. Impact of number of parameters involved in proposed model

CONCLUSION

In this paper, the concept of cross-docking in distribution network was proposed. The model utilised the concept of cross-docking. The mathematical formulation of proposed model was established. The Benders Decomposition was utilised to optimise the proposed model. The proposed model was tested on 15 test instances. The experimental result reveal that the proposed model is found to be superior to the CPLEX solver in terms of computational time, objective function and optimal gap. For future studies, the proposed model can be used in a parallel environment.

ACKNOWLEDGMENTS

REFERENCES

Boloori Arabani, A. R., FatemiGhomi, S. M., & Zandieh, M. (2010). A multi-criteria Cross-docking scheduling with just-in-time approach. *The International Journal of Advanced Manufacturing Technology, 49*(5), 741-756.

Cintron, A., Ravindran, A. R., & Ventura, J. A. (2010). Multi-criteria mathematical model for designing the distribution network of a consumer goods company. *Computers and Industrial Engineering, 58*(4), 584–593.

Dondo, R., & Cerda, J. (2012). A Sweep- Heuristic based formulation for the Vehicle Routing Problem with Cross-docking. *Computers and Chemical Engineering, 48*, 293-311.

Erengüç, S. S., Simpson, N. C., & Vakharia, A. J. (1999). Integrated production-distribution planning in supply chains: An invited review. *European Journal of Operational Research, 115*(2), 219–236.

Geoffrion, A. M., & Graves, G. W. (1974). Multicommodity distribution system design by benders decomposition. *Management Science, 20*(5) 822–844.

Hindi, K. S., & Basta, T. (1994). Computationally efficient solution of a multiproduct, two-stage distribution location problem. *Journal of the Operational Research Society, 45*(11), 1316–1323.

Kellar, G. M., Polak, G. G., & Zhang, X. (2015). Synchronization, Cross-docking, and decoupling in supply chain networks. *International Journal of Production Research, 54*(9), 2585–2599.

Klose, A. (2000). A Lagrangean relax-and-cut approach for the two-stage capacitated facility location problem. *European Journal of Operational Research, 126*(2), 408–421.

Lee, Y. H., Jung, J. W., & Lee, K. M. (2006). Vehicle Routing Scheduling for Cross-docking in the supply chain. *Computers and Industrial Engineering, 51*(2), 247-256.

Melo, M. T., Nickel, S., & Saldanha-da-Gama, F. (2009). Facility location and supply chain management—A review. *European Journal of Operational Research, 196*(2), 401–412.

Mousavi, S. M., Vahdani, B., Tavakkoli-Moghaddam, R., & Hashemi, H. (2014). Location of Cross-docking centres and vehicle routing scheduling under uncertainty: A fuzzy possibilistic-stochastic programming model. *Applied Mathematical Modelling, 38*(7), 2249–2264.

Ross, A., & Jayaraman, V. (2008). An evaluation of new heuristics for the location of Cross-docks distribution centers in supply chain network design. *Computers and Industrial Engineering, 55*(1), 64-79.

Sun, D., & Wang, G. (2016, May). Benders approach for a bulk material distribution network design problem. In *Information Technology, Networking, Electronic and Automation Control Conference, IEEE* (pp. 256-259). IEEE.

Üster, H., Easwaran, G., Akçali, E., & Çetinkaya, S. (2007). Benders Decomposition with alternative multiple cuts for a multi-product closed-loop supply chain network design model. *Naval Research Logistics, 54*(8), 890–907.

Van Belle, J., Valckenaers, P., & Cattrysse, D. (2012). Cross-docking: State of the art. *Omega, 40*(6), 827-846.

Distance Correlation between Plaintext and Hash Data by Genetic Algorithm

Farjami, Y.*, Rahbari, D. and Hosseini, E.

Department of Computer Engineering and Information Technology, University of Qom, Iran

ABSTRACT

The hash function is used as a one-way cryptography method for digital signature and message authentication. Hash values are provided using a mathematical and logical process, so they are different from the generators of random numbers. The position analysis of bits in plaintext and its hash is very suitable to show their relationship. The focus of this paper is to point to the best relations between the plaintext and hash bits, in which the difference between hash methods will be proven. In this work, we use distance correlation (dCorr) as a measurement function of precision statistical dependency between two vectors. The genetic algorithm (GA) is used to find a set of optimal positions between plaintext and its hash data with maximum dCorr. The results of the experiment regarding dCorr indicate that MD5 as a message digest method is different from the random function. Also, the proposed method compared with Tabu search (TS) and Simulated Annealing (SA) algorithms has a lower average execution time for 1000 pairs of plaintext and hash data.

Keywords: Distance correlation, genetic algorithm, hash, message digest

INTRODUCTION

The advancement of science and technology has put security and privacy issues in network communications in the spotlight (Zhang et al., 2014; Zhang & Xiao, 2013). Messages are always subject to attack and access when they are transmitted on the network. Message authentication can be used to resist the attacks (Reshef et al., 2011). Hash is a one-way method for producing fixed-length values (Schneier, 2004) that is used for security and data availability in communication systems. Strong hash functions do not produce the same output values for different inputs, but some cases point to the strength algorithms that challenge the hash functions (Kwok, 2005). Other applications of hashing are approximate matching in forensics system (Zhai, 2017;

Choi et al, 2017; Choi et al 2016). In these systems, user assertion is recognised by comparing two hash files. The file format can be text, audio, and video. If an attacker can find two different files with the same hash, then the security of hashing method is compromised.

The dCorr is a recent method for finding the correlation between data with any size of dimension (Kwok, 2005). The Pearson correlation coefficient (PCC) generally cannot detect nonlinear relations and its value often is zero for dependent variables, thus a new measure is introduced called dCorr coefficient to address the shortcomings of the Pearson coefficient (Kwok, 2005). Moreover, unlike the Pearson coefficient, dCorr coefficient value is zero if and only if the variables are independent (Székely & Rizzo, 2009).

In this paper, we show best relationships between plaintext and its hash data with maximum dCorr by GA; the latter is a metaheuristic method with the natural selection process that is used to find optimal solutions (Mitchell, 1998). Based on dCorr, the correlation between plaintext and its hash data is calculated in certain positions, also GA finds best positions with maximum dCorr. The key contributions of this paper are as follows: 1) Prove the existence of a logical process for calculating MD5-based hash unlike the random function; 2) Find the maximum dCorr value and best relationship positions between plaintext and its hash data; and 3) Comparison of our proposed algorithm with TS and SA algorithms is based on average execution time.

The rest of the paper is organised as follows. The dCorr function calculates the relationship between plaintext and its hash by MD5 and random methods. The details of the GA are presented for a quick and optimal calculation of dCorr. Findings are discussed in the results and discussion section especially comparing the proposed algorithm with TS and SA. The final section concludes the paper.

LITERATURE REVIEW OF RELATED WORK

In Lipowski and Lipowska (2012), the authors proposed a cyber black box system for analysis of MD5 (Rivest, 1992) and SHA256 hashing methods on the network packets. They recognised transmission packets in real time and optimised their method by logging and storing of network traffic and related information generation. Traffic analysis in mobile networks as an excellent source of knowledge is a popular research field (Naboulsi et al., 2016; Martínez-Gómez et al., 2014). Detecting the relationship between data and breaking the hash is one of the important steps to find encrypted information. The related works are categorised as Maximal Information Coefficient (MIC), Distance correlation (dCorr), and other correlation measures.

Maximal Information Coefficient (MIC)

Reshef et al. (2011) used MIC to compare two random variables which is based on the mutual information with an easy implementation and strong against noise (Székely & Rizzo, 2013). Linge et al. (2014) showed how MIC applies in the particular case of the side channel attacks. They offered ways to compute the MIC with low complexity. The results show high performance of this approach when the leakage is noisy and poorly modelled (Linge, 2014).

In Gorgens et al. (2017), the airborne laser scanning was introduced for computation of height and density metrics. The MIC will be used in this study to find a relationship between

forests attributes. In Wang et al. (2018), an intelligent MIC (iMIC) is offered for quadratic optimisation to create a finer normal matrix. The authors used 169 data set indices from 202 member countries of World Health Organization (WHO) to discover the local optimal value by iMIC with lower iterations and fine performance. Since the material on the y-axis is a branch of quadratic optimisation with the same partition on the x-axis, iMIC method is a better alternative than the proposed searching method by Reshef et al. (2011) In (Shao & Li., 2016), authors created the graph model with railway nodes that shows influencing factors of accidents by edges between each pairs of nodes. They proposed MIC method to discover the relationship among railway nodes to prevent accidents.

An MIC-based approach was proposed in Sun et al. (2017) where entropy epistasis was used to measure the level of factor interactions in a constant optimisation. Shuliang et al. offered an improved estimate algorithm for MIC, named the IAMIC. Their method has the ability to search for extreme values using few iterations. Reshef et al. (2014) provided a comparison between the improved algorithm and the original MIC algorithm. The MIC has lower power than dCorr-based methods (Wang et al., 2017; Székely & Rizzo, 2009) in many relationships, and it is claimed that this fault could cause MIC to lose its advantage for general use.

Distance Correlation (dCorr)

In Wang and Zhonghua (2017), dCorr is used for outlier recognition in high-dimensional regression problems and leave-one-out method. The researchers used the bootstrap method and a threshold rule to determine whether an individual is an outlier. The Monte Carlo simulations with a real data show the offered outlier recognition method can detect outlier survey and enhance the filtering accuracy in attribute hide problem.

In Górecki et al. (2017), the dCorr and canonical correlation coefficient with Fourier series are used to determine the relationship between daily temperatures for each day of the year and average daily rainfall data in Canada. The results show the values of dCorr coefficients are independent of size but the value of canonical correlation coefficient increases a little with the size of the regularisation. As a result, these methods are not suitable for univariate data. As a second experiment, they used multivariate data obtained from Professor R. Nadulski of Department of Food Machinery and Engineering. They used permutation tests that included p-values equal to zero for both methods of infrigidation (the process of cooling or freezing for preservative purposes) and all basis sizes. The authors suggested dCorr to find correlation relations.

Wang et al. (2016) studied the intrusion detection and variable screening in large-scale regression. The dCorr was used to discover and analysis correlation of actual data with high dimensional. They found the correlation between responses and predicted data with dCorr and pointed that the proposed method can reduce unfavourable effects after removing powerful data tips. The filtering accuracy can be refined in variable screening problem. In (Pan et al., 2017), the authors established conditional local dCorr coefficient to determine a nonlinear relationship between multivariate data. In Wang et al. (2017), the dCorr method and G2 statistic are used for Bayesian modelling approach, so the authors analysed and compared the dependence and

independence of nonparametric data. In Heller et al. (2016), the data space is divided into 2 * 2 tables to compare the dependencies and independence of the parts. Their strategy is similar to Fisher's exact scanning (FES) method based on the dCorr approach for 2 * 2 spaces.

The dCorr coefficient which has more information compared with Pearson coefficient (Martínez-Gómez, 2014) is used to show nonlinear associations by variables with any dimension, and it's possible to determine mini sets of variables that give same astrophysical information. For experiments, all values are given for Pearson's correlation coefficient. As a result, dCorr method is stronger than the MIC (De Siqueira Santos et al., 2013) to detect astrophysical information into most concentrated horseshoe or V-shaped, and also to increase the power of classification and pattern recognition (Linge, 2014).

Other Correlation Measures

In Budhathoki and Vreeken (2017), the Minimum Description Length (MDL) method programmed in the Python environment is used to discover related variables in data analysis. In this work, the Kolmogorov complexity is applied to compact and find a correlation. The MDL model shows data correlation with high significant power, especially for binary data but this method can't easily be extended to other data types. In another work, Exceptional Model Mining (EMM) was used to find a linear relation between subgroups of the dataset whose attributes have an unusual relationship. In Downar and Duivesteijn (2017), experiments were conducted on several datasets as Windsor Housing, the South African Heart Disease Study, the Ozone, Iris, and the Higgs Boson ML challenge. The results of correlation quality methods show that the Filler–Spearman and the Filler–Kendall rank overlap with Pearson's correlation. The authors proposed the dCorr and MIC approach to find a strong relationship between data and subgroups of data.

PROPOSED METHOD

Distance Correlation

The dCorr takes values in [0, 1] and is equals to zero if and only if independence holds. It is more general than the classical Pearson product moment correlation. It provides a scalar measure of multivariate independence that typifies independence of two random vectors. It is well defined in both the population and in the sample and it is easy to compute (Geerligs & Henson, 2016; Székely & Rizzo, 2009). Given random variables X, Y $\in$ R let X'Ý and X'', Y'' be independent pairs of random variables taken from the same joint distribution as that of X and Y. The distance covariance of X and Y is defined as the square root of Eq. (1).

$$dCov^2(X,Y) = E[|X - x'||Y - Y'|] + E[|X - X'|]E[|Y - Y'|] - E[|X - X'||Y - Y''|] - E[|X - X''||Y - Y'|] = E[|X - X'||Y - Y'|] + E[|X - X'|]E[|Y - Y'|] - 2E[|X - X'||Y - Y''|] \quad (1)$$

The distance variance and dCorr are defined as Eqs. (2) and (3).

$$dVar^2(X) = dCov^2(X, X) \tag{2}$$

$$dCor^2(X, Y) = \frac{dCov^2(X,Y)}{\sqrt{dVar^2(X)} \sqrt{dVar^2(Y)}} \tag{3}$$

Algorithm Structure

The calculation of dCorr between two vectors is simple when their positions are given. The purpose of this research is to find the best positions of the plaintext and its hash data with maximum dCorr value, which requires a lot of computing.
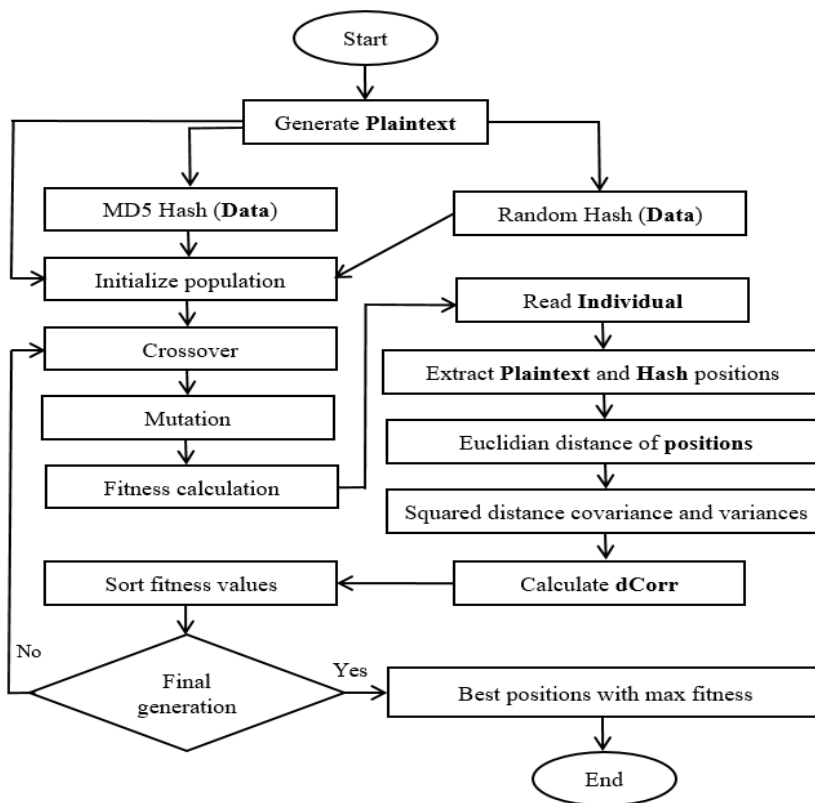

Figure 1. Flowchart of proposed algorithm

The GA is based on the idea of evolution and it can find best solutions after running several generations. GA searches the answer space with a directed random method. Flowchart of our proposed algorithm is shown in Figure 1. We generate a plaintext (P) then calculate its hash value (H) by MD5 and random methods. The pair of (P, H) sends to GA and maximum dCorr value and best bit positions of P and H are calculated. Details of the proposed method are as Algorithm 1.

Algorithm 1. **Proposed Algorithm**
1:     **For** dc = 2 to 32 **do**
2:         **For** t = 1 to 1000 **do**
3:             Generate plaintext as P.
4:             H1 = MD5 (P).
5:             H2 = RandomHash (P).
6:             [dCorr1, BestPositions1] = **GA** (P, H1, dc).
7:             [dCorr2, BestPositions2] = **GA** (P, H2, dc).
8:             Print best positions of P, H1, and H2 with Maximum dCorr value.
9:         **End For**
10:    **End For**

Our proposed method as Algorithm 1 is executed by different numbers of dCorr positions as dc in line 1. We also use this parameter for analysis of outputs. For example, to calculate the dCorr value between 2 positions of plaintext and its hash data, dc is equal to 2, also for 8 positions, dc is equal to 8. The algorithm is run by 1000 pairs of plaintext and their hash in line 2. In Line 3, a random plaintext as P with 32 characters is generated. The MD5 and random hash values of P are calculated as H1 and H2 in lines 4 and 5. The GA is called twice to find the maximum dCorr value and best positions of bits between P and H1 (MD5) in line 6, also between P and H2 (Random) in line 7. The best answers are printed in line 8. The steps of the GA as Algorithm 2 include population initialisation, crossover, mutation, selection, fitness evaluation in a certain number of iterations. These steps are explained below:

Population

The structure of a chromosome or an individual is based on a binary string, so that it has fast speed in GA operations and converts to the decimal bit positions. The string length is based on the number of dCorr positions. For example, 2 positions in plaintext and 2 positions in its hash require 32 bits in the chromosome. Because each position is a decimal value or a binary string of 8 bits. Thus, by multiplying the number 4 by 8, the string length is 32 bits, so an individual has 32 bits or genes. Additionally, for 3 positions of plaintext and its hash data, an individual has 48 bits, and so on. In the model, the first half of each individual is considered for the plaintext and the second half belongs to its hash data.
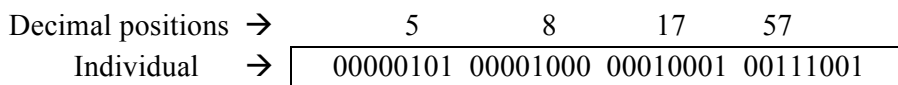
Decimal positions →   5   8   17   57

Individual →   00000101 00001000 00010001 00111001

Figure 2. One chromosome or an individual of population

We compute the main position of bits in the plaintext or hash by dCorr-based fitness function. As shown in Figure 2, by separating every 8 bits in the chromosome and converting them to a decimal value, positions 5 and 8 in the plaintext, as well as 17 and 57 in its hash data are obtained.

Crossover

In GA, the crossover operator is important for convergence towards the solutions. In Figure 3, a single-point crossover is used so that the second half of the two parent chromosomes are exchanged. The high rate of crossover operator increases the convergence speed. The parent selection stage of crossover operator is performed using roulette wheel method (Lipowski & Lipowska, 2012).

Before	Parent 1	00000101000001000 0001000100111001
	Parent 2	1001000010010001 1101010011111010
After	Child 1	00000101000001000 1101010011111010
	Child 2	1001000010010001 0001000100111001

Figure 3. The parents and childs before and after crossover

In roulette wheel, all individuals have a fitness value titled as p_i, which is divided by the sum of the fitness values of population, then a random value as r is generated for each individual. If p_i value is greater than r, then the chromosome is selected as the parent for the crossover operation. If the fitness value of two child chromosomes is better than rest individuals in the population, they will be selected to produce the next generation.

Mutation

As Figure 4, the mutation operator changes a bit of population to include plaintext and its hash positions from 0 to 1 or vice versa. The diversity rate of the population is proportional to the probability of the mutation operator rate. In fact, if we assume that the population size is 10, with a mutation rate of 30%, then the number of selected chromosomes for the mutation operation is equal to 3. After mutation operations, the plaintext or its hash positions are changed and then added to the population. If the fitness of the mutated chromosome is better than rest individuals, then it can be selected for the next generation of the GA.

Before	00000101000010000001000100111001
After	00000101000010000001001100111001

Figure 4. A chromosome before and after the mutation operator

Fitness Function

The fitness function is based on dCorr as Eq. (3). First, the plaintext and its hash data are read as an individual. Second, the plaintext and its hash positions are extracted as X and Y vectors. Third, dCorr (X, Y) is calculated. Since the purpose of the GA is to find the maximum dCorr value, thus the fitness values are in descending order. The GA is presented as Algorithm 2.

Algorithm2. **GA**

Input: (P = plaintext, H = Hash, dc = dCorr positions).

Output: dCorr value and best positions.

1:     Initialise Population_size, Max_Iterations, Crossover, and Mutation probability.

2:     L = **dc * 8 * 2**.

3:     **For** t =1 to Population_size **do**

4:         Generate L bits for each individual.

5:     **End For**

6:     **For** i =1 to Max_Iterations **do**

7:         Popc = Crossover two chromosomes with roulette wheel selection.

8:         Popm = Mutation population.

9:         Replace Popc and Popm by worst individuals.

10:     **Fitness calculation (dCorr):**

11:         Read an individual.

12:         [Plaintext_pos, Hash_pos] = Extract positions of P and H.

13:         Calculate $X = P(Plainttex\_pos)$ and $Y = H(Hash\_pos)$.

14:         Calculate the Euclidian distance for P and H.

15:         Calculate squared distance covariance and variances.

16:         Calculate dCorr as Eq. (3).

17:     Sort Fitness values.

18:     **End For**

19:     Return dCorr and best positions.

In Algorithm 2, the initial values of the parameters are presented in line 1 by the number of population individuals, the rate of mutation, and crossover operators. In line 2, the length of each individual is calculated as L. In the calculation of L, **dc** is the number of positions, **8** is the number of bits for a character, and **2** is for two inputs as plaintext and its hash. For example, we want to calculate the dCorr value for 4 positions of P and 4 positions of H, thus dc is equal 4. As Algorithm 1 (line2), the plaintext is 32 characters and each character is 8 bits then 32 * 8 = 256 bits or positions, thus, we need to **8** bits for addressing to 256 positions. Also, because of the two different sets of positions for P and H, we multiply (**dc*8**) to **2**. The main circle of GA is executed from line 6 to line 18. The crossover and mutation operators are presented in lines 7 and 8. After these operations, the worst individuals in the population are replaced by new child individuals in line 9. The fitness function as dCorr is calculated by lines 10 to 16. An individual include P and its H positions is read in line 11 and is extracted in line 12.

Based on the above example, we proof the decimal values of an individual are (5, 8, 12, 23, 2, 9, 16, and 28). Thus, values (5, 8, 12, and 23) are for P positions and values (2, 9, 16, and 28) are for H positions. Algorithm 2 in Line 12, gain the corresponding bit values in P and H. In line 13, two vectors are gained for dCorr calculation. The lines 14 to 16 calculate dCorr using the Euclidian distance, squared distance covariance, and variances as Eq. (3). In line 17, the fitness values of all individuals are descending. The GA is finished in line 18. In line 19, the maximum fitness value (dCorr) and best relation positions between P and H are returned to Algorithm 1.

RESULTS AND DISCUSSION

The proposed algorithm is implemented by Matlab software on a PC with CPU of Intel Core i5 2.67 GHz, 3.0 GB of RAM in Windows 10 operating system 32 bits. The selected hash functions for result evaluation are MD5 and random methods. At the start of the GA, some parameters are initialised, so that the number of genes of each chromosome is equal dc*16 as proposed algorithm description, the maximum number of iterations is equal 50, the population size is equal 10, the crossover operator probability is equal 0.8, the mutation operator probability is equal 0.3. The dc is a number of positions that is used to compute dCorr value between plaintext and its hash data.

Table 1
A sample input and output of the GA

Plaintext	Hash	P positions	H positions	dc
'Password123*()'	**MD5:**35acc07d9d81171d94afff007b2a69fa	4, 8, 12, 14	15, 23, 26, 27, 32	6
	Random:/Ñ&M_m□ Îa0Ìd)ë¡T² ìúAZáØÍí³	1, 5, 7, 8, 9	4, 7, 12, 13, 18, 27	

The number of dCorr position is equals to 4, so that the GA ultimately achieves 4 best positions. As Figure 5 (A), GA reaches to optimised answer in the fourth generation, thus the proposed method is very suitable for calculating of dCorr. Figure 5 (B) shows the values of dCorr between the plaintext and its hash in different generations of the GA, so that these values are calculated after 14 generations of the GA.
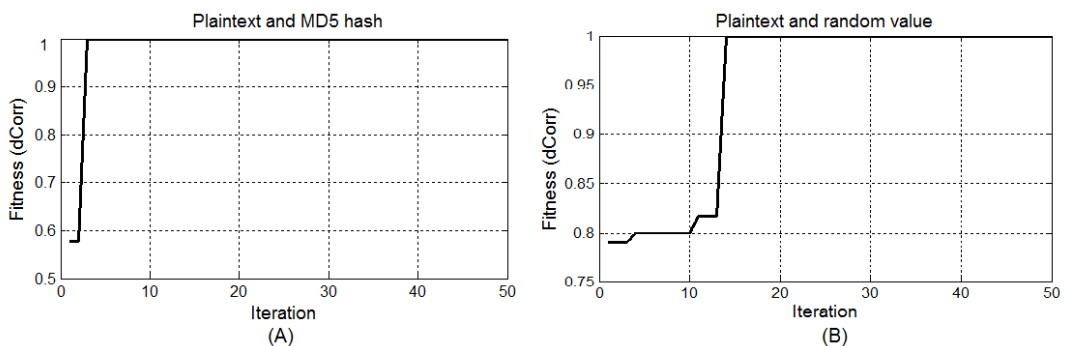

Figure 5. GA output for one plaintext with MD5 Hash in (A) and Random Hash in (B)

According to the results, the convergence rate and dCorr values of MD5 is less than the random function and the proposed algorithm finds the best MD5 dCorr values at the primitive generations. The diagrams in Figure 5 are based on the plaintext (P) and its hash (H) of Table 1. The proposed method is tested by 1000 pairs of plaintext (Each plaintext generates by 32 random characters) and their hash data by MD5 and random methods. This test is based on the different number of dCorr positions, so that their average values are given in Figure 6.

The average dCorr values for 1000 plaintexts and hash data are different in MD5 and random hash methods as seen in Figure 6. The values in horizontal axis represent dCorr value for certain position count of plaintext and its hash data. The MD5 hash function has a known process and it isn't based on random process. As per Figure 3, for positions greater than 3, dCorr values for MD5 are more than the random values. As the first result of this diagram, MD5 hash, unlike the random function, is based on a mathematical and logical process. In second, there are correlations between the plaintext and its MD5 hash data in different positions, thus, attack to MD5 hash function is feasible. If a relationship is found between the plaintext and its hash, then other relationships will be found, which will result in the hash method longer having security.
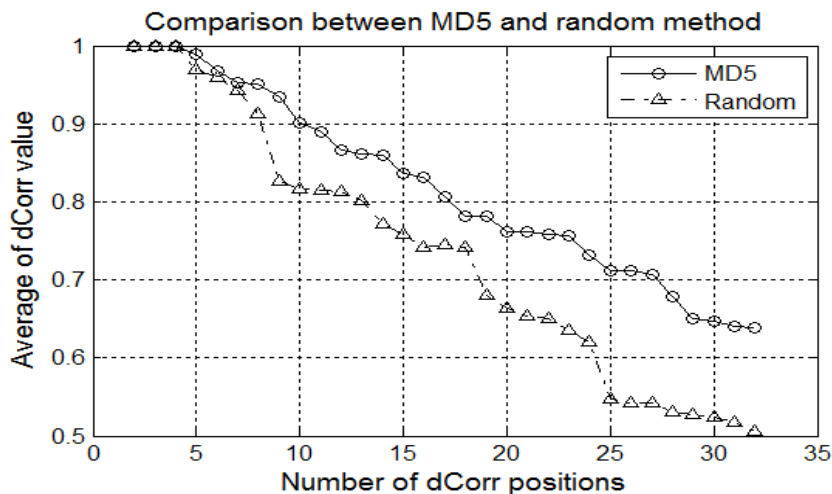
Figure 6. Comparison of dCorr for MD5 and Random Hash

The hash operation should be a one-way function but the MD5 has shown that it does not have enough security and it is not suitable for use. The main purpose of this paper is to calculate dCorr between plaintext and its hash data as well as the proof of difference with the random method. Table 2 shows different plaintext (P) and its hash (H) data as well as their best relationship positions in both MD5 and random methods, which are obtained from the GA.

Table 2
Details of GA output

Plaintext	Hash	dc	P Positions	H Positions	dCorr
'password'	**MD5:** 5f4dcc3b5aa765d61d8327deb882cf99	2	3, 6	9, 19	1
	Random: IþRq($"dxÀ 2□§¨@8Àù½'v	7, 8	3, 30		0.99
'12345678'	**MD5:** 25d55ad283aa400af464c76d713c07ad	3	4, 7, 8	1, 10, 29	1
	Random: AùnÅÀ¶∏□ò¡@/«·\Ò¸Û		2, 4, 8	7, 25, 30	0.98
'20172017'	**MD5:** 29ce822e8b7e51a318e382b0e34f4360	4	3, 4, 8	11, 18, 28, 29	1
	Random: Eý÷9ÈSVØ÷Èÿ{¦në'¡»Èí%o0[°w		1, 2, 5	1, 19, 22, 23	0.97
'pass123'	**MD5:** 32250170a0dca92d53ec9624f336ca24	5	1, 2, 6, 8	5, 12, 21, 29, 30	0.98
	Random: ;)Uæ2Ï¥¡Á4Ý&÷¿7õNBàVvÞ		3, 4, 5, 6	5, 6, 11, 17, 30	0.96
'!@#pass123'	**MD5:** 0dbe23f840b9443a25b9b6b493759925b	6	1, 3, 5, 6, 7, 9	5, 16, 19, 25, 28, 31	0.96
	Random: ÁÜpco¥Ä¸>»□ÁÒCjk{täëÈc;¢þ&o		3, 4, 5, 8, 10	11, 17, 22, 29, 30	0.96
'123$%^789'	**MD5:** 0f246b38c49c50b7460f82cbe15871e9	7	2, 3, 5, 6, 8, 9	4, 9, 11, 13, 22, 24, 27	0.96
	Random: ³-Ò"+¥□ é:½S$3Õpnyß¾8DGÑGb		1, 2, 6, 7, 8	11, 14, 15, 17, 30, 31	0.90
'Password123*()'	**MD5:** 35acc07d9d81171d94afff007b2a69fa	8	2, 4, 8, 9, 13, 14	5, 13, 14, 21, 23, 26, 27, 30	0.90
	Random: ò3s?Ð>w¢Bÿó'òÛ7ÅEÐ©»È{æË		2, 4, 7, 9, 12, 13	3, 7, 18, 21, 24, 27, 30	0.89
'123Pass#$%^'	**MD5:** d91b9fb6a4c0a5c1d2a8c33b7d8cd261	9	1, 3, 4, 6, 7, 9, 11	5, 8, 9, 16, 17, 18, 20	0.85
	Random: Ü0UÜíí'¡fêfà Àð'yk⇐»»5vu·		3, 4, 5, 6, 8, 10, 11	3, 6, 9, 14, 15, 17, 24, 27	0.83
'P1a@s#s$w%o^r&d'	**MD5:** f0ec1af343ebd56e7aa23c14f99fb7b6	10	4, 8, 9, 13, 15, 23	1, 2, 3, 5, 8, 10, 12 14	0.81
	Random: o9ëAm¶∏À´ÏÅÍ?âî°«ßo¤tüëÀ¸5g>ë		4, 9, 10, 11, 12, 14, 15	3, 8, 10, 14, 18, 21, 22, 25	0.81

In some rows of Table 2, the GA obtained less number of positions than dc value, due to dCorr values being optimal. The dc is maximum position counts in each iteration of the algorithm. For example, in Table 2, the row with plaintext equal 'Password123*()' whose dc is 8, in the result of the simulation, the number of positions in plaintext is 6, in hash positions by MD5 is 8, and in the random hash is 7. The difference in the number of positions is due to the maximum dCorr value between plaintext and its hash data that is calculated using the proposed algorithm.
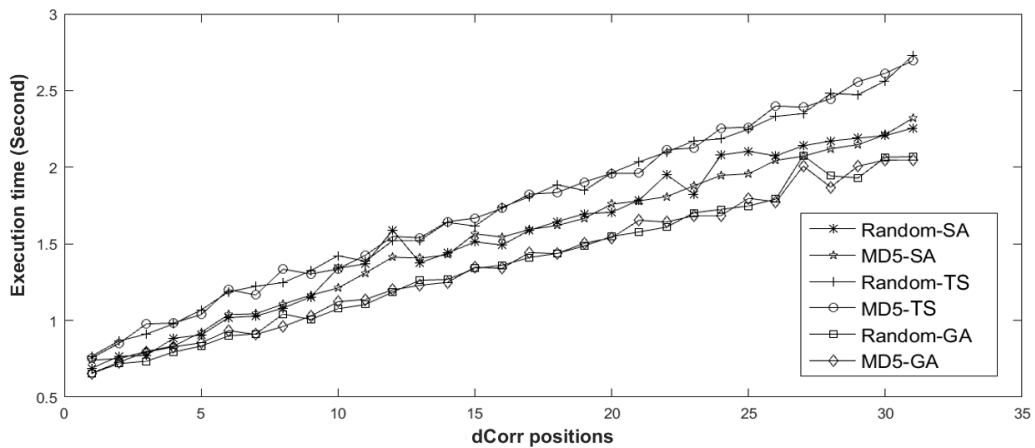
Figure 7. Comparison of Execution Times

Another parameter for the analysis of the proposed method is the execution time. To compare of the proposed algorithm with other methods, TS (Glover & Laguna, 2013) and SA are programmed (Siarry, 2016). The proposed method is first run by GA and the best value of dCorr is calculated before it is run using TS and SA to reach best dCorr value. The plaintexts and their hashes are considered equal for each of three algorithms. Figure 7 shows 32 columns include the execution time (Second) of the proposed algorithm as GA, TS, and SA. In this chart, the horizontal axis as dc is the number of dCorr positions and the vertical axis is execution time of all methods. The executed random hash function has a less time than MD5. In fact, the MD5 method has a longer process than random function. The average of execution time in two methods (MD5, Random) in GA is (1.3596, 1.3647), in TS is (1.7358, 1.72.86), and in SA is (1.5225, 1.5430). As our experiments for three algorithms (GA, TS, and SA), the execution time parameter for GA is minimum and TS has maximum value.

CONCLUSION

In this paper, we prove the existence of a mathematical and logical process for calculating MD5 hash method, which is different with the random function. To do this work, it is very important to find the correlation between the plaintext and its hash in breaking the hash method.

The study has shown different methods are used for correlation measuring between two data vectors. These methods include MIC, dCorr and other correlation measures. The dCorr is

the best. We calculate the maximum dCorr value of the best relationship positions in plaintext and its hash by two hash functions as MD5 and random methods. The GA is used for optimal positions. The proposed method is executed for 1000 pairs of plaintexts and their hashes. The results of dCorr values show that MD5 method is better than the random method for many different positions. The execution time of MD5 is also longer than random function due to its long process. This method is proposed with TS and SA algorithms. As results, the average execution time by GA is 22% less than TS and 11% less than SA.

The results of this paper prove a strong relationship between the plaintext and its hash that is computed by dCorr for MD5 and random method. The hash functions with a high dCorr value between plaintext and its hash are not suitable for the message digest and it is simply broken by the attacker. Future research will focus on discovering the relationship in big data by dCorr measurement.

REFERENCES

Budhathoki, K., & Vreeken, J. (2017, June). Correlation by compression. In *Proceedings of the 2017 SIAM International Conference on Data Mining* (pp. 525-533). Society for Industrial and Applied Mathematics.

Choi, J., Park, J., & Kim, H. (2017). Forensic analysis of the backup database file in KakaoTalk messenger. In *Big Data and Smart Computing (BigComp), IEEE International Conference* (pp. 156-161). IEEE.

Choi, Y., Lee, J. Y., Choi, S., Kim, J. H., & Kim, I. (2016). Introduction to a network forensics system for cyber incidents analysis. In *Advanced Communication Technology (ICACT), 2016 18th International Conference on IEEE* (pp. 50-55). IEEE.

De Siqueira Santos, S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2013). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics, 15*(6), 906-918.

Downar, L., & Duivesteijn, W. (2017). Exceptionally monotone models—The rank correlation model class for Exceptional Model Mining. *Knowledge and Information Systems, 51*(2), 369-394.

Geerligs, L., & Henson, R. N. (2016). Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *NeuroImage, 135*, 16-31.

Glover, F., & Laguna, M. (2013). Tabu Search*. In *Handbook of Combinatorial Optimization* (pp. 3261-3362). New York, NY: Springer.

Gorgens, E. B., Valbuena, R., & Rodriguez, L. C. E. A. (2017). Method for optimizing height threshold when computing airborne laser scanning metrics. *Photogrammetric Engineering and Remote Sensing, 83*(5), 343-350.

Górecki, T., Mirosław, K., & Woły´nski, W. (2017). Correlation analysis for multivariate functional data. In *Data Science* (pp. 243-258). Springer, Cham.

Heller, R., Heller, Y., Kaufman, S., Brill, B., & Gorfine, M. (2016). Consistent distribution-free k-sample and independence tests for univariate random variables. *The Journal of Machine Learning Research, 17*(1), 978-1031.

Kwok, H. S., & Tang, W. K. (2005). A chaos-based cryptographic hash function for message authentication. *International Journal of Bifurcation and Chaos, 15*(12), 4043-4050.

Linge, Y., Dumas, C., & Lambert-Lacroix, S. (2014). Mutual information coefficient analysis. *International Association for Cryptologic Research, 12*, 2012.

Lipowski, A., & Lipowska, D., (2012). Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications, 391*(6), 2193-2196.

Martínez-Gómez, E., Richards, M. T., & Richards, D. S. P. (2014). Distance correlation methods for discovering associations in large astrophysical databases. *The Astrophysical Journal, 781*(1), 39-49.

Mitchell, M. (1998). *An introduction to genetic algorithms*. USA: MIT press.

Naboulsi, D., Fiore, M., Ribot, S., & Stanica, R. (2016). Large-scale mobile traffic analysis: A survey. *IEEE Communications Surveys and Tutorials, 18*(1), 124-161.

Pan, W., Wang, X., Wen, C., Styner, M., & Zhu, H. (2017). Conditional local distance correlation for manifold-valued data. In *International Conference on Information Processing in Medical Imaging* (pp. 41-52). Springer, Cham.

Reshef, D., Reshef, Y., Finucane, H., Grossman, S., McVean, G., Turnbaugh, P., … & Sabeti, P. (2011). Detecting novel associations in large datasets. *Science, 334*(662), 1518-1524.

Rivest, R. (1992). The MD5 message-digest algorithm. *MIT Laboratory for Computer Science and RSA Data Security.* Network Working Group.

Schneier, B. (2004, August 19). Cryptanalysis of MD5 and SHA: Time for a New Standard. In *Computer World, Crypto 2004 conference* (pp. 1-2). Santa Barbara, California.

Shao, F., & Li, K. (2016). A graph model for preventing railway accidents based on the maximal information coefficient. *International Journal of Modern Physics B, 31*(3), 1750010.

Siarry, P. (2016). Simulated Annealing. *Metaheuristics. Springer, Cham*, 19-50.

Sun, Y., Kirley, M., & Halgamuge, S. K. (2017). Quantifying variable interactions in continuous optimization problems. *IEEE Transactions on Evolutionary Computation, 21*(2), 249-264.

Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics, 3*(4), 1236-1265.

Székely, G. J., & Rizzo, M. L., (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis, 117*, 193-213.

Wang, S., Zhao, Y., Shu, Y., & Shi, W. (2017). Improved approximation algorithm for maximal information coefficient. *International Journal of Data Warehousing and Mining (IJDWM), 13*(1), 76-93.

Wang, S., Zhao, Y., Shu, Y., Yuan, H., Geng, J., & Wang, S. (2018). Fast search local extremum for maximal information coefficient (MIC). *Journal of Computational and Applied Mathematics, 327*(1), 372-387.

Wang, T., & Li, Z. (2017). Outlier detection in high-dimensional regression model. *Communications in Statistics-Theory and Methods, 46*(14), 6947-6958.

Wang, T., Zheng, L., Li, Z., & Liu, H. (2017). A robust variable screening method for high-dimensional data. *Journal of Applied Statistics, 44*(10), 1839-1855.

Wang, X., Jiang, B., & Liu, J. S. (2017). Generalized R-squared for Detecting No independence. *Biometrika, 104*(1), 129-139.

Zhai, A., Xu, F., Pan, H., Shi, J., & Xiong, G. (2017). Design for network file forensics system based on approximate matching. In *Advanced Multimedia and Ubiquitous Engineering* (pp. 426-431). Springer, Singapore.

Zhang, Y., & Xiao, D. (2013). Cryptanalysis of S-box-only chaotic image ciphers against chosen plaintext attack. *Nonlinear Dynamics, 72*(4), 751-756.

Zhang, Y., Xiao, D., Wen, W., & Li, M. (2014). Breaking an image encryption algorithm based on hyper-chaotic system with only one round diffusion process. *Nonlinear Dynamics, 76*(3), 1645-1650.

Simulation of Fermentation Compounds for Bioethanol Production Using Different Separating Agents

S. M. Anisuzzaman*, D. Krishnaiah, A. Bono, F. A. Lahin and Syazryn R. I.

Chemical Engineering Programme, Faculty of Engineering, Universiti Malaysia Sabah, 88400 UMS, Kota Kinabalu, Sabah, Malaysia

ABSTRACT

This study focuses on the simulation of ethylene glycol ($C_2H_6O_2$) – glycerol ($C_3H_8O_3$) and ethylene glycol ($C_2H_6O_2$) – calcium chloride ($CaCl_2$) as separating agents in bioethanol production from fermentation effluent. The entire process was simulated using Aspen HYSYS V7.3 software, but the main focus is the extractive distillation where the mixture compounds were utilised. Response Surface Methodology (RSM) was used to optimise the process variables in extractive distillation column with the separating agent ethylene glycol, temperature, solvent to feed molar ratio and reflux ratio. Non-random two-liquid (NRTL) model was used for activity coefficients of mixture from Aspen properties databank. Results show that both mixture compounds values on solvent to feed molar ratio, reflux ratio and reboiler energy consumption were slightly different as separating agent temperature maintained at 80 oC and ethanol composition in distillate was 99.89 mole%. The separating agents show that better ethanol-water separation with lower energy consumption compared with a well-known single compound such as ethylene glycol. Thus, this study is important to improve extractive distillation column operating conditions by studying the effect of mixture compounds as separating agents in bioethanol production.

Keywords: Calcium chloride, extractive distillation, ethylene glycol, glycerol, simulation

INTRODUCTION

Bioethanol is a promising alternative new energy to ensure adequate supply of low cost fuel that burns completely and without unwanted exhaust emissions (Manivannan and Narendhirakannan, 2015; Govumoni et al., 2013; Nikzad et al., 2012). The current source of bioethanol generation is crop as well as agricultural residues and they serve as a cheap and abundant feedstock to produce bioethanol at reasonable cost (Won et al.,

2012; Balat, 2011; Goh et al., 2010; Hii et al., 2010; Fenning et al., 2008; Basiron, 2005). Bioethanol is generated from biomass by hydrolysis process, followed by sugar fermentation process. Biomass contains a complex combination of carbohydrate polymers from plant cell walls known as cellulose, hemi cellulose and lignin. Sugar is extracted from biomass by pre-treating it with acids or enzymes to reduce the particle sizes and to opening up the fibrous structure (Nitayavardhana et al., 2010). The cellulose and hemi cellulose are broken down by enzymes or acids into glucose, and then into bioethanol by anaerobic fermentation (Ajibola et al., 2012).

However, in alcoholic fermentation process, a large quantity of polluted water is generated due to presence of oxygen (Nikzad et al., 2012; Lin et al., 2012). Thus, purification process for final fermented product of bioethanol is required to recover remaining ethanol with the aid of third component, in this case extractive compounds are useful in order to produce dry bioethanol production (Anwar & Neni, 2012; Gil et al., 2008).

There are many alternatives to distillation for recovering ethanol from aqueous solutions, such as membrane permeation, vacuum stripping, gas stripping, solvent extraction, adsorption and various hybrid processes (Offeman et al., 2008). For ethanol dehydration, extractive distillation is commonly used to recover ethanol. Extractive distillation is used to separate azeotropic mixture by applying third solvent, known as entrainer or separating agent to alter the relative volatility of the compounds of the mixture (Gil et al., 2008). However, many studies have been done using a pure component as entrainer to improve the value of bioethanol (Gil et al., 2014; Gil et al., 2012; Llano-Restrepo & Aguilar-Arias, 2003). Hence, the purpose of this work was to investigate the effect of entrainer as solvent in extractive distillation process of bioethanol from fermentation effluent in order to produce high purity bioethanol. Optimisation of ethanol dehydration is also important due to high energy consumption and capital investment cost (Bastidas et al., 2010).

There are three main objectives for this study. First, to develop a simulation process of bioethanol production from fermentation effluent using Aspen HYSYS V7.3 software and the main focus for the simulation is the extractive distillation. Second, to study the effect of process variables in extractive distillation which are separating agents concentration glycerol ($C_3H_8O_3$) and calcium chloride ($CaCl_2$) in ethylene-glycol ($C_2H_6O_2$), feed temperature, solvent to feed molar ratio and reflux ratio (Gil et al., 2008; Gil et al., 2014). The responses are ethanol composition in distillate and energy consumption by reboiler. Third, to design an experiment using response surface methodology (RSM) for analysing the contribution of process variables in order to optimise the extractive distillation column.

METHODS

Process Flow

There are two main steps to obtain dry ethanol: removal of solid and liquid substance from effluent and bioethanol refining. The bioethanol dehydration process is shown in Figure 1.
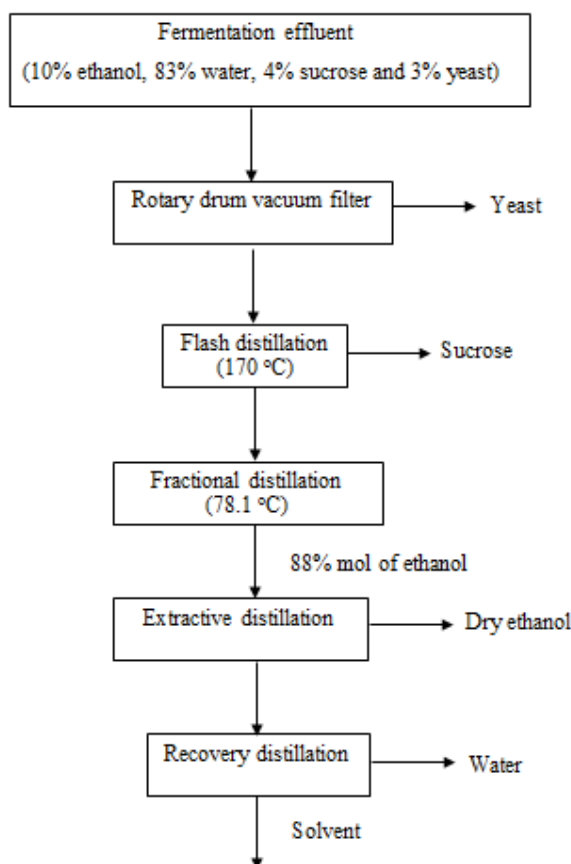
Figure 1. Process flow diagram of bioethanol production from fermentation effluent

Materials

The fermented feed proposed in this process contains yeast, sucrose, ethanol and water. By taking composition of dissolved carbohydrate from molasses and carbohydrate conversion into ethanol entering distillation column, the proportion of water and ethanol are be 83% and 10% respectively (75% sugar converted) respectively. The rest of the unconverted sugar consists of sucrose (4%), glucose and fructose. The remaining is assumed as the equivalent amount of yeast. To conclude, the compositions are 10% ethanol, 83% water, 4% sucrose and 3% yeast (Amin et al., 2013; Arvinius et al., 2010). The components are set according to their name in HYSYS except for yeast which is named carbon. Carbon serves the same function as yeast that can settle in the rotary drum vacuum filter.

Process Description

Starting from rotary drum vacuum filter, the temperature of the unit is 25°C and yeast is removed from the effluent that leaves ethanol, water and sucrose as the by-product. The stream is feed to flash distillation where the feed temperature is adjusted to 170°C and sucrose is eliminated. Then, ethanol and water are fed to fractional distillation and the feed temperature is lowered to

78.1°C. The column parameters such as number of stages, distillate flow and reflux ratio are 10, 100 kmol/hr and 0.3 respectively and adjusted until 88% ethanol is achieved in the top stream. Lastly, ethanol-water or azeotrope mixture enters the extractive distillation and it is mixed with the separating agent. The feed temperature is maintained at 78.1°C. Other parameters such, as distillate rate, azeotrope molar flow, number of theoretical stages, separating agent stage and azeotrope stage are 86.8 kmol/hr, 100 kmol/hr, 18, 3 and 12 respectively. Table 1 shows the initial process variables for extractive column (Gil et al., 2014; Gil et al., 2008).

Table 1
Initial value of process variables for extractive column

Parameter	Initial value
Separating agent concentration in ethylene glycol, mole%	$C_3H_8O_3 = 0.4$; $CaCl_2 = 0.05$
Separating agent temperature, °C	80
Solvent-feed ratio	0.8
Reflux ratio	0.5

Simulation Process

Aspen HYSYS V7.3 software is used to develop simulation for bioethanol production from fermentation effluent, while *non-random two-liquid* (NRTL) thermodynamic model is used to calculate the activity coefficients for the mixture (Llano-Restrepo and Aguilar-Arias, 2003; Gil et al., 2012; Gil et al., 2014; Lars, 2012). Initial data as discussed previously, is used and adjusted in the simulation to satisfy the end result.

Statistical Analysis

In order to study the interaction between e process variables - third component feed temperature, solvent concentration in ethylene glycol ($C_2H_6O_2$) – glycerol ($C_3H_8O_3$) and ethylene glycol ($C_2H_6O_2$) – calcium chloride ($CaCl_2$), solvent to feed molar ratio and reflux ratio, a statistical analysis is required, and to study the interaction between variables, considering the composition of ethanol in distillate and energy consumption by reboiler in order to optimise the process and to obtain anhydrous ethanol. RSM in Design Expert 7 software is used to perform the process variables relation (Myers & Montgomery, 1995).

RESULTS AND DISCUSSION

Simulation

The simulation is done by parts based on HYSYS and Aspen Properties databank at steady state (Julio, n.d.; Smejkal & Soos, 2002; Seider et al., 1999). The first part is the rotary drum vacuum filter until fractional distillation, the thermodynamic model general NRTL is used in HYSYS. The second part is extractive and recovery distillation column where NRTL is

used to estimate the activity coefficients of mixture from Aspen properties. The reason is that carbon may not be pure component as solid form in Aspen properties. Hence, separation of carbon cannot be done. General NRTL is performed because it analyses the system according to molecular interaction. However, carbon atom is considered as solid. Additionally, the NRTL model is the best fit for ethanol-water system or exhibits phase splitting (Ngema, 2010). Figure 2 shows the simulation process of bioethanol production from fermentation effluent. In Figure 2, the blue line indicates the simulation is working. The red dash line region shows an extractive distillation where ethanol-water separation occurs - the focus of this study. The bottom distillate for recovery column is not recycled into the extractive column in order to maintain the composition and temperature of initial separating agent entering the column as it is studied in a particular range.
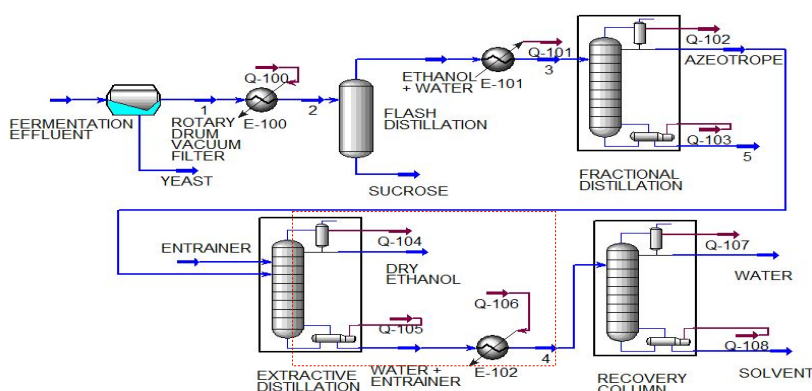
Figure 2. Simulation process of bioethanol production from fermentation effluent

In order to study the effect of process variables related to separating agent in extractive distillation process, these variables are set in a range, solvent concentration in $C_2H_6O_2$ for $C_3H_8O_3$ is 0.3 to 0.7; meanwhile $CaCl_2$ is 0.05 to 0.1. The separating agent temperature, solvent to feed molar ratio and reflux ratio are 60 to 110°C, 0.3 to 1 and 0.3 to 0.8 respectively. As for responding variables, molar fraction of ethanol in distillate and reboiler energy consumption, kJ/kg of ethanol is chosen. The four ranges of process variables are important which are used in Design Expert 7.0 to help in interpretation of the multi-factor to produce comprehensive data (Buxton, 2007; Myers and Montgomery, 1995). The responses values are keyed in based on Aspen HYSYS V7.3 result according to values of factors as shown in Table 2 and 3.

Table 2

Design process variables data for $C_2H_6O_2$ and $C_3H_8O_3$

Factor 1	Factor 2	Factor 3	Factor 4	Response 1	Response 2
$C_3H_8O_3$ concentration in $C_2H_6O_2$, mole fraction	Separating agent temperature, °C	Solvent-feed molar ratio	Reflux ratio	Molar fraction of ethanol in distillate	Reboiler energy consumption, (kJ/kg of ethanol)
0.50	110.00	0.65	0.55	0.999	1473.22
0.50	85.00	0.65	0.55	0.9992	1552.91
0.70	60.00	0.30	0.30	0.9807	1218.12
0.30	60.00	0.30	0.30	0.9711	1203.87
0.70	60.00	1.00	0.30	0.9999	1723.93
0.30	110.00	1.00	0.30	0.9956	1361.24
0.30	60.00	0.30	0.80	0.9804	1629.71
0.70	60.00	1.00	0.80	1.0000	2146.29
0.50	85.00	0.65	0.55	0.9992	1551.91
0.50	85.00	0.65	0.55	0.9992	1551.91
0.50	85.00	0.65	0.30	0.9975	1336.84
0.70	110.00	0.30	0.80	0.9864	1571.86
0.50	85.00	0.65	0.55	0.9992	1552.41
0.30	110.00	0.30	0.60	0.9803	1568.10
0.70	110.00	1.00	0.80	0.9999	1898.47
0.30	60.00	1.00	0.30	0.9994	1595.05
0.50	85.00	0.30	0.55	0.9821	1393.68
0.50	85.00	1.00	0.55	0.9999	1752.19
0.50	85.00	0.65	0.80	0.9993	1765.57
0.50	60.00	0.65	0.55	0.9992	1625.97
0.50	85.00	0.65	0.55	0.9992	1552.66
0.70	85.00	0.65	0.55	0.9996	1583.79
0.30	60.00	1.00	0.80	0.9998	2018.76
0.70	60.00	0.30	0.80	0.9863	1648.01
0.30	85.00	0.65	0.55	0.9983	1516.40
0.70	110.00	0.30	0.30	0.9749	1142.97
0.70	110.00	1.00	0.30	0.9994	1473.24
0.50	85.00	0.65	0.55	0.9992	1553.16
0.30	110.00	1.00	0.80	0.9998	1799.40
0.30	110.00	0.30	0.30	0.9658	1135.09

The tables show that reboiler energy consumption increases as molar fraction of ethanol in distillate increases. The result is analysed to verify the factors of significant effect toward the responses. The analysis is called ANOVA (analysis of variance) used as alternative method to highlight the active factor (Myers & Montgomery, 1995).

Table 3
Design process variables data for $C_2H_6O_2$ and $CaCl_2$

Factor 1	Factor 2	Factor 3	Factor 4	Response 1	Response 2
$CaCl_2$ concentration in $C_2H_6O_2$, mole fraction	Separating agent temperature, °C	Solvent-feed molar ratio	Reflux ratio	Molar fraction of ethanol in distillate	Reboiler energy consumption, (kJ/kg of ethanol)
0.08	60.0	0.65	0.55	0.9951	1535.74
0.05	110.00	1.00	0.30	0.9870	1284.96
0.08	85.00	0.30	0.55	0.9689	1372.16
0.10	85.00	0.65	0.55	0.9937	1472.14
0.05	60.00	1.00	0.30	0.9976	1493.49
0.08	85.00	0.65	0.80	0.9962	1687.72
0.05	110.00	1.00	0.80	0.9992	1724.04
0.10	60.00	1.00	0.30	0.9967	1484.72
0.08	85.00	0.65	0.55	0.9943	1474.28
0.08	85.00	0.65	0.55	0.9943	1474.28
0.10	110.00	1.00	0.30	0.9850	1282.54
0.08	85.00	1.00	0.55	0.9985	1610.51
0.05	60.00	0.30	0.30	0.9622	1190.68
0.08	85.00	0.65	0.55	0.9943	1474.28
0.10	60.00	0.30	0.30	0.9594	1192.05
0.10	60.00	1.00	0.80	0.9992	1911.43
0.08	85.00	0.65	0.55	0.9942	1473.29
0.05	60.00	0.30	0.80	0.9742	1620.43
0.10	60.00	0.30	0.80	0.9716	1612.98
0.10	110.00	0.30	0.30	0.9553	1136.25
0.05	110.00	0.30	0.80	0.9737	1561.26
0.05	85.00	0.65	0.55	0.9950	1474.54
0.08	85.00	0.65	0.55	0.9943	1471.77
0.08	85.00	0.65	0.30	0.9845	1255.43
0.05	110.00	0.30	0.30	0.9577	1134.72
0.10	110.00	1.00	0.80	0.9989	1722.22
0.08	85.00	0.65	0.55	0.9942	1472.77
0.10	110.00	0.30	0.80	0.9711	1562.23
0.05	60.00	1.00	0.80	0.9994	1919.94
0.08	110.00	0.65	0.55	0.9928	1408.09

In ANOVA, if the p-value is less than 0.05, it means the model (or term) is statistically significant. A value of more than 0.10 indicates the model is not significant and if it is in between 0.05 and 0.10, then the model may be significant (Ngema, 2010; Anderson & Whitcomb, 2005). The analysed result is a quadratic model that allows for curvature in the effect of a control on the response (Buxton, 2007; Myers & Montgomery, 1995). Tables 4 shows the ANOVA for response surface quadratic model on both responses.

Table 4
ANOVA for $C_2H_6O_2$- $C_3H_8O_3$ on responses

Source	Response 1					Response 2				
	Sum of squares	df	Mean square	F value	p-value Prob > F	Sum of squares	df	Mean square	F Value	p-value Prob > F
Model	3.012×10^{-3}	14	2.152×10^{-4}	139.13	< 0.0001	1.583×10^{6}	14	1.131×10^{5}	17672.78	< 0.0001
A	7.442×10^{-5}	1	7.442×10^{-5}	48.13	< 0.0001	18499.33	1	18499.33	2891.49	< 0.0001
B	1.369×10^{-5}	1	1.369×10^{-5}	8.86	0.0094	1.067×10^{5}	1	1.067×10^{5}	16683.45	< 0.0001
C	1.916×10^{-3}	1	1.916×10^{-3}	1238.89	< 0.0001	5.894×10^{5}	1	5.894×10^{5}	92123.70	< 0.0001
D	1.275×10^{-4}	1	1.275×10^{-4}	82.43	< 0.0001	8.260×10^{5}	1	8.260×10^{5}	1.291×10^{5}	< 0.0001
AB	5.256×10^{-7}	1	5.256×10^{-7}	0.34	0.5685	274.31	1	274.31	42.88	< 0.0001
AC	4.258×10^{-5}	1	4.258×10^{-5}	27.53	< 0.0001	11198.95	1	11198.95	1750.42	< 0.0001
AD	7.156×10^{-6}	1	7.156×10^{-6}	4.63	0.0482	12.86	1	12.86	2.01	0.1767
BC	2.806×10^{-6}	1	2.806×10^{-6}	1.81	0.1980	28054.58	1	28054.58	4384.99	< 0.0001
BD	1.463×10^{-5}	1	1.463×10^{-5}	9.46	0.0077	34.51	1	34.51	5.39	0.0347
CD	7.966×10^{-5}	1	7.966×10^{-5}	51.51	< 0.0001	4.18	1	4.18	0.65	0.4314
A^2	8.243×10^{-7}	1	8.243×10^{-7}	0.53	0.4766	0.43	1	0.43	0.067	0.7986
B^2	4.441×10^{-7}	1	4.441×10^{-7}	0.29	0.5999	9.24	1	9.24	1.47	0.2438
C^2	1.878×10^{-4}	1	1.878×10^{-4}	121.45	< 0.0001	1189.91	1	1189.91	185.99	< 0.0001
D^2	3.216×10^{-6}	1	3.216×10^{-6}	2.08	0.1699	0.23	1	0.23	0.036	0.8530

Whereby, A = $C_3H_8O_3$ concentration in $C_2H_6O_2$, mole fraction; B = Separating agent temperature, °C; C = Solvent-feed molar ratio and D = Reflux ratio for both equations

The ANOVA analysis for $C_2H_6O_2$- $C_3H_8O_3$ of responses with variation of factors is represented by polynomial equations 1 and 2.

$$X = 1 + 2.033 \times 10^{-3} A - 8.722 \times 10^{-4} B + 0.010 C + 2.661 \times 10^{-3} D + 1.813 \times 10^{-4} AB - 1.631 \times 10^{-3} AC - 6.687 \times 10^{-4} AD + 4.188 \times 10^{-4} BC + 9.563 \times 10^{-4} BD - 2.231 \times 10^{-3} CD - 5.640 \times 10^{-4} A^2 - 4.140 \times 10^{-4} B2 - 8.514 \times 10^{-3} C^2 - 1.114 \times 10^{-3} D^2 \qquad (1)$$

$$X = 1552 + 32.06 A - 77.01 B + 180.95 C + 214.21 D - 4.14 AB + 26.46 AC - 0.90 AD - 41.87 BC + 1.47 BD - 0.51 CD - 0.41 A^2 - 1.91 B2 + 21.43 C^2 - 0.30 D^2 \qquad (2)$$

Based on Table 4 for reboiler energy consumption, the four factors show significant p-value that is less than 0.05. In other words, the effects are significant for the response (CAMO Software AS, 2016). As for Table 4 for molar fraction of ethanol in distillate, only separating agent feed temperature is not significant, and it can be concluded that the investigated factor does not affect the response. In contrast, the concentration of calcium chloride in ethylene glycol is not showing significant result for both responses as can be seen in Table 5.

The ANOVA analysis for $C_2H_6O_2$ - $CaCl_2$ of responses with variation of factors is represented by polynomial equations 3 and 4.

$$X = 0.99 - 8.389 \times 10^{-4} A - 1.928 \times 10^{-3} B + 0.015 C + 5.45 \times 10^{-3} D - 5 \times 10^{-5} AB + 4.375 \times 10^{-4} AC + 1.5 \times 10^{-4} AD - 8.25 \times 10^{-4} BC + 1.838 \times 10^{-3} BD - 1.6 \times 10^{-3} CD + 3.535 \times 10^{-4} A^2 - 4.649 \times 10^{-5} B^2 - 0.01 C^2 - 3.646 \times 10^{-3} D^2 \qquad (3)$$

$$X = 1473.19 - 1.53 A - 63.62 B + 113.95 C + 214.86 D + 1.35 AB - 1.12 AC - 0.53 AD - 35.88 BC + 1.72 BD + 1.80 CD + 0.40 A^2 - 1.03 B^2 + 18.40 C^2 - 1.37 D^2 \qquad (4)$$

Both solvents with the four factors studied has shown a model with significant effect on the response molar fraction of ethanol in distillate and reboiler energy consumption. Then, model graphs can illustrate the relationship between factors and responses in 3D surface. Two variables and one response for three tests are selected to standardise the study on the effect for both mixture compounds as separating agent.

Effect of Separating Agent Concentration in $C_2H_6O_2$

The effect of separating agent concentration is analysed by fixing two factors at constant such as separating agent temperature, and solvent to feed ratio, 80°C and 0.65 respectively. Figure 3 shows the ethanol in distillate curvature purity increases moderately from the lowest value of reflux ratio and glycerol to the highest. Figure 4 shows a linear behaviour of negative slope when the reflux ratio value decreases at constant glycerol concentration in $C_2H_6O_2$. Meanwhile, increasing the value of concentration of $C_3H_8O_3$ in $C_2H_6O_2$ consumes higher energy in reboiler at constant reflux ratio.

Table 5
ANOVA for $C_2H_6O_2$ - $CaCl_2$ on responses

Source	Response 1					Response 2				
	Sum of squares	df	Mean square	F value	p-value Prob > F	Sum of squares	df	Mean square	F Value	p-value Prob > F
Model	5.902×10^{-3}	14	4.215×10^{-4}	411.96	< 0.0001	1.160×10^{6}	14	82885.31	10081.05	< 0.0001
A	1.267×10^{-5}	1	1.267×10^{-5}	12.38	0.0031	42.02	1	42.02	5.11	0.0391
B	6.689×10^{-5}	1	6.689×10^{-5}	65.37	< 0.0001	72854.29	1	72854.29	8861.01	< 0.0001
C	3.972×10^{5}	1	3.972×10^{-3}	3882.00	< 0.0001	2.337×10^{5}	1	2.337×10^{5}	28427.01	< 0.0001
D	5.346×10^{-4}	1	5.346×10^{-4}	522.48	< 0.0001	8.309×10^{5}	1	8.309×10^{5}	1.011×10^{5}	< 0.0001
AB	4×10^{-8}	1	4×10^{-8}	0.039	0.8459	29.21	1	29.21	3.55	0.0790
AC	3.063×10^{-6}	1	3.063×10^{-7}	2.99	0.1041	20.09	1	20.09	2.44	0.1389
AD	3.6×10^{-7}	1	3.6×10^{-7}	0.35	0.5619	4.53	1	4.53	0.55	0.4692
BC	1.089×10^{-5}	1	1.089×10^{-5}	10.64	0.0052	20601.88	1	20601.88	2505.74	< 0.0001
BD	5.402×10^{-5}	1	5.402×10^{-5}	52.79	< 0.0001	47.11	1	47.11	5.73	0.0302
CD	4.096×10^{-5}	1	4.096×10^{-5}	40.03	< 0.0001	51.60	1	51.60	6.28	0.0243
A^2	3.238×10^{-7}	1	3.238×10^{-7}	0.32	0.5821	0.41	1	0.41	0.050	0.8260
B^2	5.6×10^{-9}	1	5.6×10^{-9}	$5.473 \times 10\text{-}3$	0.9420	2.74	1	2.74	0.33	0.5724
C^2	$2.747 \times 10\text{-}4$	1	2.747×10^{-4}	268.43	< 0.0001	876.87	1	876.87	106.65	< 0.0001
D^2	3.445×10^{-5}	1	3.445×10^{-5}	33.67	< 0.0001	4.86	1	4.86	0.59	0.4541

Whereby, A = $CaCl_2$ concentration in $C_2H_6O_2$, mole fraction; B = Separating agent temperature, °C; C = Solvent-feed molar ratio; D = Reflux ratio
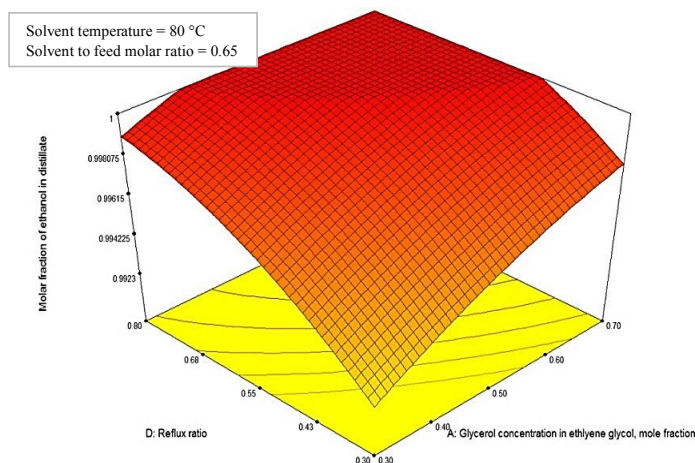
Figure 3. Effect of $C_3H_8O_3$ concentration in $C_2H_6O_2$ and reflux ratio on response 1
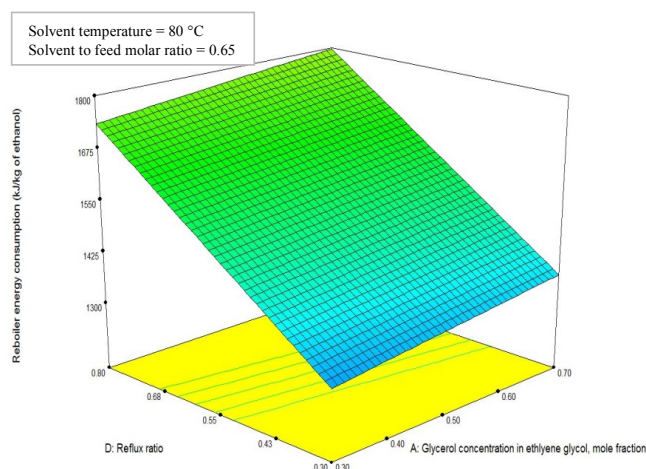
Figure 4. Effect of $C_3H_8O_3$ concentration in $C_2H_6O_2$ and reflux ratio on response 2

For $C_2H_6O_2$ and $C_3H_8O_3$, the ethanol in distillate purity increases moderately until it reaches the maximum concentration from the lowest value of reflux ratio and $C_3H_8O_3$ concentration until the highest. In terms of energy, it presents a linear behaviour of negative slope when reflux ratio value decreases at constant $C_3H_8O_3$ concentration and increasing value of $C_3H_8O_3$ concentration consumes higher energy at reboiler by constant reflux ratio. The most adequate composition of $C_3H_8O_3$ in $C_2H_6O_2$ is between 40 until 50 mol% of $C_3H_8O_3$. Hence, 40 mol% of $C_3H_8O_3$3 is chosen.

The reduction of energy consumption for reboiler is explained by lowering the reflux ratio compared with $C_3H_8O_3$ concentration. Therefore, the most adequate composition of $C_3H_8O_3$ in $C_2H_6O_2$ is between 40 until 50 mol% of glycerol. Hence, 40 mol% of $C_3H_8O_3$ is chosen. This concentration is within the range as reported by Gil et al. (2014) and by setting 60 mol% of

C_2H6O_2 and 40 mol% of $C_3H_8O_3$, a distillate in steam with 0.999 or higher, molar composition of ethanol can be achieved.

Figures 5 and 6 show the $CaCl_2$ concentration increases at constant reflux ratio, ethanol in distillate slightly decreases and energy consumption slope almost approaches zero. This pattern of $CaCl_2$ concentration affecting the distillate purity is also reported by Gil et al., (2008) that distillate purity is higher for the 0.05 and 0.075 g of $CaCl_2$/mL of C_2H6O_2 compared with 0.075 and 0.1 g of $CaCl_2$/mL of C_2H6O_2. Energy consumption can be minimised as benefit of salt in solvent that helps to reduce number of theoretical stages needed for the separation, which is lower than only using C_2H6O_2 from previous study. Presence of $CaCl_2$ combined with low reflux ratio diminishes the amount of energy consumption. Hence, the $CaCl_2$ concentration is fixed at 0.075 mol% to achieve lower energy utilisation at reboiler.
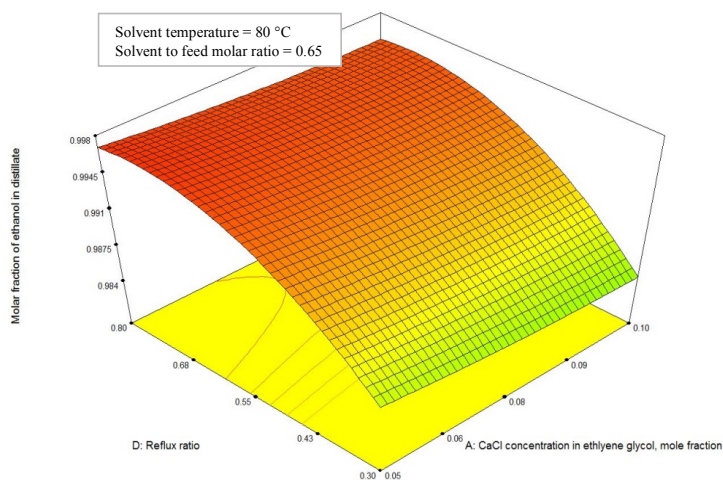
Figure 5. Effect of $CaCl_2$ concentration in $C_2H_6O_2$ and reflux ratio on response 1
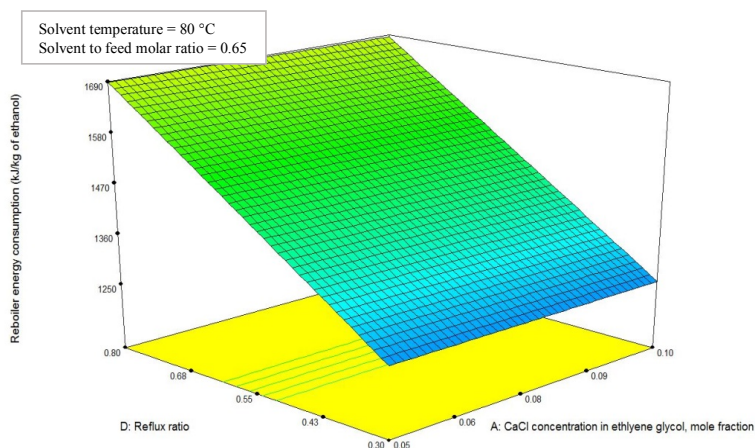
Figure 6. Effect of $CaCl_2$ concentration in $C_2H_6O_2$ and reflux ratio on response 2

Effect of Separating Agent Feed Temperature

The initial feed temperature of separating agent has a significant effect on purity of ethanol in distillate and energy utilisation compared with reflux ratio. Increasing the temperature to separate the agent entering the column requires high reflux ratio to reach high ethanol composition in distillate as shown in Figures 7, 8, 9, and 10. This is because as the temperature of separating agent increases, water at the top stage vaporises and it condenses together with ethanol that leads to reduction of ethanol purity (Gil et al., 2008). Thus, higher reflux ratio is needed to balance this effect. For both cases, 80°C was used as the temperature of separating agent.
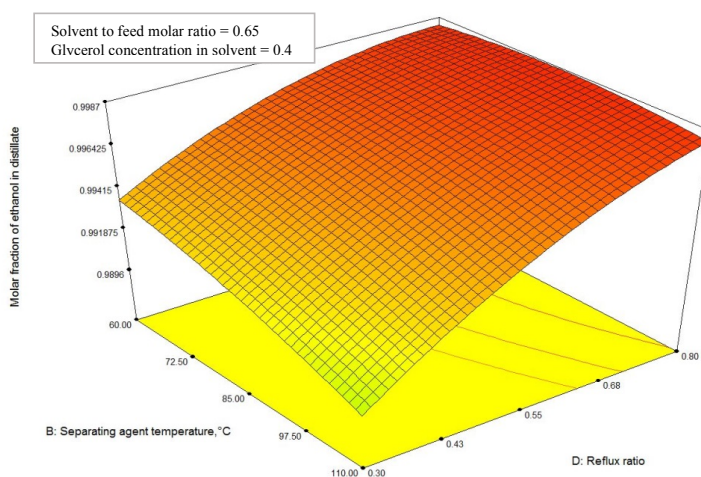
Figure 7. Effect of $C_2H_6O_2$-$C_3H_8O_3$ temperature and reflux ratio on response 1
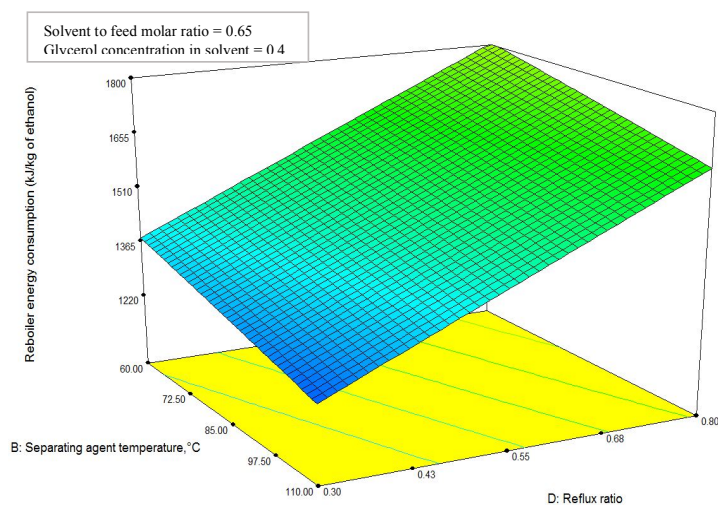
Figure 8. Effect of $C_2H_6O_2$-$C_3H_8O_3$ temperature and reflux ratio on response 2
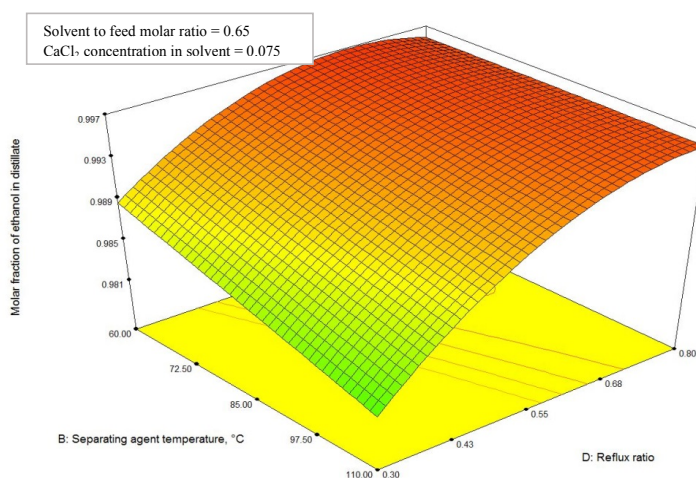
Figure 9. Effect of $C_2H_6O_2$- $CaCl_2$ temperature and reflux ratio on response 1
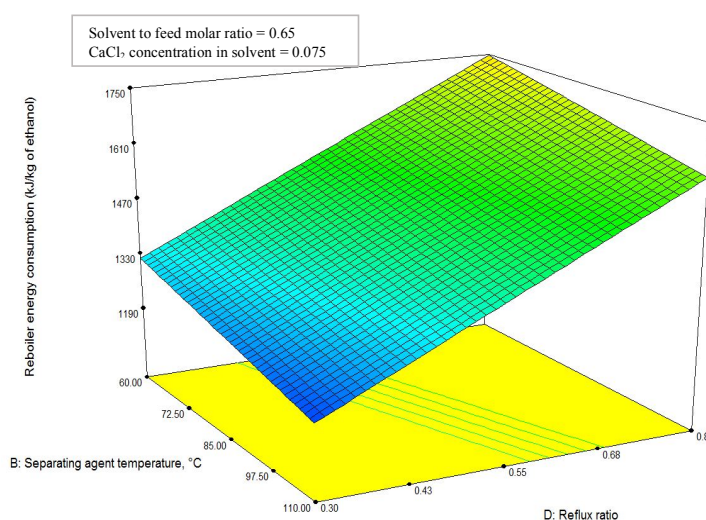
Figure 10. Effect of $C_2H_6O_2$- $CaCl_2$ temperature and reflux ratio on response 2

Effect of Solvent to Feed Molar Ratio

The concentration of $C_3H_8O_3$ and $CaCl_2$ in $C_2H_6O_2$ was chosen constant at 0.4 mol% and 0.075 mol% respectively and the separating agent temperature was set at 80°C. Based on Gil et al., (2008), solvent to feed ratio causes a direct effect on the distillate purity. In order to evaluate this effect, the solvent to feed molar ratio with a of range 0.3 to 1.0 is analysed. Figures 11, 12, 13, and 14 represent the results of the assessment.
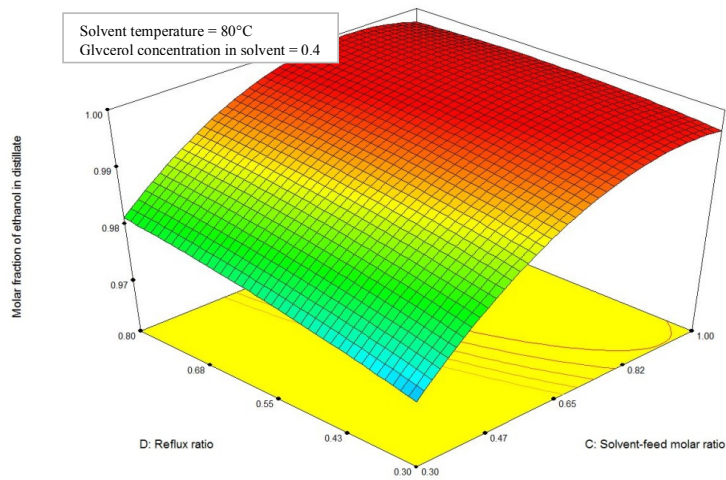
Figure 11. Effect of $C_2H_6O_2$ -$C_3H_8O_3$ to feed ratio and reflux ratio on response 1
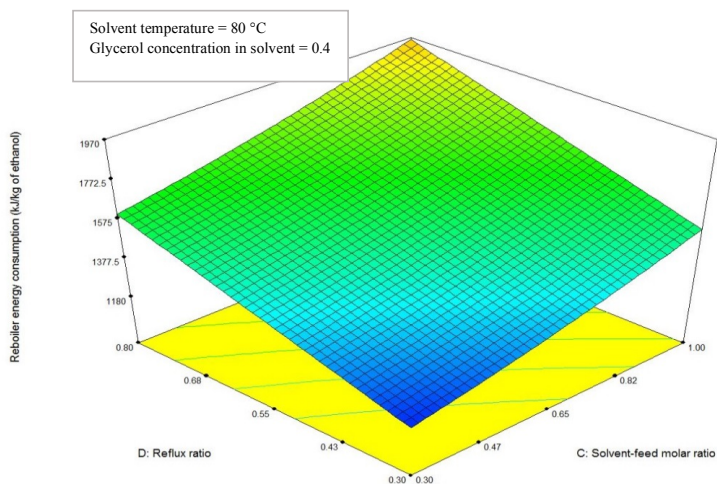
Figure 12. Effect of $C_2H_6O_2$ -$C_3H_8O_3$ to feed ratio and reflux ratio on response 2

Different values of solvent to feed ratio at constant reflux ratio can either increase or reduce energy consumption. In order to maintain the solvent to feed ratio with increasing reflux ratio causes increase in energy consumption. However, by making solvent to feed ratio constant and setting the reflux ratio until targeted distillate is achieved needs more energy in reboiler. For this reason, the reflux ratio must be performed in the lowest value, hence, the solvent to feed ratio can be controlled to obtain desired ethanol purity with low energy consumption.
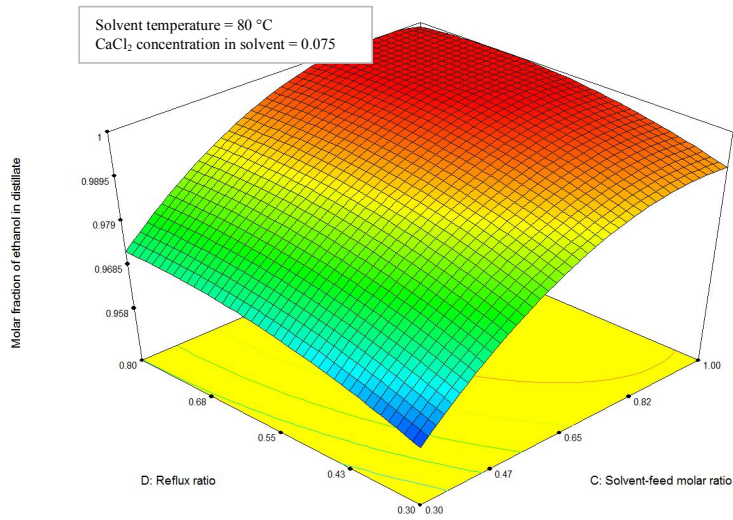
Figure 13. Effect of $C_2H_6O_2$-$CaCl_2$ to feed ratio and reflux ratio on response 1
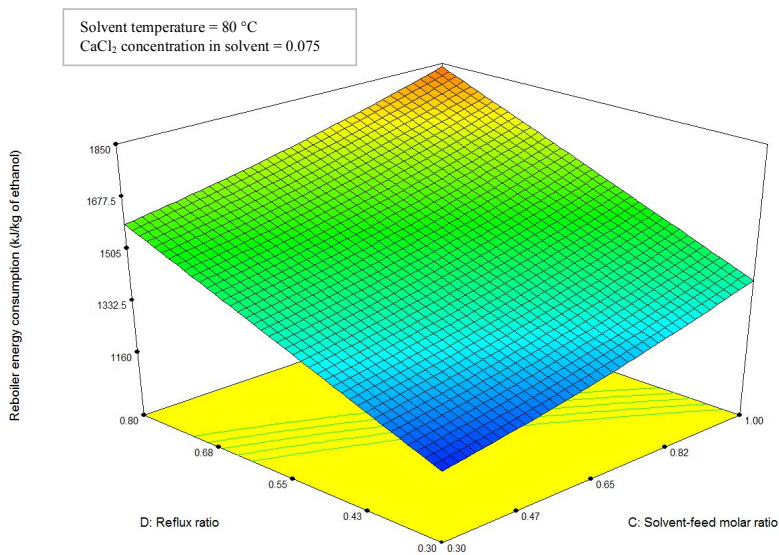
Figure 14. Effect of $C_2H_6O_2$-$CaCl_2$ to feed ratio and reflux ratio on response 2

Summary of the Operating Conditions

A summary of the operating conditions achieved for the extractive column using the two mixtures is shown in Table 6. Using only ethylene glycol as solvent in extractive distillation to separate ethanol-water, the amount of energy consumed to produce dry ethanol is 1760 kJ/kg of ethanol (Meirelles et al., 1992). Based on Table 6, the amount of energy required to obtained dry ethanol using proposed mixture is lower than using ethylene glycol as solvent

alone. The relative volatility of ethanol - water when using ethylene glycol mixture with CaCl2 with salt concentration of 0.2 g/ml of solvent is 2.56 rather than only $C_2H_6O_2$ is 1.85 (Lei et al., 2003). This is due to the presence of ions in solid salt for a stronger effect compared with molecules of a liquid agent including both in the strength of attractive forces they can apply on feed component molecules and degree of selectivity applied. In contrast, the strength of using glycerol in $C_2H_6O_2$ is primarily physical force including hydrogen bonding. Using mixture of liquid solvents, the most important criteria is relative volatility. Hence, finding a better additive to increase the relative volatility of ethanol – water leads to decrease in the solvent to feed ratio. Furthermore, the energy consumed at reboiler to reach about 99.9% molar concentration of ethanol in distillate using $CaCl_2$ is higher compared with glycerol. The difference is due to solvent to feed molar ratio. This is because increasing the reflux ratio leads to the dilution of solvent, reducing its function on relative volatility of the mixture ethanol-water. Thus, high solvent to feed ratio is necessary to offset the condition (Gil et al., 2008).

Table 6

Operating condition of extractive column for $C_2H_6O_2$-$C_3H_8O_3$ and $C_2H_6O_2$-$CaCl_2$

Parameter	Solvent	
	$C_3H_8O_3$	$CaCl_2$
Azeotrope feed flow (kmol/h)		100
Separating agent flow (kmol/h)		80
Distillate flow (kmol/h)		86.8
Temperature of azeotrope feed (°C)		78.1
Number of theoretical stages		18
Pressure (atm)		1
Separating agent feed stage		3
Mole % of solvent in ethylene glycol	0.40	0.075
Solvent to feed ratio	0.74	0.87
Temperature of separating agent (°C)		80
Molar reflux ratio	0.4	0.5
Mole % of ethanol in distillate	99.89	99.89
Reboiler energy consumption in extractive column (kJ/kg of ethanol)	1470.85	1527.05

CONCLUSION

In this study, four parameters which are separating agent concentration in $C_2H_6O_2$, separating agent to feed molar ratio, separating agent feed temperature and reflux ratio are analysed and optimised. It can be concluded that adding a salt and $C_3H_8O_3$ in liquid solvent provide good separation ability to achieve high ethanol in distillate with lower energy consumption compared with liquid solvent alone. Additionally, reflux ratio has shown a significant effect corresponding with energy consumption whereby as reflux ratio increases, energy required in reboiler also increases. So, reflux ratio must be performed at lowest possible value along with fixed additional solvent in $C_2H_6O_2$ and temperature at 80°C. At the same time, solvent to feed

molar ratio has become the manipulating variable for balancing changes in column operating conditions that enables the regulation of the distillate purity and reboiler energy consumption.

REFERENCES

Ajibola, F. O., Edema, M. O., & Oyewole, O. B. (2012). Enzymatic production of ethanol from cassava starch using two strains of *Saccharomyces Cerevisiae*. *Nigerian Food Journal, 30*(2), 114-121.

Amin, R. M., Hossain, S. M., & Sarker, M. (2013). Simulation of ethanol production by fermentation of molasses. *Journal of Engineering (JOE), 1*(4), 69-73.

Anderson, M. J., & Whitcomb, P. J. (2005). *RSM simplified optimizing process using response surface methods for design of experiments* (pp. 1-70). New York, NY: Productivity press.

Anwar, M., & Neni, D. (2012). Synthesis of zeolite pellets from natural zeolite and starch as adsorbent for fuel grade bioethanol production. In *Proceeding of International Conference on Chemical and Material Engineering* (pp. 1-5). Semarang, Indonesia.

Arvinius, E., Bondesson, P. M., Komorin, I., Navasa, M., & Svensson, P. (2010). Integration of first and second generation of ethanol production from wheat. *Projektering, KET050*, 1-66.

Balat, M. (2011). Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review. *Energy Conversion and Management, 52*, 858-875.

Basiron, Y. (2005). Biofuel: an alternative fuel in the Malaysia scenario. *Palm Oil Developments, 42*, 1-4.

Bastidas, P. A., Gil, I. V., & Rodriguez, G. (2010). Comparison of the main ethanoldehydration technologies through process simulation. In S. Pierucci & G. B. Ferraris (Eds.), *20th European Symposium on Computer Aided Process Engineering – ESCAPE* (pp. 1-6). Naples, Italy.

Buxton, R. (2007). *Design Expert 7: Introduction. Mathematics learning support center.* http://www.lboro.ac.uk/media/wwwlboroacuk/content/mlsc/downloads/Design%20Expert%207.pdf

CAMO Software AS. (n.d.). *Color coded table for easy interpretation.* Retrieved from http://www.camo.com/resources/anova.html.

Fenning, T. M., Walter, C., & Gartland, K. M. A. (2008). Forest biotech and climate change. *Nature Biotechnology, 26*(6), 615-617.

Gil, I. D., Uyazan, A. M., Aguilar, J. L., Rodriguez, G., & Caicedo, L. A. (2008). Separation of ethanol and water by extractive distillation with salt and solvent as entrainer: Process simulation. *Brazilian Journal of Chemical Engineering, 25*(1), 207-215.

Gil, I. D., Gomez, J. M., & Rodriguez, G. (2012). Control of an extractive distillation process to dehydrate ethanol using glycerol as entrainer. *Computer and Chemical Engineering, 39*, 129-142.

Gil, I. D., Garcia, L. C., & Rodriguez, G. (2014). Simulation of ethanol extractive distillation with mixed glycols as separating agent. *Brazilian Journal of Chemical Engineering, 31*(1), 260-269.

Goh, C. S., Tan, K. T., Lee, K. T., & Bhatia, S. (2010). Bioethanol from lignocellulose: status, perspective and challenges in Malaysia. *Bioresource Technology, 101*(13), 4834-4841.

Govumoni, S. P., Koti, S., Kothagouni, S. Y., Venkateshwar, S., & Linga, V. R. (2013). Evaluation of pretreatment methods for enzymatic saccharification of wheat straw for bioethanol production. *Carbohydrate Polymer, 91*(2), 646-650.

Hii, S., Ding, T., & Ong, L. (2010). Comparative evaluation of pretreatment strategies on enzymatic saccharification of hylocereus polyrhizus's pericarps for bioethanol production. *Iranica Journal of Energy and Environment, 1*, 275-279.

Julio, A. D. C. (n.d.). *Comparison of chemical process simulator: Aspen vs HYSYS.* Instituto Superior Técnico, Lisboa, Portugal. Retrieved from https://fenix.tecnico.ulisboa.pt/downloadFile/395137861652/Aspen_vs_HYSYS.pdf

Lars, E. (2012). Comparison of Aspen HYSYS and Aspen Plus simulation of CO2 Absorption into MEA from Atmospheric Gas. *Energy Procedia, 23*, 360-369.

Lei, Z., Li, C., & Chen, B. (2003). Extractive distillation: A review. *Separation and Purification Reviews, 32*(2), 121-213.

Lin, Y., Zhang, W., Li, C., & Kong, H. (2012). Factors affecting ethanol fermentation using Saccharomyces cerevisiae BY4742. *Biomass Bioenergy, 47*, 1-7.

Llano-Restrepo, M., & Aguilar-Arias, J. (2003). Modeling and simulation of saline extractive distillation columns for the production of absolute ethanol. *Computer and Chemical Engineering, 27*(4), 527-549

Manivannan, A., & Narendhirakannan, R. T. (2015). Bioethanol production from aquatic weed water hycainth (*Eichhornia crassipes*) by yeast fermentation. *Waste Biomass Valor, 6*(2), 209-216.

Meirelles, A., Weiss, S., & Herfurth, H. (1992). Ethanol dehydration by extractive distillation. *Journal of Chemical Technology and Biotechnology, 53*(2), 181-188.

Myers, R. H., & Montgomery, D. C. (1995). *Response surface methodology.* New York, NY: Wiley.

Ngema, P. T. (2010). *Separation processes for high purity ethanol production* (Master's thesis). Chemical Engineering Department, Durban University of Technology.

Nikzad, M., Movagharnejad, K., Najafpour, G., & Talebnia, F. (2012). Comparative studies on the effect of pretreatment of rice husk on enzymatic digestibility and bioethanol production. *IJE Transactions B: Applications, 26*, 455-464.

Nitayavardhana, S., Shrestha, P., Rasmussen, M. L., Lamsal, B. P., Van Leeuwen, J. H., & Khanal, S. K. (2010). Ultrasound improved ethanol fermentation from cassava chips in cassava-based ethanol plants. *Bioresource Technology, 101*(8), 2741-2747.

Offeman, R. D., Stephenson, S. R., Franqui, D., Cline, F. L., Robertson, G. H., & Orts, W. J. (2008). Extraction of ethanol with higher alcohol solvents and their toxicity to yeast. *Separation and Purification Technology, 63*(2), 444-451.

Seider, W. D., Seader, J. D., & Lewin, D. R. (1999). *Process design principles—synthesis, analysis, and evaluation.* New York, NY: John Wiley & Sons. Inc.

Smejkal, Q., & Soos, M. (2002). Comparison of computer simulation of reactive distillation using ASPEN PLUS and HYSYS software. *Chemical Engineering and Processing: Process Intensification, 41*(5), 413-418.

Won, K. Y., Kim, Y. S., & Oh, K. K. (2012). Comparison of bioethanol production of simultaneous saccharification and fermentation and separation hydrolysis and fermentation from cellulose-rich barley straw. *Korean Journal of Chemical Engineering, 29*(10), 1341-1346.

The Study of Time Lag on Plant Growth Under the Effect of Toxic Metal: A Mathematical Model

Kalra, P. and Kumar, P.*

Department of Mathematics, Lovely Professional University, Punjab, India

ABSTRACT

In this paper, a mathematical model is proposed for analytical study of plant growth subjected to the effect of toxic metal. The associated state variables are plant biomass, concentration of nutrients and concentration of toxic metal in the soil. The assumption is that the toxic metals present in the soil interferes with nutrient availability and hence, adversely affect the plant's growth. This effect is studied by introducing the time-lag (delay) in consumption and utilisation coefficient due to less availability of nutrients in the presence of toxic metal. The inclusion of delay disturbed the stability of the interior equilibrium of the system and Hopf bifurcation occurred at the critical value of delay parameter. Further, the direction, stability and period of these bifurcating periodic solutions are also studied as well as sensitivity analysis of state variables with respect to model parameters. Numerical simulation support analytical results using MATLAB.

Keywords: Concentration of nutrients, delay, Hopf bifurcation, plant biomass, toxic metal, simulation

INTRODUCTION

Plants need carbon, hydrogen, and oxygen, water and other nutrients which come from soil for growth. Nutrients are components in food that an organism uses to survive and grow. Plant-soil interaction means the mechanism in which the plants take essential nutrients from the soil through their roots which leads to growth of plants. A low concentration of heavy metals is necessary for growth of plants, but excess of these metals adversely affect the soil quality and hence retards plant growth. Thornley (1976) is the first to apply mathematical modelling to wide range of subjects in plant physiology to predict effect of factors such as temperature, humidity, radiation input and concentration of CO_2 on process rates of respiration, photosynthesis, transpiration, fluid transport and stomatal

responses. Lacointe (2000) reported that that models suggested by Thornley were designed for a particular plant species and under particular conditions, so they could not be applicable to a wider range of conditions. Deleo et al. (1993) gave a simple model that coupled the effect of toxic metal and soil chemistry to study the adverse effect of toxic metal on biomass of trees. Guala et al. (2010, 2013) further modified the parameters of the model given by Deleo to show the model was applicable to not only trees, but all plants in general. Misra and Kalra (2012) studied how the toxicity of heavy metals could adversely affect the growth of a plant using a two-compartment mathematical model. Rouches theorem (1960) explains distribution of roots of exponential polynomials. Ruan and Wei (2001) used Rouches theorem for the discussion of distribution of roots of exponential polynomials. Kubiaczyk and Saker (2002) studied stability and oscillations in system of non-linear delay differential equations of population dynamics. Ruan and Wei (2003) used Rouches theorem for the discussion of distribution of roots of exponential polynomials for study of stability involving delays. Naresh et al (2014) studied the effect of toxicant on plant biomass with time delays. Shukla et al (1996) studied the effect of environmentally degraded soil by rain water and wind on crop yield. Sikarwar (2012) studied the effect of time delay on the dynamics of a multi team prey predator system. Naresh et al. (2006) studied the effect of an intermediate toxic product formed by uptake of a toxicant on plant biomass. Huang et al. (2016) studied analysis for global stability of system of non-linear delay differential equations involving population growth. Zhang et al. (2009) studied the distribution of the roots of a fifth-degree exponential polynomial with applications to a delayed neural network model. The explicit formulae is derived for determining the properties of the Hopf-bifurcation at the critical value using the normal form theory and manifold reduction (Hassard et al.,1981). Bocharov and Rihan (2000) came up with adjoint and direct methods for sensitivity analysis in numerical modelling in biosciences using delay differential equations. Rihan's (2003) sensitivity analysis for dynamic systems with time-lags using adjoint equations and direct methods when the parameters appearing in the model showed they were not only constants but also variables of time. Banks, Robbins and Sutton (2012) presented theoretical foundations for traditional sensitivity and generalised sensitivity functions for a general class of nonlinear delay differential equations. Theoretical results for sensitivity are presented with respect to the delays. Ingalls, Mincheva and Russel (2017) developed a parametric sensitivity analysis for periodic solutions of delay differential equations. Kalra and Kumar (2017) studied the role of time lag in plant growth dynamics using a two-compartment mathematical model. Over the last decade, a lot of work has been done in the field of plant soil interaction under the effect of toxic metals. Till date, delay differential equations have not been prominently used in the field of soil-plant dynamics and agriculture. In view of the above, a mathematical model is proposed for the study of plant growth by introducing delay parameter in containing consumption and utilisation coefficient and complex behaviour giving rise to Hopf bifurcation.

MATHEMATICAL MODEL

The plant growth dynamics is governed by the following system of non-linear delay differential equations involving three state variables: Concentration of nutrients N in the plant, amount of plant biomass W and concentration of toxic metal M in the soil .

$$\frac{dN}{dt} = (K_N - K_{NM}M) - \alpha N(t-\tau)W - \delta_1 N \qquad [1]$$

$$\frac{dW}{dt} = \beta N(t-\tau)W - \delta_2 W \qquad [2]$$

$$\frac{dM}{dt} = I - \gamma NM - \delta_3 M \qquad [3]$$

With initial conditions $N(0)>0$, $W(0)>0$, $M(0)>0$ for all t and $N(t-\tau)=$ constant for $t \in [-\tau, 0]$.

The parameters defined are : K_N is the availability of total nutrients and (K_N-K_{NM} M) is the supply of nutrients hindered due to presence of toxic metal. α is the consumption coefficient. β is the utilisation coefficient for nutrients. γ is the depletion of M due to interaction between M and N. I is the intake of toxic metal M in the soil. δ_1, δ_2, δ_3 are natural decay rates of N, W and M respectively. Here, all the parameters α, β, γ, K_N, K_{NM}, I, δ_1, δ_2, δ_3 are taken as positive.

BOUNDEDNESS

The boundedness of solutions of the model given by (1) -(3) is given by the lemma stated below:

Lemma 1. The model has all its solution lying in the region $C = \left[(N, W, M) \in R_+{}^3 : 0 \le N + W \le \frac{(\delta_3 K_N - IK_{NM})}{\delta_3 \varphi}, 0 \le M \le \frac{I}{\delta_3}\right]$, as t→∞, for all positive initial values $\{N(0), W(0), M(0), N(t-\tau)=$Constant for all $t \in [-\tau, 0]\} \in C \subset R_+{}^3$, where $\varphi = \min(\alpha, \beta, \delta_1, \delta_2)$.

Proof: Consider the following function: $B(t) = N(t) + W(t)$

$$\frac{dB(t)}{dt} = \frac{d}{dt}[N(t) + W(t)]$$

Using Equations (1) -(2) and $\varphi = \min(\alpha, \beta, \delta_1, \delta_2)$ assuming that $N_R(t) \approx N_R(t-\tau)$ as $t \to \infty$ $\quad \frac{dW(t)}{dt} \le \left(K_N - \frac{IK_{NM}}{\delta_3}\right) - \varphi B(t).$

Applying the comparison theorem, as $t \to \infty$ $\quad B(t) \le \frac{(\delta_3 K_N - IK_{NM})}{\delta_3 \varphi}$

So, $0 \le N(t) + W(t) \le \frac{(\delta_3 K_N - IK_{NM})}{\delta_3 \varphi}$

From equation (3): $\frac{dM}{dt} = I - \gamma MN - \delta_3 M$

$\frac{dM}{dt} \leq I - \delta_3 M$, then by usual comparison theorem, when $t \to \infty$: $M \leq \frac{I}{\delta_3}$

So, $0 \leq M \leq \frac{I}{\delta_3}$

POSITIVITY OF SOLUTIONS

Positivity means that the system sustains. For positive solutions, one needs to show that all solution of system given by Equations. (1)– (3), where initial condition is $N(0) > 0, W(0) > 0, M(0) > 0$ for all t and $N(t - \tau)$ = constant for $t \in [-\tau, 0]$, the solution $(N(t), W(t), M(t))$ of the model stays positive for all time $t > 0$.

From equation (3): $\frac{dM}{dt} \geq -(\gamma N + \delta_3)M$ i.e. $\frac{dM}{dt} \geq -\left(\frac{\gamma(\delta_3 K_N - IK_{NM}) + \delta_3{}^2 \varphi}{\delta_3 \varphi}\right)$

$M \geq c_1 e^{-\left(\frac{\gamma(\delta_3 K_N - IK_{NM}) + \delta_3{}^2 \varphi}{\delta_3 \varphi}\right)t}$, here c_1 is constant of integration. So, $M > 0$ for all t.

Similar argument holds for N and W.

INTERIOR EQUILIBRIUM OF MODEL

We calculate an interior equilibrium E^* of model. The system of equations (1) -(3) has one feasible equilibrium E^* (N^*,W^*,M^*) where $N^* = \frac{\delta_2}{\beta}$, $W^* = \left(\frac{K_N(\beta \delta_3 + \gamma \delta_2) - K_{NM}\beta I}{(\beta \delta_3 + \gamma \delta_2)}\right) - \frac{\delta_1}{\alpha}$, $M^* = \frac{\beta I}{(\beta \delta_3 + \gamma \delta_2)}$.

Theorem 1. Consider the exponential polynomial:

$f\left(\lambda, e^{-\lambda \tau_1}, \dots, e^{-\lambda \tau_m}\right) = \lambda^n + P_1{}^0 \lambda^{n-1} + \cdots + P_{n-1}{}^0 \lambda^n + P_n{}^0 + [P_1{}^1 \lambda^{n-1} +$

$\cdots + P_{n-1}{}^1 \lambda^n + P_n{}^1]e^{-\lambda \tau_1} + \cdots + [P_1{}^m \lambda^{n-1} + \cdots + P_{n-1}{}^m \lambda^n + P_n{}^m]e^{-\lambda \tau_m}$,

Where $\tau_i \geq 0$ (i = 0,1,2,…,m) and $P_j{}^i$ (i = 0,1,2,…m :j = 1,2,…n) are constants. As $(\tau_1, \tau_2, \dots, \tau_m)$ vary, the sum of the orders of the zeros of exponential polynomial f(λ,$^{e-\lambda \tau 1}$,….,$e^{-\lambda \tau m}$) on the open right half plane can change only if a zero appears on or crosses the imaginary axis.

This result has been proved by Ruan and Wei [10,12] by using Rouches theorem.

STUDY OF INTERIOR EQUILIBRIUM AND LOCAL HOPF-BIFURCATION

Here, the dynamic behaviour of the interior equilibrium points E^* (N^*, W^*, M^*) of the model given by (1) -(3) is analysed. The exponential characteristic equation about equilibrium E^* is given by:

$$\lambda^3 + m_1 \lambda^2 + m_2 \lambda + m_3 + (n_1 \lambda^2 + n_2 \lambda + n_3)e^{-\lambda \tau} = 0 \qquad [4]$$

Where $m_1 = (\delta_1 + \delta_2 + \delta_3 + \gamma N^*), m_2 = (\delta_1\delta_2 + \delta_2\delta_3 + \delta_3\delta_1 + \delta_1\gamma N^* + K_{NM}\gamma M^*)$,

$$m_3 = (\delta_1\delta_2\delta_3 + \delta_2 K_{NM}\gamma M^*), n_1 = \alpha W^*, n_2 =$$
$$\alpha W^*(\delta_1 + \delta_3 + \gamma N^*), n_3 = \delta_2\delta_3\alpha W^*.$$

Clearly $m_1, m_2, m_3, n_1, n_2, n_3$ are all positive.

Clearly $\lambda = i\omega$ is a solution of equation (4) if and only if

$$(i\omega)^3 + m_1(i\omega)^2 + m_2(i\omega) + m_3 + (n_1(i\omega)^2 + n_2(i\omega) + n_3)e^{-i\omega\tau} = 0 \qquad [5]$$

Separating real and imaginary parts:

$$m_3 - m_1\omega^2 + (n_3 - n_1\omega^2)\cos\omega\tau + n_2\omega\sin\omega\tau = 0 \qquad [6]$$

$$m_2\omega - \omega^3 + n_2\omega\cos\omega\tau - (n_3 - n_1\omega^2)\sin\omega\tau = 0 \qquad [7]$$

W h i c h g i v e s : $\omega^6 + (m_1{}^2 - n_1{}^2 - 2m_2)\omega^4 + (m_2{}^2 - n_2{}^2 + 2n_1n_3 - 2m_1m_3)\omega^2 + (m_3{}^2 - n_3{}^2) = 0$ [8]

Let a $= a = (m_1{}^2 - n_1{}^2 - 2m_2), b = (m_2{}^2 - n_2{}^2 + 2n_1n_3 - 2m_1m_3), c = (m_3{}^2 - n_3{}^2).$

Let $\omega^2 = y$, then equation (8) becomes: $y^3 + ay^2 + by + c = 0.$ [9]

Claim 1. If $c<0$, Equation (9) contains at least one positive real root.

Proof. Let $h(y) = y^3 + ay^2 + by + c$

Here $h(0) = c < 0$, $\lim_{y\to\infty} h(y) = \infty$ o, $\exists$ $y_0 \in (0, \infty)$ such that $h(y_0)=0$. Proof completed.

Claim 2. If $c \geq 0$, then necessary condition for equation (9) to have positive real roots is $D = a^2 - 3b \geq 0$.

Proof. Since $h(y)=y^3+ay^2+by+c$, therefore $h'(y)=3y^2+2ay+b$

$h'(y)=0$ implies $3y^2+2ay+b=0$ [10]

The roots of equation (10) can be expressed as $y_{1,2} = \frac{-2a \mp \sqrt{4a^2 - 12b}}{6} = \frac{-a \mp \sqrt{D}}{3}$ [11]

If $D<0$, then equation (10) does not have any real roots. So, the function $h(y)$ is monotone increasing function in y. It follows from $h(0)=c\geq0$ that equation (9) has no positive real roots.

Clearly if $D\geq0$, then $y_1 = \frac{-a+\sqrt{D}}{3}$ is local minima of $h(y)$. Thus the following claim.

Claim 3. If $c\geq0$, then equation (9) has positive roots if and only if $y_1>0$ and $h(y_1)\leq0$.

Proof. The sufficiency is obvious. Only necessity needs to be proved. Otherwise, assume that either $y_1\leq0$ or $y_1>0$ and $h(y_1)>0$ .If $y_1\leq0$, since $h(y)$ is increasing for $y\geq y_1$ and $h(0)=c\geq0$, it follows that $h(y)$ has no positive real zeros. If $y_1>0$ and $h(y_1)>0$, since $y_2 = \frac{-a-\sqrt{D}}{3}$ is the local maxima value, it follows that $h(y_1)\leq h(y_2)$. Hence, $h(0)=c\geq0$, As $h(y)$ does not have positive real roots.

Lemma 2. Suppose y_1 is defined by equation (11).

(I)   If $c<0$, Equation (9) contains at least one positive real root.

(II) If $c\geq0$ and $D=a^2-3b<0$, then equation (9) has no positive roots.

(III)If $c\geq0$, then equation (9) has positive roots if and only if $y_1>0$ and $h(y_1)\leq0$. **Proof.** Suppose that equation (9) has positive roots. Without loss of generality, assume that it has three positive roots, denoted by y_1,y_2,y_3. Then equation (8) has three positive roots, say $\omega_1 = \sqrt{y_1}$, $\omega_2 = \sqrt{y_2}$, $\omega_3 = \sqrt{y_3}$.

From (7) $\sin\omega\tau = \frac{m_2\omega-\omega^3}{d}$ Which gives $\tau = \frac{1}{\omega}\left[sin^{-1}\left(\frac{m_2\omega-\omega^3}{d}\right) + 2(j-1)\pi\right]$; $j = 1,2,3,-$

Let $\tau_k{}^{(j)} = \frac{1}{\omega_k}\left[sin^{-1}\left(\frac{m_2\omega_k-\omega_k{}^3}{d}\right) + 2(j-1)\pi\right]$; $k = 1,2,3.$; $j = 0,1,2,----$

Then is a pair of purely imaginary roots of equation (8)

Where $\tau = \tau_k{}^{(j)}, k = 1,2,3.$; $j = 0,1,2,---,$ $\lim_{j\to\infty}\tau_k{}^{(j)} = \infty, k = 1,2,3,4.$

Thus, define $\tau_0 = \tau_{k_0}{}^{(j_0)} = \min_{1\leq k\leq3,j\geq1}\left[\tau_k{}^{(j)}\right]$, $\omega_0 = \omega_{k_0}$, $y_0 = y_{k_0}$ [12]

Lemma 3. Suppose that $m_1>0,(m_3+d)>0,m_1 m_2-(m_3+d)>0$.

(I)   If $c\geq0$ and $D=a^2-3b<0$, then all the roots of equation (4) have negative real parts for all $\tau\geq0$.

(II) If $c<0$ or $c\geq0$, $y_1>0$ and $h(y_1)\leq0$, then all the roots of equation (4) have negative real parts for all $\tau \in [0,\tau_0)$.

Proof. When $\tau=0$, equation (4) becomes

$$\lambda^3 + (m_1 + n_1)\lambda^2 + (m_2 + n_2)\lambda + (m_3 + n_3) = 0$$ [13]

Based Routh-Hurwitz's criteria, **(H1):** All roots of equation (5) have negative real parts if and only if $(m_3+n_3)>0,(m_1+n_1)(m_2+n_2)-(m_3+n_3)>0$.

If $c \geq 0$ and $D=a^2-3b<0$, Lemma 2 (II) shows that equation (4) has no roots with zero real part for all $\tau \geq 0$. When $c<0$ or $c \geq 0$, $y_1>0$ and $h(y_1) \leq 0$, Lemma 2 (I) and (III) imply that when $\tau \neq \tau_k^{(j)}$, $k=1,2,3.;j \geq 1$, equation (4) has no roots with zero real part and τ_0 is the minimum value of τ so that the equation(4) has purely imaginary roots. Applying theorem 1, the conclusion of the lemma is obtained.

$$\text{Let } \lambda(\tau) = \psi(\tau) + i\omega(\tau) \tag{14}$$

be the roots of equation (4) satisfying: $\psi(\tau_0) = 0, \ \omega(\tau_0) = \omega_0$

In order to guarantee that $\mp\omega_0$ are simple purely imaginary roots of equation (4), with $\tau=\tau_0$ and $\lambda(\tau)$ satisfies transversality condition, assume that $h'(y_0) \neq 0$.

Lemma 4. Suppose $y_0=\omega_0^2$. If $\tau=\tau_0$, Then Sign $[\psi'(\tau_0)]$ ==Sign $[h'(y_0)]$

Proof. Putting $\lambda(\tau)$ in equation (4) and differentiating *w.r.t* τ, it follows that

$$\frac{d\lambda}{d\tau}\left[3\lambda^2 + 2m_1\lambda + m_2 + \left((n_1\lambda^2 + n_2\lambda + n_3)(-\tau) + (2n_1\lambda + n_2)\right)e^{-\lambda\tau}\right] =$$

$$\lambda(n_1\lambda^2 + n_2\lambda + n_3)e^{-\lambda\tau}$$

Then $\left(\frac{d\lambda}{d\tau}\right)^{-1} = \frac{(3\lambda^2+2m_1\lambda+m_2)e^{\lambda\tau}}{\lambda(n_1\lambda^2+n_2\lambda+n_3)} + \frac{(2n_1\lambda+n_2)}{\lambda(n_1\lambda^2+n_2\lambda+n_3)} - \frac{\tau}{\lambda}$

From equations (6) -(8):

$$\mu'(\tau_0) = Re\left[\frac{(3\lambda^2+2m_1\lambda+m_2)e^{\lambda\tau}}{\lambda(n_1\lambda^2+n_2\lambda+n_3)}\right] + Re\left[\frac{(2n_1\lambda+n_2)}{\lambda(n_1\lambda^2+n_2\lambda+n_3)}\right] \qquad =\frac{1}{\Delta}[3\omega_0^6 +$$

$$2a\omega_0^4 + b\omega_0^2]$$

Where $\Delta= [(n_3 - n_1\omega^2)^2 + (n_2\omega)^2]$. Here $\Delta >0$ and $\omega_0>0$.

It is concluded that Sign $[\psi'(\tau_0)]$=Sign $[h'(y_0)]$.This proves the lemma.

DIRECTION AND STABILITY OF HOPF-BIFURCATING SOLUTION

In the previous section, a family of periodic solutions is obtained that bifurcate from the positive steady state E^* at the critical values of τ. It is also worthwhile to determine the direction, stability and period of these bifurcating periodic solutions. In this section, an explicit formula will be derived to determining the properties of the Hopf-bifurcation at the critical value τ_j, using the normal form theory and manifold reduction due to Hassard, Kazarion and Wan (1981).

Let $u_1 = N - N^*, u_2 = W - W^*, u_3 = M - M^*$ and normalising the delay τ by time scaling $t \to \frac{t}{\tau}$, equations (1) -(3) are transformed into

$$\frac{du_1}{dt} = -\delta_1 u_1 - K_{NM} u_3 - \alpha W^* u_1(t-1) - \alpha u_1(t-1)u_2 \qquad [15]$$

$$\frac{du_2}{dt} = -\delta_2 u_2 + \beta W^* u_1(t-1) + \beta u_1(t-1)u_2 \qquad [16]$$

$$\frac{du_3}{dt} = -\gamma M^* u_1 - (\gamma N^* + \delta_3)u_3 - \gamma u_1 u_3 \qquad [17]$$

Thus, work can be done in phase $C = C\left((-1,0), R_+^3\right).$ Without loss of generality, denote the critical value τ_j by τ_0. Let $\tau = \tau_0 + \mu$, then $\mu = 0$ is a Hopf-bifurcation value of the system given by equations (17) -(19). For the simplicity of notations, rewrite this system as

$$u'(t) = L_\mu(u_t) + F(\mu, u_t) \qquad [18]$$

Where $u(t) = (u_1(t), u_2(t), u_3(t))^T \in R^3, u_t(\theta) \in C$ is defined by $u_t(\theta) = u_t(t+\theta),$ and

$L_\mu : C \to R, \quad F : R \times C \to R$ are given, respectively by

$$L_\mu \emptyset =$$

$$(\tau_0 + \mu)\begin{bmatrix} -\delta_1 & 0 & -K_{NM} \\ 0 & -\delta_2 & 0 \\ -\gamma M^* & 0 & -(\gamma N^* + \delta_3) \end{bmatrix}\begin{bmatrix} \emptyset_1(0) \\ \emptyset_2(0) \\ \emptyset_3(0) \end{bmatrix} +$$

$$(\tau_0 + \mu)\begin{bmatrix} -\alpha W^* & 0 & 0 \\ \beta W^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \emptyset_1(-1) \\ \emptyset_2(-1) \\ \emptyset_3(-1) \end{bmatrix}$$

And $F(\mu, \emptyset) = (\tau_0 + \mu)\begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}$ respectively where $F_1 = -\alpha\emptyset_1(-1)\emptyset_2(0),$

$F_2 = \beta\emptyset_1(-1)\emptyset_2(0), F_3 = -\gamma\emptyset_1(0)\emptyset_3(0)$

$$\emptyset(\theta) = \left(\emptyset_1(\theta), \emptyset_2(\theta), \emptyset_3(\theta)\right)^T \in C\left((-1,0), R\right).$$

Based on the Riesz representation theorem, there exist a function $\eta(\theta,\mu)$ of bounded variation for $\theta \in [-1,0],$, such that $L_\mu \emptyset = \int_{-1}^0 d\,\eta(\theta, 0)\emptyset(\theta)$ for $\emptyset \in C.$

$$\eta(\theta, \mu) =$$

Instead, choose $(\tau_0 + \mu) \begin{bmatrix} -\delta_1 & 0 & -K_{NM} \\ 0 & -\delta_2 & 0 \\ -\gamma M^* & 0 & -(\gamma N^* + \delta_3) \end{bmatrix} \delta(\theta) +$

$$(\tau_0 + \mu) \begin{bmatrix} -\alpha W^* & 0 & 0 \\ \beta W^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \delta(\theta + 1)$$

Here δ is the Dirac delta function. For $\emptyset \in C([-1,0], R_+{}^3)$, define

$$A(\mu)\emptyset = \begin{cases} \dfrac{d\emptyset(\theta)}{d\theta}, & \theta \in [-1,0) \\ \int_{-1}^0 d\,\eta(\theta, 0)\emptyset(\theta), & \theta = 0. \end{cases}$$ And $R(\mu)\emptyset =$

$$\begin{cases} 0, & \theta \in [-1,0) \\ F(\mu, \emptyset) & \theta = 0. \end{cases}$$

Then the system (18) is equivalent to

$$u'(t) = A(\mu)\emptyset + R(\mu)u_t \qquad [19]$$

For $\psi \in C^1([-1,0], R_+{}^3)$, define

$$A^*\psi(s) = \begin{cases} -\dfrac{d\psi(s)}{ds}, & s \in [-1,0) \\ \int_{-1}^0 d\,\eta^T(-t, 0)\psi(-t), & s = 0. \end{cases}$$ And bilinear inner product

$$< \psi(s), \emptyset(\theta) > = \overline{\psi(0)}\emptyset(0) - \int_{-1}^0 \int_{\xi=\theta}^{\theta} \overline{\psi}(\xi - \theta)d\eta(\theta)\phi(\xi)\,d\xi \qquad [20]$$

Sine A^* and $A=A(0)$ are adjoint operators and $i\omega_0$ are eigen values of $A(0)$. Thus they are eigen values of A^*. Suppose that $q(\theta) = q(0)e^{i\omega_0\theta}$ is an eigen vector of $A(0)$ corresponding to the eigen value $i\omega_0$. Then $A(0) = i\omega_0\, q(\theta)$. When $\theta = 0$,

$$\left[i\omega_0 I - \int_{-1}^0 d\eta(\theta)e^{i\omega_0\theta} \right] q(0) = 0, \text{ which yields } q(0) = (1, \sigma_1, \rho_1)^T \text{ where}$$

$\sigma_1 = \dfrac{(\alpha W^* + \delta_1 + i\omega_0)}{K_{NM}}$ and $\rho_1 = \dfrac{\beta W^*(\delta_2 - i\omega_0)}{\delta_2{}^2 + \omega_0{}^2}$

Similarly, it can be verified that $q^*(s) = D(1, \sigma_2, \rho_2)e^{i\omega_0\tau_0 s}$ is the eigen value of A^* corresponding to $-i\omega_0$, where

$\sigma_1 = \dfrac{(\alpha W^* + \delta_1 - i\omega_0)}{K_{NM}}$ and $\rho_1 = \dfrac{\beta W^*(\delta_2 + i\omega_0)}{\delta_2{}^2 + \omega_0{}^2}$

In order to assure $<q^*(s), q(\theta)> = 1$, the value of D needs to be determined.

From equation (22), $<q^*(s), q(\theta)>$

$$= \overline{D}(1, \overline{\sigma_2}, \overline{\rho_2})(1, \sigma_1, \rho_1)^T - $$

$$\int_{-1}^{0} \int_{\xi=0}^{\theta} \overline{D}(1, \overline{\sigma_2}, \overline{\rho_2}) e^{-i\omega_0 \tau_0 (\xi-\theta)} d\eta(\theta)(1, \sigma_1, \rho_1)^T e^{i\omega_0 \tau_0} d\xi$$

$$= \overline{D}\left\{1 + \sigma_1\overline{\sigma_2} + \rho_1\overline{\rho_2} - \int_{-1}^{0}(1, \overline{\sigma_2}, \overline{\rho_2}) \theta e^{i\omega_0 \tau_0 \theta}(1, \sigma_1, \rho_1)^T\right\}$$

$$= \overline{D}\{1 + \sigma_1\overline{\sigma_2} + \rho_1\overline{\rho_2} + \tau_0\overline{\sigma_2}W^*(\beta\rho_1 - \alpha\sigma_1)e^{i\omega_0 \tau_0}\}$$

Hence, choose $\overline{D} = \frac{1}{(1+\sigma_1\overline{\sigma_2}+\rho_1\overline{\rho_2}+\tau_0\overline{\sigma_2}W^*(\beta\rho_1-\alpha\sigma_1)e^{i\omega_0\tau_0})}$

such that $< q^*(s), q(\theta) > = 1, < q^*(s), \overline{q(\theta)} > = 0$.

Following the algorithm given by Hassard et al., (1981) and using the same notations as there to compute the coordinates describing the centre manifold C_0 at $\mu=0$. Let u_t be the solution of equation (18) with $\mu=0$. Define

$$z(t) = < q^*(s), u_t(\theta) >, \quad W(t, \theta) = u_t(\theta) - 2Re(z(t)q(\theta)) \qquad [21]$$

On the centre manifold C_0, $W(t, \theta) = W\left(z(t), \overline{z(t)}, \theta\right)$

Where $W(z, \overline{z}, \theta) = W_{20}(\theta)\frac{z^2}{2} + W_{11}(\theta)z\overline{z} + W_{02}(\theta)\frac{\overline{z}^2}{2} + \cdots,$

z and $\overline{z}$ are local coordinates for centre manifold C_0 in the direction of q^* and $\overline{q^*}$. Note that W is real if u_t is real. Consider only real solution. For solution $u_t \in C_0$ of equation (20), since $\mu=0$,

$$z'(t) = i\omega_0\tau_0 z + < \overline{q^*}(\theta), F\left(0, W(z, \overline{z}, \theta) + 2Re\left(z(t)q(\theta)\right)\right) >$$

$$= i\omega_0\tau_0 z + \overline{q^*}(0) F\left(0, W(z, \overline{z}, 0) + 2Re\left(z(t)q(\theta)\right)\right)$$

$$\equiv i\omega_0\tau_0 z + \overline{q^*}(0)F_0(z, \overline{z})$$

Rewrite this equation as

$$z'(t) = i\omega_0\tau_0 z(t) + g(z, \overline{z}) \qquad [22]$$

Where $g(z, \overline{z}) = \overline{q^*}(0)F_0(z, \overline{z}) = g_{20}(\theta)\frac{z^2}{2} + g_{11}(\theta)z\overline{z} + g_{02}(\theta)\frac{\overline{z}^2}{2} +$

$$g_{21}(\theta)\frac{z^2\overline{z}}{2} + \cdots \qquad (23)$$

As $u_t(\theta) = (u_{1t}, u_{2t}, u_{3t}) = W(t, \theta) + z\, q(\theta) + \bar{z}\overline{q(\theta)}$

$(1, \sigma_1, \rho_1)^T e^{i\omega_0\tau_0\theta}$, so

$$u_{1t}(0) = z + \bar{z} + W_{20}{}^{(1)}(0)\frac{z^2}{2} + W_{11}{}^{(1)}(0)\, z\bar{z} + W_{02}{}^{(1)}(0)\frac{\bar{z}^2}{2} + \cdots,$$

$$u_{2t}(0) = \sigma_1 z + \overline{\sigma_1}\,\bar{z} + W_{20}{}^{(2)}(0)\frac{z^2}{2} + W_{11}{}^{(2)}(0)\, z\bar{z} + W_{02}{}^{(2)}(0)\frac{\bar{z}^2}{2} + \cdots,$$

$$u_{3t}(0) = \rho_{1_1} z + \overline{\rho_{1_1}}\,\bar{z} + W_{20}{}^{(3)}(0)\frac{z^2}{2} + W_{11}{}^{(3)}(0)\, z\bar{z} + W_{02}{}^{(3)}(0)\frac{\bar{z}^2}{2} + \cdots,$$

$$u_{1t}(-1) = ze^{-i\omega_0\tau_0} + \bar{z}e^{i\omega_0\tau_0} + W_{20}{}^{(1)}(-1)\frac{z^2}{2} + W_{11}{}^{(1)}(-1)\, z\bar{z} +$$

$$W_{02}{}^{(1)}(-1)\frac{\bar{z}^2}{2} + \cdots,$$

$$u_{2t}(-1) = \sigma_1 e^{-i\omega_0\tau_0} z + \overline{\sigma_1}\, e^{i\omega_0\tau_0}\,\bar{z} + W_{20}{}^{(2)}(-1)\frac{z^2}{2} + W_{11}{}^{(2)}(-1)\, z\bar{z} +$$

$$W_{02}{}^{(2)}(-1)\frac{\bar{z}^2}{2} + \cdots,$$

Thus, comparison of coefficients with equation (23) gives:

$$g_{20} = \overline{D}(1, \sigma_1, \rho_1)f_{z^2},\ g_{02} = \overline{D}(1, \overline{\sigma_1}, \overline{\rho_2})f_{\bar{z}^2},$$

$$g_{11} = \overline{D}(1, \overline{\sigma_1}, \overline{\rho_2})f_{z\bar{z}}, g_{21} = \overline{D}(1, \overline{\sigma_1}, \overline{\rho_2})f_{z^2\bar{z}}$$

In order to determine g_{21}, focus needs to be on computation of $W_{20}(\theta)$ and $W_{11}(\theta)$. From equations (19) and (21):

$$W' = u_t{}' - z'q - \bar{z}'q = \begin{cases} AW - 2Re\big[\overline{q^*}(0)F_0 q(\theta)\big], & \theta \in [-1,0) \\ AW - 2Re\big[\overline{q^*}(0)F_0 q(0)\big] + F_0, & \theta = 0 \end{cases}$$

Let $W' = AW + H(z, \bar{z}, \theta)$ [24]

Where $H(z, \bar{z}, \theta) = H_{20}(\theta)\frac{z^2}{2} + H_{11}(\theta)z\bar{z} + H_{02}(\theta)\frac{\bar{z}^2}{2} + H_{21}(\theta)\frac{z^2\bar{z}}{2} + \cdots,$ [25]

On the other hand, on C_0 near the origin $W' = W_z z' + W_{\bar{z}}\bar{z}'$

Expanding the above series and computing the coefficients, we get

$$[A - 2i\omega_0 I]W_{20}(\theta) = -H_{20}(\theta),\ AW_{11}(\theta) = -H_{11}(\theta)$$ [26]

By equation (22), for $\theta \in [-1,0)$,

$$H(z,\bar{z},\theta) = -\overline{q^*}(0)\overline{F_0}q(\theta) - \overline{q^*}(0)\overline{F_0}\,\overline{q}(\theta) = -gq(\theta) - \overline{g}\,\overline{q}(\theta)$$

Comparing the coefficients with (23) for $\theta \in [-1,0]$ that

$$H_{20}(\theta) = -g_{20}q(\theta) - \overline{g_{02}}\,\overline{q}(\theta), \ \ H_{11}(\theta) = -g_{11}q(\theta) - \overline{g_{11}}\,\overline{q}(\theta).$$

From equations (22), (25) and the definition of A we obtain

$$W_{20}(\theta) = 2i\omega_0\tau_0 W_{20}(\theta) + g_{20}q(\theta) + \overline{g_{02}}\,\overline{q}(\theta)$$

Solving for $W_{20}\,(\theta)$:

$$W_{20}(\theta) = \frac{ig_{20}}{\omega_0\tau_0}q(0)e^{i\omega_0\tau_0\theta} + \frac{i\overline{g_{02}}}{3\omega_0\tau_0}\overline{q}(0)e^{-i\omega_0\tau_0\theta} + E_1 e^{2i\omega_0\tau_0\theta},$$

And similarly

$$W_{11}(\theta) = \frac{-ig_{11}}{\omega_0\tau_0}q(0)e^{i\omega_0\tau_0\theta} + \frac{i\overline{g_{11}}}{\omega_0\tau_0}\overline{q}(0)e^{-i\omega_0\tau_0\theta} + E_2$$

Where E_1 and E_2 are both three dimensional vectors, and can be determined by setting $\theta=0$ in H. In fact since $H(z,\bar{z},\theta) = -2Re\left[\overline{q^*}(0)F_0 q(0)\right] + F_0$,

So

$$H_{20}(\theta) = -g_{20}q(\theta) - \overline{g_{02}}\,\overline{q}(\theta) + F_{z^2},$$

$$H_{11}(\theta) = -g_{11}q(\theta) - \overline{g_{11}}\,\overline{q}(\theta) + F_{z\bar{z}}$$

Where $F_0 = F_{z^2}\frac{z^2}{2} + F_{z\bar{z}}z\bar{z} + F_{\bar{z}^2}\frac{\bar{z}^2}{2} + \cdots$

Hence combining the definition of A,

$$\int_{-1}^{0} d\,\eta(\theta)W_{20}(\theta) = 2i\omega_0\tau_0 W_{20}(0) + g_{20}q(0) + \overline{g_{02}}\,\overline{q}(0) - F_{z^2} \text{ and}$$

$$\int_{-1}^{0} d\,\eta(\theta)W_{11}(\theta) = g_{11}q(0) - \overline{g_{11}}\,\overline{q}(0) - F_{z\bar{z}}$$

Notice that

$$\left[i\omega_0\tau_0 I - \int_{-1}^{0} e^{i\omega_0\tau_0\theta} d\eta(\theta)\right]q(0) = 0 \text{ and}$$

$$\left[-i\omega_0\tau_0 I - \int_{-1}^{0} e^{-i\omega_0\tau_0\theta} d\eta(\theta)\right]\overline{q}(0) = 0$$

implies

$$\left[2i\omega_0\tau_0 I - \int_{-1}^{0} e^{2i\omega_0\tau_0\theta}\,d\eta(\theta)\right]E_1 = F_{z^2} \ \text{ and } \ -\left[\int_{-1}^{0} d\eta(\theta)\right]E_2 = F_{z\bar{z}}$$

Hence,

$$\begin{bmatrix} \left(2i\omega_0 + \delta_1 + \alpha W^* e^{-2i\omega_0\tau_0}\right) & 0 & -K_{NM} \\ \beta W^* e^{-2i\omega_0\tau_0} & (2i\omega_0 + \delta_2) & 0 \\ \gamma M^* & 0 & (2i\omega_0 + \gamma N^* + \delta_3) \end{bmatrix} E_1 =$$

$$-2\begin{bmatrix} \alpha\sigma_1 e^{-i\omega_0\tau_0\theta} \\ -\beta\sigma_1 e^{-i\omega_0\tau_0\theta} \\ -\gamma\rho_1 \end{bmatrix} \text{ and }$$

$$\begin{bmatrix} (\delta_1 + \alpha W^*) & 0 & -K_{NM} \\ \beta W^* & \delta_2 & 0 \\ \gamma M^* & 0 & (\gamma N^* + \delta_3) \end{bmatrix} E_2 = -2\begin{bmatrix} \alpha Re\{\sigma_1\}e^{i\omega_0\tau_0\theta} \\ -\beta Re\{\sigma_1\}e^{i\omega_0\tau_0\theta} \\ -\gamma Re\{\rho_1\} \end{bmatrix}$$

Thus g_{21} can be expressed by the parameters.

Based on the above analysis, each g_{ij} can be determined by the parameters. Thus, following quantities can be computed:

$$C_1(0) = \frac{i}{2\omega_0\tau_0}\left(g_{11}g_{20} - 2|g_{11}|^2 - \frac{|g_{02}|^2}{3}\right) + \frac{g_{21}}{2} \quad , \quad \mu_2 = -\frac{Re\{C_1(0)\}}{Re\{\lambda'(\tau_0)\}},$$

$$\beta_2 = 2Re\{C_1(0)\}, \hspace{6cm} [27]$$

$$T_2 = -\frac{Im\{C_1(0)\} + \mu_2 Im\{\lambda'(\tau_0)\}}{\omega_0\tau_0}$$

Theorem 2. The value of μ_2 determines the direction of the Hopf bifurcation: if $\mu_2>0(\mu_2<0)$, then the Hopf bifurcation is supercritical (subcritical) and the bifurcating periodic solutions exist for $\tau>\tau_0$ ($\tau<\tau_0$). The value of β_2 determines the stability of bifurcating solutions: the bifurcating periodic solutions are orbitally asymptotically stable (unstable) if $\beta_2<0$ ($\beta_2>0$). The value of T_2 determines the period of the bifurcating periodic solutions: the period increases (decreases) if $T_2 >0$ ($T_2 <0$).

SENSITIVITY ANALYSIS OF STATE VARIABLES WITH RESPECT TO MODEL PARAMETERS

In this paper, the model includes constant parameters. The 'Direct Method' to estimate the general sensitivity coefficients is used. The direct method is based on considering all parameters as constants and then the sensitivity coefficients are estimated by solving sensitivity equations simultaneously with the original system. If all the parameters (α,β,γ) appearing in the system model (1)–(3) are taken to be constants, then sensitivity analysis, in this case, may just entail

finding the partial derivatives of the solution with respect to each parameter. As an illustration if consider parameter β, then partial derivatives of the solution (N,W,M) with respect to β give rise to following set of sensitivity equations:

$$\frac{dS_1}{dt} = -\delta_1 S_1 - \alpha N(t-\tau)S_2 - K_{NM}S_3 - \alpha W S_1(t-\tau) \qquad [28]$$

$$\frac{dS_2}{dt} = (\beta N(t-\tau) - \delta_2)S_2 + \beta W S_1(t-\tau) + W N(t-\tau) \qquad [29]$$

$$\frac{dS_3}{dt} = -\gamma M S_1 - (\gamma N + \delta_3)S_3 \qquad [30]$$

Where $S_1 = \frac{\partial N}{\partial \beta}, S_2 = \frac{\partial W}{\partial \beta}, S_3 = \frac{\partial M}{\partial \beta}$

Then, this system of sensitivity equations (28) – (30) along with the original system of equations (1) – (3) is solved to estimate the sensitivity of the state variables (N,W,M) to the parameter β. The similar procedure and argument holds for estimating the sensitivity of the state variables with respect to the parameters α and γ.

Sensitivity of Variables to Parameter α

As shown by Figures 1 and Figure 2, the parameter consumption coefficient α does not lead to much of variation and change in the values of state variables nutrient concentration N and concentration of toxic metal M which ultimately remain stable and tens to zero, as we decrease the values of α from $\alpha=0.9$ to $\alpha=0.5$. It predicts the lesser sensitivity of state variables N and M to the parameter α. However, for the same range of values of α, the state variable amount of plant biomass W undergoes under considerable change as shown by Figure 3. It shows increase in the rate of plant biomass with decrease in the delayed value of consumption coefficient. It remains stable as well.
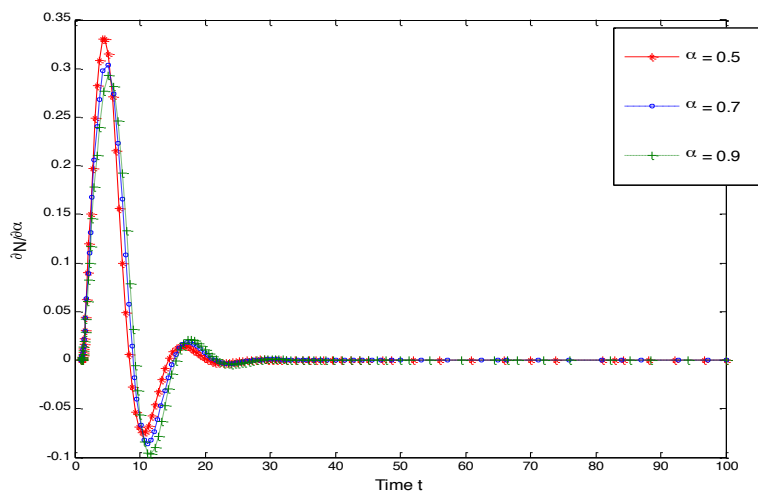
Figure 1. Time series graph between partial changes in nutrient concentration *N* for different values of consumption coefficient α
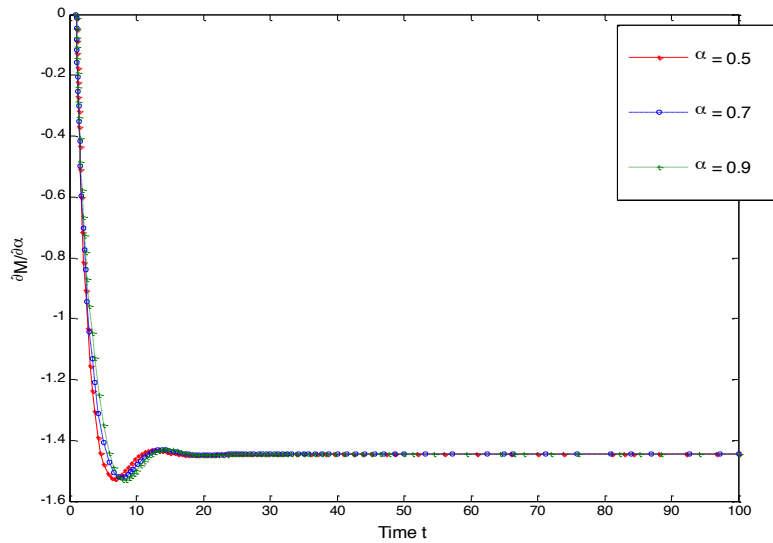
Figure 2. Time series graph between partial changes in concentration of toxic metal M for different values of consumption coefficient α
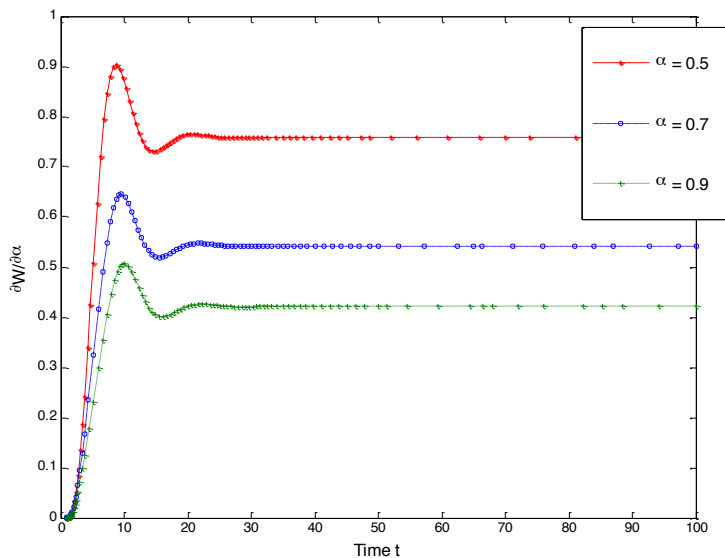
Figure 3. Time series graph between partial changes in plant biomass W for different values of consumption coefficient α

Sensitivity of Variables to Parameter β

Initially, the rate of nutrient concentration starts losing stability with decrease in value of utilisation coefficient, but finally becomes stable and tends to be zero as we decrease the values of parameter utilization coefficient β from β=0.7 to β=0.3 as shown by Figure 4. Figure 5 shows the increase in rate of concentration of toxic metal M with decrease in value of utilisation

coefficient β from β=0.7 to β=0.3. It starts losing stability as well. Decrease in the rate of plant biomass W with decreased value of delayed utilisation coefficient is shown in Figure 6. It also starts losing stability.
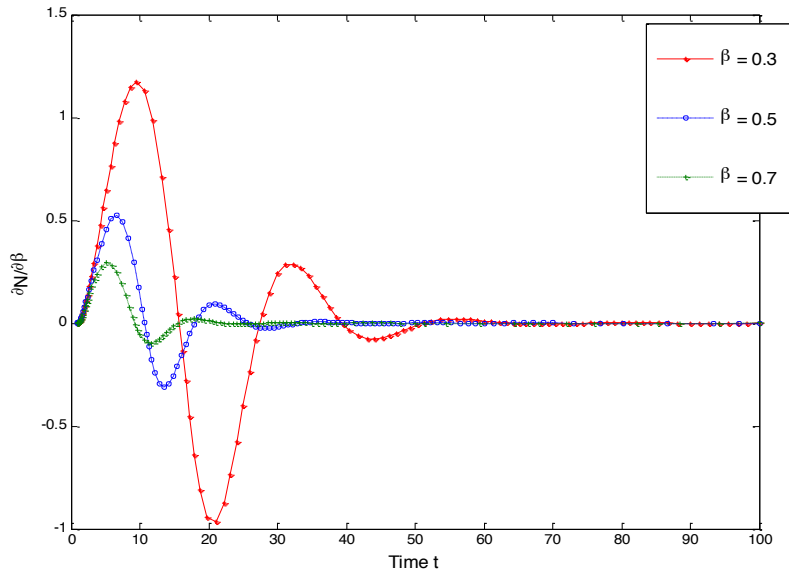
Figure 4. Time series graph between partial changes in nutrient concentration *N* for different values of utilisation coefficient *β*
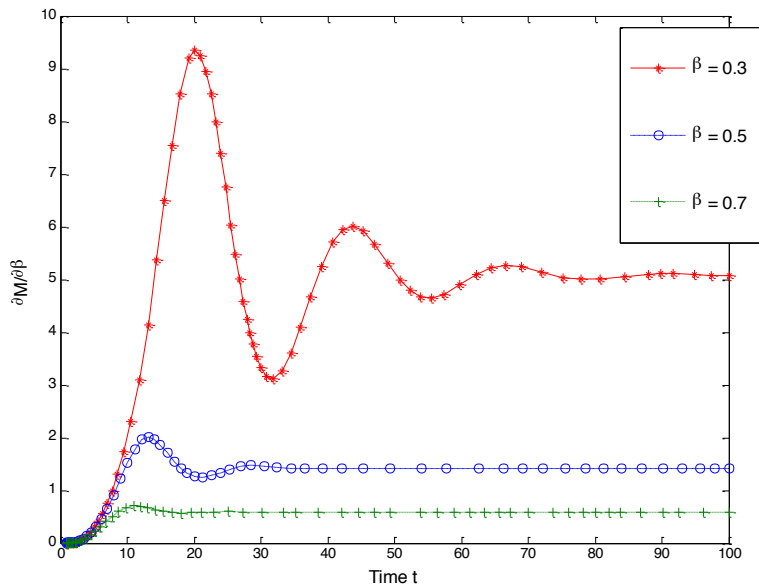
Figure 5. Time series graph between partial changes in concentration of toxic metal *M* for different values of utilisation coefficient *β*
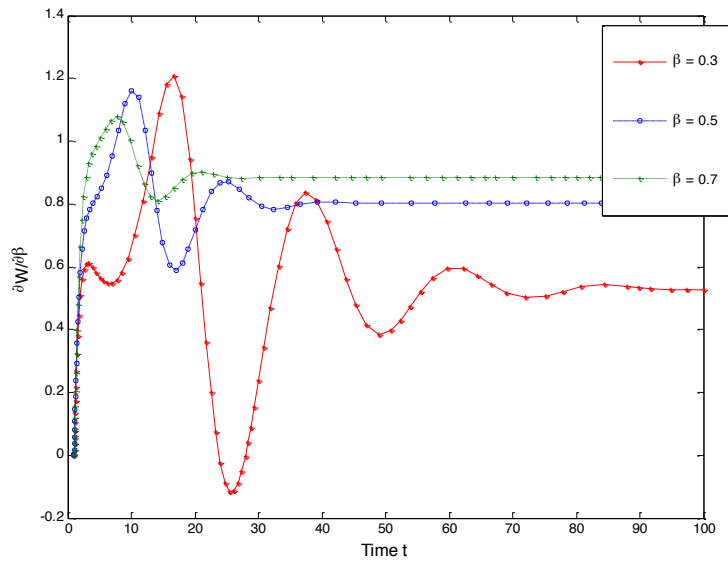
Figure 6. Time series graph between partial changes in plant biomass *W* for different values of utilisation coefficient β

Sensitivity of Variables to Parameter γ

Figure 7 shows nutrient concentration *n* is not very much affected by decease in values of depletion coefficient γ from γ=0.2 to γ=0.05. It does not lose stability. Figure 8 shows decrease in rate of concentration of toxic metal with decrease in values of depletion coefficient of toxic metal due to interaction with nutrients. It stays stable. Figure 9 shows increase in rate of plant biomass with decrease in values of depletion coefficient of toxic metal due to interaction with nutrients. It remains stable as well.
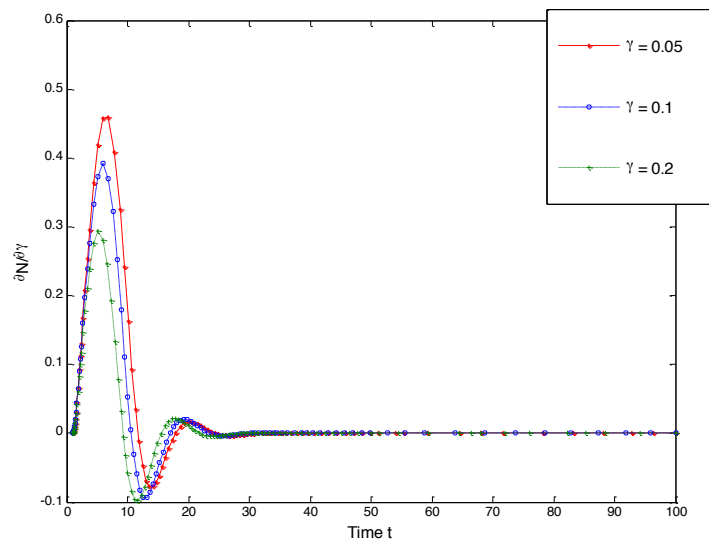
Figure 7. Time series graph between partial changes in nutrient concentration *N* for different values of depletion coefficient γ
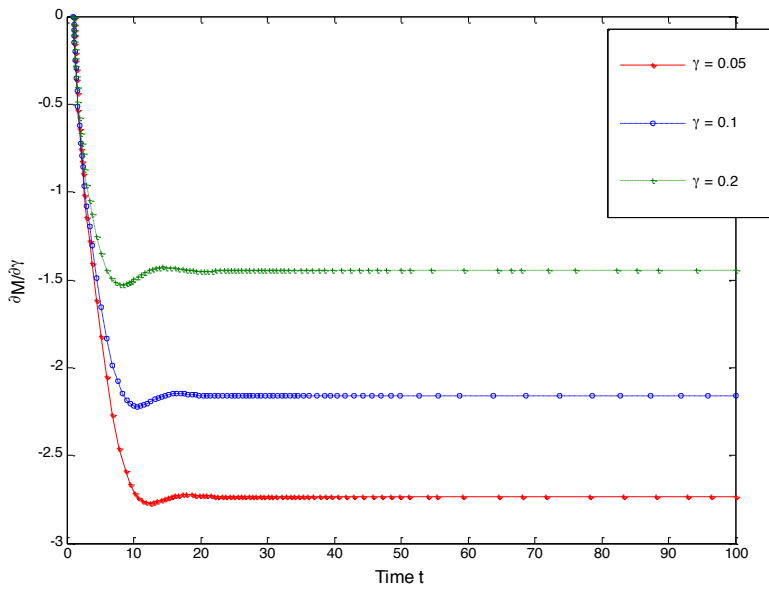
Figure 8. Time series graph between partial changes in concentration of toxic metal *M* for different values of depletion coefficient γ
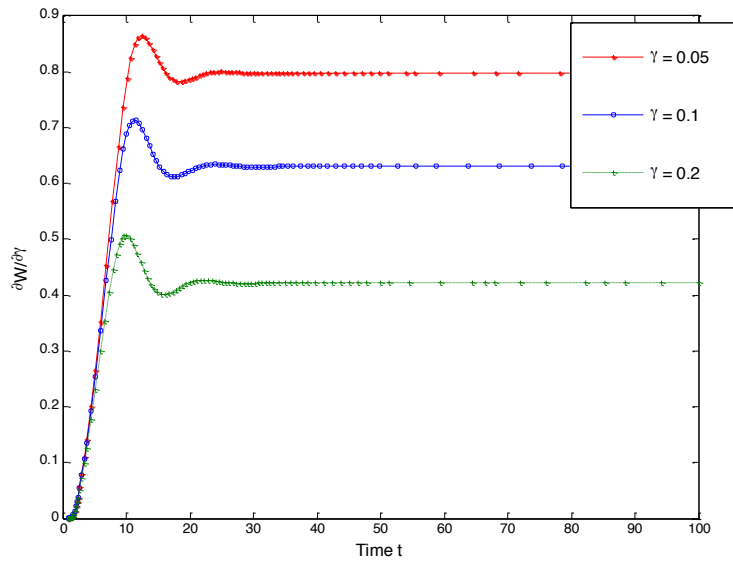
Figure 9. Time series graph between partial changes in plant biomass *W* for different values of depletion coefficient γ

NUMERICAL EXAMPLE

To consolidate the analytical result with the help of a numerical, simulation is done with MATLAB. For the following set of values, the behaviour shown by the system is as follows:

$K_N=1, K_{NM}=0.3, \alpha=0.9, \delta_1=0.2, \beta=0.7, \delta_2=0.8, I=0.5, \gamma=0.2, \delta_3=0.4.$

The behaviour of the system for different values of delay is expressed as:

E_1 ($N^*=1.1426, W^*= 0.5181, M^*= 0.7950$)
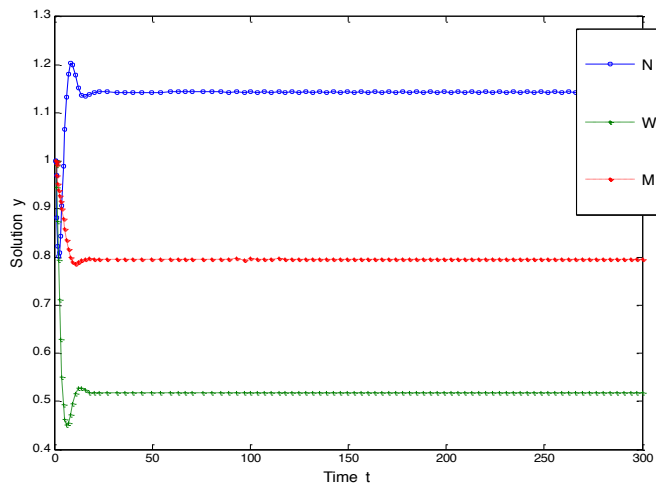
Figure 10. The interior equilibrium point E_1 (1.1426,0.5181,0.7950) of the system is stable when there is no delay that is τ=0
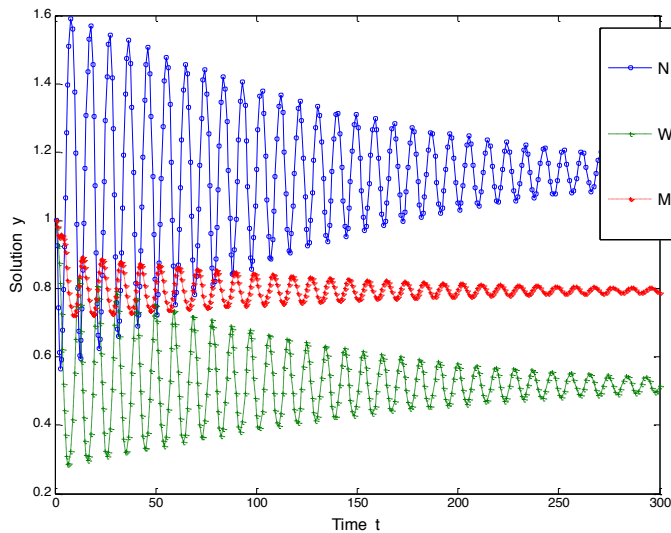
Figure 11. When delay τ<1.373, the interior equilibrium point E_1 (1.1426,0.5181,0.7950) is asymptotically stable
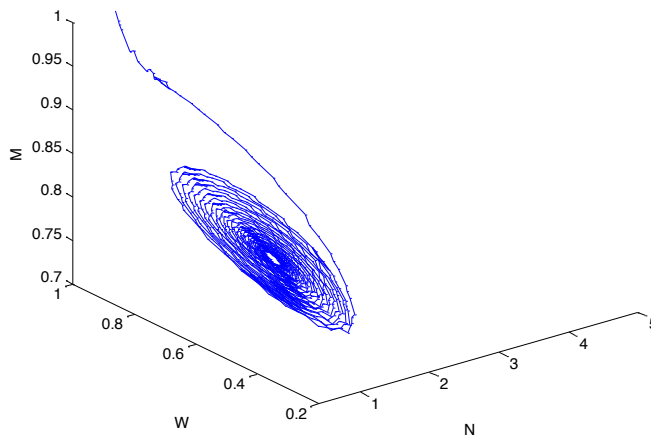
Figure 12. Phase space diagram of Nutrient N, Plant Biomass W and Toxic Metal M when delay τ<1.373
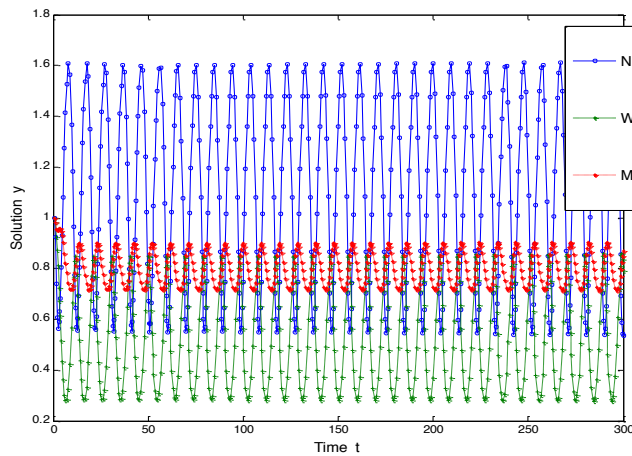
Figure 13. The interior equilibrium point E_1 (1.1426,0.5181,0.7950)losses its stability and Hopf- bifurcation occurs when delay τ≥1.373
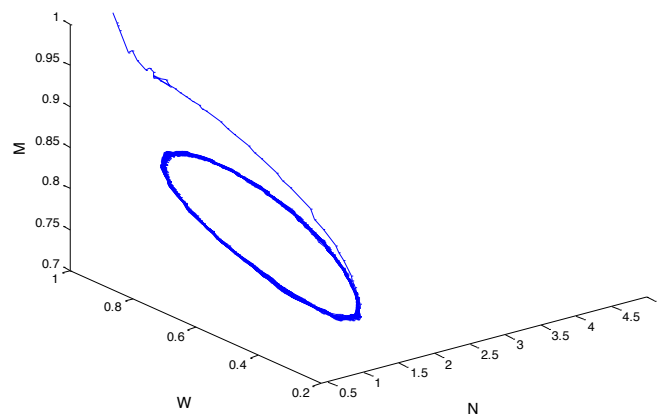
Figure 14. Phase space diagram of Nutrient N, Plant Biomass W and Toxic Metal M when delay τ≥1.373. The bifurcating periodic solution is orbitally, asymptotically stable

CONCLUSION

In this paper, a mathematical model is proposed to study the role of delay on plant growth dynamics under the effect of toxic metal. The stability and Hopf- bifurcation about the interior equilibrium is studied. It has been concluded that when there is no time delay, interior equilibrium E_1 (1.1426,0.5181,0.7950) is completely stable (Figure 10) as proved by lemma 3 using Routh-Hurwitz's criteria. But under the same set of parameters, a critical value of the parameter delay is obtained below which the system is asymptotically stable (Figure 11 and Figure 12) and unstable above that critical value of parameter (Figure 13 and Figure 14) as proved by lemma 2 and lemma 4. While passing through the critical value, the system shows oscillations that is Hopf bifurcation.

In this paper, the sensitivity of model solutions due to perturbing the parameters appearing in delay differential systems is also investigated using direct method. It is shown how the sensitivity functions enable one in identification of specific parameters and improve the understanding of the role played by specific model parameters. The oscillation and change in values accompanied by the sensitivity of state variables to parameters means that the solution is sensitive to changes in the parameter and that parameter plays an important role in the model. Sensitivity analysis reveals that the state variable nutrient concentration N is least sensitive to all parameters (α,β,γ) compared with other two state variables W and M who show considerable amount of change in their rates for different sets of values of the parameters. Rate of plant biomass shows increase with decrease in the delayed value of consumption coefficient and stays stable (Figure 3) and decreases with decrease in delayed value of utilisation coefficient and loses stability (Figure 6).

This theoretical model as well as numerical results show that for a certain threshold of parameters, the system possesses asymptotic stability around positive interior equilibrium. Further from stability analysis and numerical simulation, it is concluded that τ is a bifurcating parameter for which the interior equilibrium point shows stable oscillatory behaviour when $\tau \geq \tau_0$. After considering the effect of time lag in the system, limit cycles appear for interior equilibrium points when time delay crosses some critical value.

In future, the efforts will be made to validate the proposed mathematical model with some existing plant growth data under the effect of toxic metals. The proposed mathematical model dealing with delay in plant soil dynamics under the effect of toxicant will be helpful to farmers, agriculturists, ecologists and scientists to use pesticides, insecticides and chemical fertilizers in an optimal way. The study of the factors due to which the delay is produced, and the components being affected will help the concerned community to plan the remedial measures. Being quantitative in nature, the mathematical model will prove to be economical in terms of time and money being invested on large scale experiments.

REFERENCES

Banks, H. T., Robbins, D., & Sutton, K. L. (2012). *Generalized sensitivity analysis for delay differential equations.* Centre for Research in Scientific Computations. North Carolina State University. Raleigh. NC 27695-8212.

Bocharov, A. G., & Rihan, F. A. (2000). Numerical modelling in biosciences using delay differential equations. *Journal of Computational and Applied Mathematics, 125*(1-2), 183-199.

Deleo, G., Delfuria, L., & Gatto, M. (1993). The interaction between soil acidity and forest dynamics: A simple-model exhibiting catastrophic behaviour. *Theoretical Population Biology, 43*(1), 31-51. doi:10.1006/tpbi.1993.1002

Dieudonne, J. (1960). *Foundations of Modern Analysis.* New York: Academic press.

Guala, S., Vega, F. A., & Covelo, E. F. (2012). Modeling the plant–soil interaction in presence of heavy metal pollution and acidity variations. *Environmental Monitoring and Assessment, 185*(1), 73-80. doi:10.1007/s10661-012-2534-z

Guala, S. D., Vega, F. A., & Covelo, E. F. (2010). The dynamics of heavy metals in plant–soil interactions. *Ecological Modelling, 221*(8), 1148-1152. doi: 10.1016/j.ecolmodel.2010.01.003

Hassard, B. D., Kazarino, N. D., & Wan, Y. H. (1981). *Theory and applications of Hopf bifurcation.* CUP, Archive.

Huang, G., Liu, A., & Foryś, U. (2015). Global stability analysis of some nonlinear delay differential equations in population dynamics. *Journal of Nonlinear Science, 26*(1), 27-41. doi:10.1007/s00332-015-9267-4

Ingalls, B., Mincheva, M., & Roussel, M. R. (2017). Parametric sensitivity analysis of oscillatory delay systems with an application to gene regulations. *Bulletin of Mathematical Biology, 79*(7), 1539-1563. doi:10.1007/s11538-017-0298-x.

Kalra, P., & Kumar, P. (2017, July). Role of delay in plant growth dynamics: A two compartment mathematical model. In *AIP Conference Proceedings* (Vol. 1860, No. 1, p. 020045). AIP Publishing. doi:10.1063/1.4990344.

Kubiaczyk, I., & Saker, S. (2002). Oscillation and stability in nonlinear delay differential equations of population dynamics. *Mathematical and Computer Modelling, 35*(3-4), 295-301. doi:10.1016/s0895-7177(01)00166-2

Lacointe, A. (2000). Carbon allocation among tree organs: A review of basic processes and representation infunctional-structural tree models. *Annals of Forest Science, 57*(5), 521-533. doi:10.1051/forest:2000139

Misra, O. P., & Kalra, P. (2013). Modelling effect of toxic metal on the individual plant growth: A two compartment model. *American Journal of Computational and Applied Mathematics, 2*(6), 276-289. doi: 10.5923/j.ajcam.20120206.06

Misra, O. P., & Kalra, P. (2013). Effect of toxic metal on the structural dry weight of a plant: A model. *International Journal of Biomathematics, 06*(05), 1350028. doi:10.1142/s1793524513500289

Naresh, R., Sharma, D., & Sundar, S. (2014). Modeling the effect of toxicant on plant biomass with time delay. *International Journal of Nonlinear Science, 17*(3), 254-267.

Naresh, R., Sundar, S., & Shukla, J. (2006). Modeling the effect of an intermediate toxic product formed by uptake of a toxicant on plant biomass. *Applied Mathematics and Computation, 182*(1), 151-160. doi: 10.1016/j.amc.2006.03.041

Rihan, F. A. (2003). Sensitivity analysis for dynamic systems with time-lags. *Journal of Computational and Applied Mathematics, 151*(2), 445–462.

Ruan, S., & Wei, J. (2001). On the zeros of a third-degree exponential polynomial with applications to a delayed model for the control of testosterone secretion. *Mathematical Medicine and Biology, 18*(1), 41-52. doi:10.1093/imammb18.1.41

Ruan, S., & Wei, J. (2003). On the zeros of a transcendental function with applications to stability of delay differential equations with two delays. *Dynamics of Continuous Discrete and Impulsive Systems Series A, 10*, 863-874.

Shukla, A., Dubey, B., & Shukla, J. (1996). Effect of environmentally degraded soil on crop yield: The role of conservation. *Ecological Modelling, 86*(2-3), 235-239. doi:10.1016/0304-3800(95)00057-7

Sikarwar, C. S. (2012). *Effect of time delay on the dynamics of a multi team prey predator system.* (Unpublished doctoral dissertation). Jiwaji University, Gwalior, India.

Thornley, J. H. (1976). *Mathematical models in plant physiology. A Quantitative approach to problems in plant and crop physiology.* London: Academic press.

Zhang, T., Jiang, H., & Teng, Z. (2009). On the distribution of the roots of a fifth-degree exponential polynomial with application to a delayed neural network model. *Neurocomputing, 72*(4-6), 1098-1104. doi: 10.1016/j.neucom.2008.03.003.

SCIENCE & TECHNOLOGY

Optimisation of Multireservoir Operation Policy using Teaching-Learning Based Optimisation Algorithm

Jayantilal N. Patel and Pranita N. Balve*

Civil Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, India

ABSTRACT

The multi reservoir water resource system has various purposes and therefore, operation planning is becoming complex and involves a number of decision variables. This paper presents an efficient and reliable teaching-learning based approach, namely teaching-learning based optimisation (TLBO) algorithm for optimisation of multireservoir operation policy. It is based on the teaching-learning process of the education system. TLBO algorithm does not require any algorithm-specific parameters for obtaining optimal results; instead it requires only the population size and number of iterations. The time required for obtaining the specific optimised algorithm parameter is reduced and results are also near the global-optimal solution. Furthermore, the number of function evaluations required is less. This TLBO algorithm is implemented at the five-reservoir model of the upper Godavari river project in the city of Nashik in Maharashtra, India. The efficiency of the results of the TLBO algorithm is compared with the genetic algorithm (GA). The results show that TLBO algorithm is considered to be a viable alternative to the operating policy of multireservoir system and it avoids the local optimal solution.

Keywords: Multireservoir operation, optimisation, TLBO algorithm

INTRODUCTION

Optimisation of reservoir operation is a challenging for water resource planners and managers. The proper utilisation of

available resources becomes necessary to develop strategies to utilise the available resources effectively and efficiently. Due to an increase in population, industrialisation, and urbanisation, the optimum utilisation of available water and proper management is becoming an important task. For optimisation of a complex reservoir system, various traditional and non-traditional techniques are implemented. Linear programming (LP), non-linear programming (NLP), and dynamic programming (DP) are the most common and principal optimisation techniques used in

water resources system analysis. Future direction for research and application of optimisation in reservoir system management and operations is reviewed (Labadie, 2004). Also, the challenges and issues of climate change in optimisation of reservoir is reviewed (Asmadi et al., 2014).

In the case of a problem faced by multiobjective and multireservoir system optimisation, the objective function is complex with a nonlinear relationship in constraints as well as in objective function, thus LP cannot be used. In regard to a problem faced in large optimisation, where a decision is to be taken at each stage and performance of the next step is dependent on the results of the previous step, and in such a case, dynamic programming can be used. However, the use of DP in problems that have more decision variables and complex objective functions and constraints, the problem becomes high dimensional. The problems that involve a non-linear relationship in their objective function and constraints can be handled by NLP, however, it faces the problem of a slow rate of convergence and also takes up a large amount of time for computation as well as a large amount of computational storage (Jyothiprakash & Shanthi, 2006). These methods solve the problem point by point and the solution obtained is also a single optimal solution. Moreover, the solution may lie on the non-convex region of the function space (Laifa & Boudour, 2009). These are some limitations of traditional techniques for solving multiobjective optimisation problems.

To obtain the Pareto optimal solution, first, all the possible Pareto fronts are derived and then the algorithm is solved step by step. Therefore, the solution for the next iteration is improved from the previous iteration. The meta-heuristic techniques like evolutionary algorithm's (EA's) and swarm intelligence techniques are used in solving single or multiobjective, or single or multireservoir system problems. These techniques give a solution using a population in every iteration in a single run. These techniques can solve the problem of local minima by searching the solution in the entire search space using randomised initialisation and stochastic search in their operation process. Many problems cannot reach to the global optimal solution in the parameter optimisation process because difficulties arise in determining the optimal controlling parameters of algorithm for optimisation. The EA's and swarm intelligent-based algorithms are probabilistic algorithms and require common controlling parameters, such as population size, number of iterations, along with common controlling parameters. These algorithms require their own algorithm-specific parameters. The problem with multi decision variables can be solved by using nested stochastic dynamic programming (nSDP) and nested reinforcement learning (nRL) algorithms to overcome the high dimensional problem (Delipetrev et al., 2017). The performance and efficiency of the newly developed metaheuristic bat algorithm is evaluated by solving the reservoir operation optimisation problem (Hadded et al., 2014).

To overcome these difficulties of determining the optimal controlling parameters of algorithms, teaching-learning based optimisation (TLBO) algorithm is introduced (Rao et al., 2011). The TLBO has the capability of determining the best solution with a logical number of population size and corresponding number of iterations. It is a population-based optimisation technique in which the population size is considered to be learners in the teaching-learning process. The most intelligent learner among all learners is considered to be the teacher. The different subjects offered to the students are considered to be decision variables of the optimisation problem. The result from the learners is analogous to the fitness value of the optimisation problem. The TLBO algorithm does not require any algorithm specific parameters

to solve the complex reservoir problems. Due to non-availability of algorithm specific parameters, less efforts are required in obtaining the optimal solution, therefore, algorithm becomes robust and powerful. Currently, TLBO algorithm has gained a wide application in the field of engineering applications for optimisation. Non-domination-based sorting multiobjective power flow problems are solved using the TLBO algorithm for minimisation (Nayak et al., 2012). Togan (2012) used the TLBO algorithm for the design of planning steel frame structures. Furthermore, multiobjective heat exchanger problems are solved using a modified version of the TLBO algorithm (Rao & Patel, 2013). TLBO algorithm is proposed (Zou et al., 2013) for multiobjective optimisation in which a non-dominating concept and computed crowding distance is adopted. For modern machining processes of manufacturing, TLBO is implemented to achieve high quality products (Rao & Kalyankar, 2013). To check the efficiency of TLBO algorithm (Rao et al., 2011), solving the constrained and unconstrained real parameter optimisation problem and performance is compared with other optimisation algorithms. Moreover, to solve large-scale non-linear optimisation problems, TLBO is proposed (Rao et al., 2012) for obtaining the global solution. The performance of the TLBO algorithm is shown in Figure 4.

METHODS

Teaching-Learning Based Optimisation (TLBO) Algorithm

The TLBO algorithm is a comparatively new algorithm (Rao et al., 2011). The performance of the algorithm is based on the teaching-learning phenomenon as seen in a classroom setting. Here, a teacher and learner are considered to be the main components in which from the teacher and from the learner's interaction, they improve the average grade of the class. It works in two phases, teacher phase and student phase. In the teacher phase, the teacher takes efforts to share his or her knowledge to improve the grade of the students. In the student phase, students use the knowledge taught by teacher and they also interact amongst themselves and improve overall knowledge. Student achievement in obtaining the best results is defined in terms of grade by learning well from the teacher (Rao et al., 2011).

TLBO is population-based method and in this optimisation method the number of the population is considered initially. This population number is considered to be the number of students in a class. The population giving the best output includes the teacher from the available students. In a class, a different subject assigned is considered to be a decision variable in the optimisation of the problem. The fitness value of an objective function is analogous to the result of the class. Working of the teacher and student phases is explained (Rao & Kalyankar, 2012).

Teacher Phase

In this phase, students learn from the teacher and the teacher tries to increase the average result of the class from value M_1 to his or her level (i.e. T_A). Practically, the teacher can move the mean result of class room M_1 to another value M_2, which shows improvement in M_1, depending upon the teaching skill of the teacher. Consider M_j as the mean value and T_i is teacher at iteration i.

Thus, T_i teacher tries to improve the mean value M_j towards it and the new mean is denoted as M_{new} and the difference between M_{new} and M_j is given by the equation as:

$$(\text{Difference of mean})_i = r_i (M_{new} - T_F M_j) \tag{1}$$

where, r_i is a random number that varies from [0,1], T_F is the teaching factor that decides the value of the mean difference. The value of T_F is either 1 or 2. Considering equal probability, the T_F value is decided randomly. When T_F is 1, this means there is no increase in the level of knowledge and 2 corresponds to complete transfer of knowledge.

$$T_F = \text{round} [1 + \text{rand} (0,1)\{2\text{-}1\}] \tag{2}$$

The capability of students decides the transfer level of knowledge. By using the difference of the mean value's existing solution is updated using the following expression as,

$$X_{new,i} = X_{old,j} + (\text{Difference of mean})_i \tag{3}$$

Student Phase

In the second phase of the algorithm, students interact between them and increase their knowledge level. After learning from the teacher, students interact randomly with other students and learn new things if the other student has more knowledge. Mathematically, the learning phenomenon of students in this phase is expressed as, two different learners X_i and X_j at any iteration i is given by:

$$X_{new,i} = X_{old,i} + r_i (X_i - X_j) \qquad \text{if } f(X_i) < f(X_j) \tag{4}$$

$$X_{new,i} = X_{old,i} + r_i (X_j - X_i) \qquad \text{if } f(X_i) > f(X_j) \tag{5}$$

In the process of the TLBO algorithm, the solution is updated two times, i.e. in the teacher phase as well as in the learner phase. If any duplicate solution is observed while obtaining the solution, it is updated randomly. Therefore, the total number of function evolution required for TLBO algorithm is [(2 X number of population X number of iteration) + (function evolution require for elimination of duplicate solution].

Application of TLBO Algorithm

In this section, TLBO is implemented to five reservoir systems. The optimal operating policy of the multireservoir system of the Upper Godavari project is considered. In this multireservoir system, Karanjwan, Waghad, Punegaon, Ozarkhed, and Palkhed peak up weir is considered. The main purpose of these reservoirs is to allocate the water for irrigation purposes. At the downstream of the Karanjwan reservoir, an industrial area is developing, thus it is becoming

necessary to provide water for industrial purposes. The schematic representation of the reservoir system is shown in Figure 1. The geological position of the reservoir system is at latitude 20°12'15" N Longitude 73°49'37.93" E in Nasik district, Maharashtra, India. The water is released for irrigation purposes from each reservoir and at the end released from each reservoir to meet the Palkhed peak up weir, which is located at a distance of 24 km downstream from the Karanjwan reservoir. Data of monthly inflow and other required data is collected from the Palkhed Irrigation department, Nashik during the period of 1985-2013. The monthly irrigation demand is calculated using FAO Penman- Monteith method.
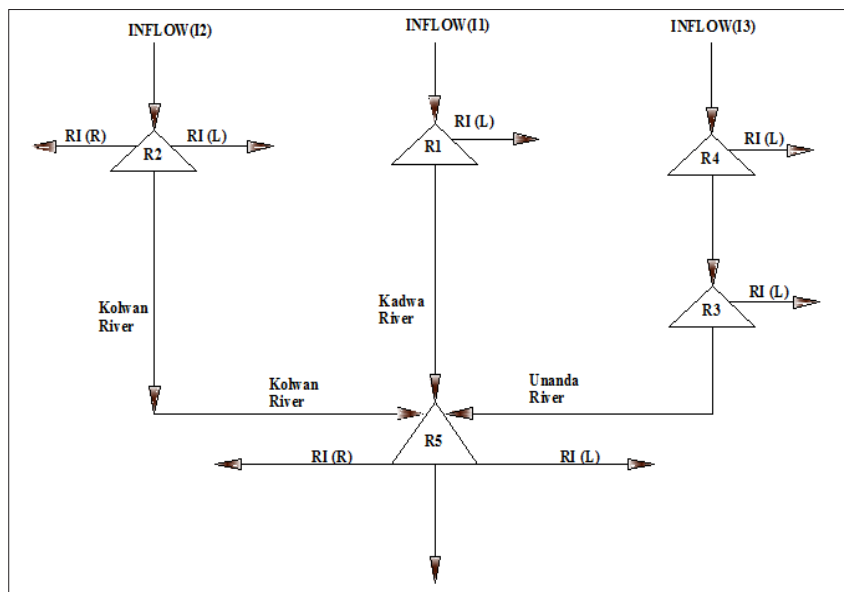
Figure 1. Schematic representation of physical system

A maximum of 90% rainfall occurs during the monsoon season (i.e. June to November). The Karanjwan Dam is constructed on Kadwa River and releases the water for irrigation from the river and from the left bank canal. The Waghad Dam is constructed along the Kolwan River, which has the left and right bank canal for irrigation purposes. Similarly, Punegaon and Ozarkhed reservoirs are in a series on the Unanda River, which is also constructed for irrigation purposes. At the end, releases from all the reservoirs meet the Palkhed peak up weir, which is also has the left and right bank canals that are constructed for irrigation purposes.

Objective Function

In formulating the problem, the main objective of a multireservoir system is to maximise the releases for irrigation. In India, the water year is considered to be from June to the end of May and in this study, the monthly allocation policy is derived.

Objective =

$$\sum_{t=1}^{t=12} RI_{1,t} + \sum_{t=1}^{t=12} RI_{2,t} + \sum_{t=1}^{t=12} RI_{3,t} + \sum_{t=1}^{t=12} RI_{4,t} + \sum_{t=1}^{t=12} RI_{5,t} \qquad [6]$$

where, objective is the maximisation of releases for irrigation. $RI_{1,t}$, $RI_{2,t}$, $RI_{3,t}$, $RI_{4,t}$ and $RI_{5,t}$ are the releases from Karanjwan, Waghad, Punegaon, Ozarkhed, and Palkhed reservoirs respectively in the period of t month in Mm^3.

Irrigation Release Constraint

Releases into the canal and river are lifted from the reservoir for irrigation (RI_t) in all time periods should be less than or equal to the maximum irrigation demand (ID_{max})t to produce the targeted yield. Moreover, the releases should be greater than or equal to minimum irrigation demand (ID_{min})t to sustain the crops for all the time periods. In the present study, minimum irrigation demand is considered to be 30% of the maximum irrigation demand for all the periods.

$8.0006 \leq RI_{1,t} \leq 16.0089$ $\qquad$ Vt = 1,2,3,….,12 $\qquad$ [7]

$31.0764 \leq RI_{2,t} \leq 52.3232$ $\qquad$ Vt = 1,2,3,….,12 $\qquad$ [8]

$10.8320 \leq RI_{3,t} \leq 18.0533$ $\qquad$ Vt = 1,2,3,….,12 $\qquad$ [9]

$9.6442 \leq RI_{4,t} \leq 64.2949$ $\qquad$ Vt = 1,2,3,….,12 $\qquad$ [10]

$54.5793 \leq RI_{5,t} \leq 181.9310$ $\qquad$ Vt = 1,2,3,….,12 $\qquad$ [11]

Storage Capacity Constraint

Reservoir storage should be less than or equal to the maximum storage, i.e. gross storage of the reservoir. Reservoir storage in time period 't' should be greater than or equal to the minimum storage, i.e. dead storage of the reservoir to be maintained in the reservoir for all time periods.

$8.78 \leq S1 \leq 175$ $\qquad$ [12]

$6.48 \leq S_2 \leq 75.1$ $\qquad$ [13]

$2.11 \leq S_3 \leq 17.57$ $\qquad$ [14]

$7.63 \leq S_4 \leq 67.95$ $\qquad$ [15]

$17.7 \leq S_5 \leq 230.1$ $\qquad$ [16]

where, S_1, S_2, S_3, S_4 and S_5 are Karanjwan, Waghad, Punegaon, Ozarkhed, and Palkhed reservoirs respectively in the water year in Mm^3.

Water Balance Equation

This constraint relates to the reservoir storage S(t), inflow I(t), release for irrigation RI(t), and losses due to leakages and evaporation L(t), as well as downstream water requirement DWR(t) for all the time periods.

$$S_1(t+1)=S_1(t)+I_1(t)-RI_1(t)-L_1(t)-DWR_1(t) \qquad \forall t=1,2,3,\ldots,12 \qquad [17]$$

$$S_2(t+1)=S_2(t)+I_2(t)-RI_2(t)-L_2(t)-DWR_2(t) \qquad \forall t=1,2,3,\ldots,12 \qquad [18]$$

$$S_3(t+1)=S_3(t)+I_3(t)-RI_3(t)-L_3(t)-DWR_3(t) \qquad \forall t=1,2,3,\ldots,12 \qquad [19]$$

$$S_4(t+1)=S_4(t)+RI_3(t)-RI_4(t)-L_4(t)-DWR_4(t) \qquad \forall t=1,2,3,\ldots,12 \qquad [20]$$

$$S_5(t+1)=S_5(t)+RI_4(t)+RI_1(t)+RI_2(t)-RI_5(t)-L_5(t)-DWR_5(t) \qquad \forall t=1,2,3,\ldots,12 \qquad [21]$$

The above formulated model is solved using the TLBO algorithm to obtain the best fitness value of the objective function.

RESULTS AND DISCUSSION

The TLBO technique is applied to the model described in the previous section to obtain the operation policy of the multireservoir system. In this study, the decision variables are the releases for irrigation purposes from the reservoirs. These decision variables are evaluated two times to obtain the results. If any constraint is in violation in satisfying the constraints, then a penalty is assigned by a proper penalty coefficient. In this problem, the total number of decision variables are 120 (number of time periods = 12 and number of decisions for each reservoir in each period is 5). The termination criterion is either to reach the maximum number of generations, or if there is no significant improvement in the solution.

The model is evaluated for a combination of the number of population size and number of generations. After satisfying all the constraints and receiving the optimal value of the objective function, the number of function evolutions required is decided. The storage constraint satisfies the condition, storages of all the reservoirs for each time period is obtained and its values are varying between dead storage to the gross storage of the reservoir. The storages obtained from GA and TLBO are shown in Figure 2(a), Figure 2(b), Figure 2(c), Figure 2(d) and Figure 2(e). In Figure 2(a) to Figure 2(e) the comparative performance of the storages for each time period is shown. The storage results of the GA are just satisfying the constraints of the storage of the five reservoirs. As per requirement, GA is not releasing the water at the downstream, therefore, less changes in the reservoir storages. The maximum storages is available in the reservoir. The storages obtained from TLBO algorithm are also satisfying the constraints of the reservoir storages. The storages at the starting of the water year i.e. in monsoon are less. In Rabi season, crops are grown on the stored water, so the storages of the reservoir are reduced.
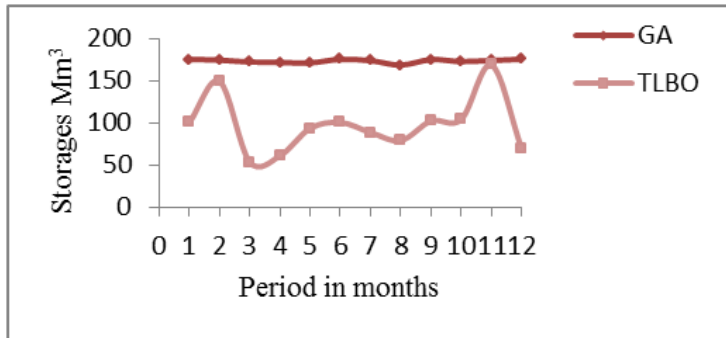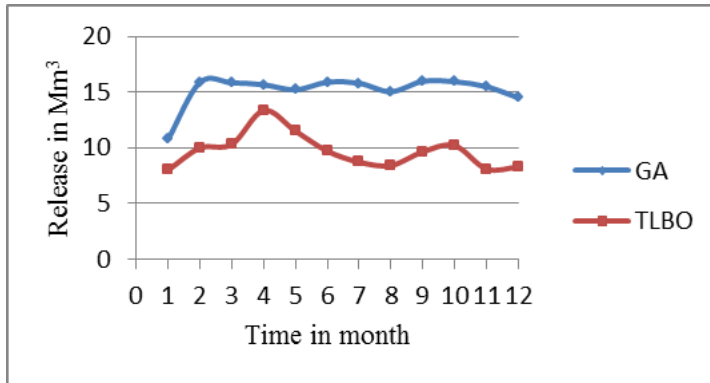
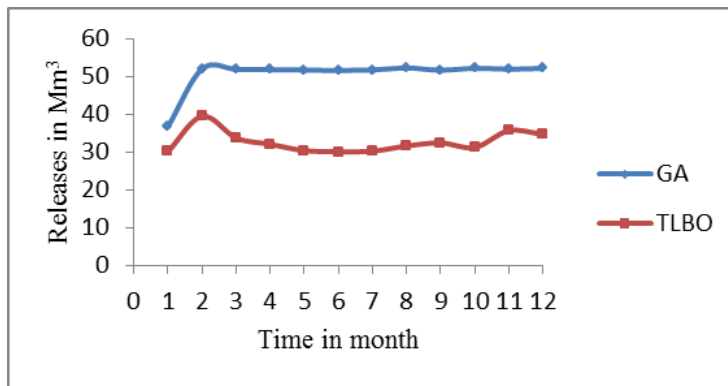*Figure 2(a)*. Karanjwan Reservoir storage from GA and TLBO in time period
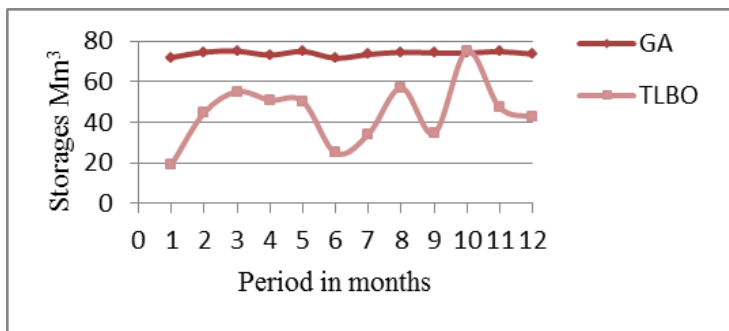
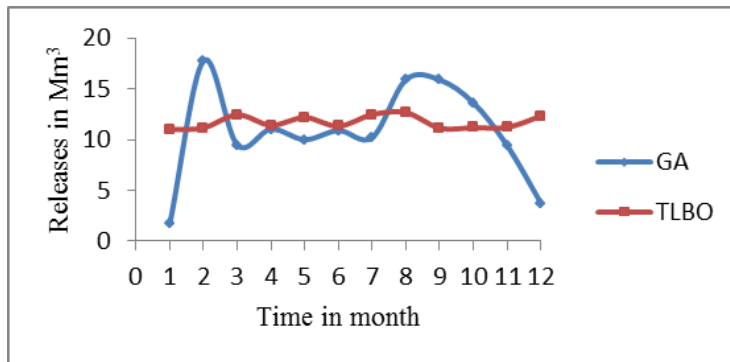*Figure 2(b)*. Waghad Reservoir storage from GA and TLBO in time period
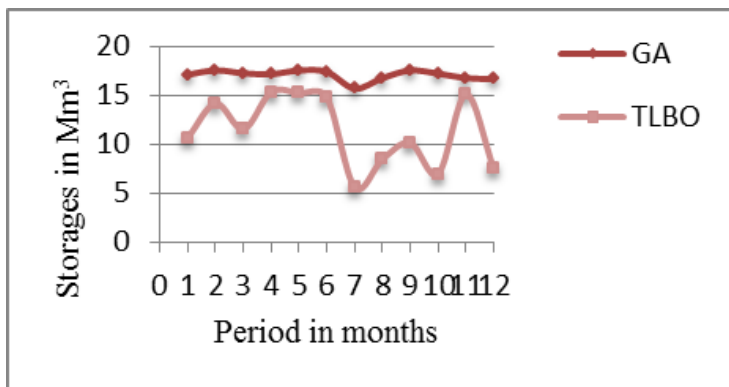
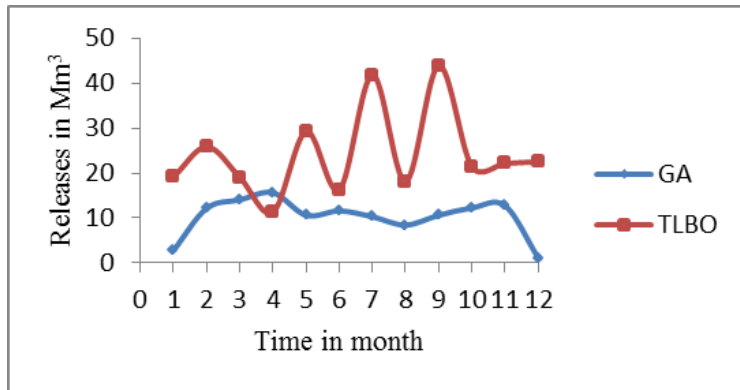*Figure 2(c)*. Punegaon Reservoir storage from GA and TLBO in time period
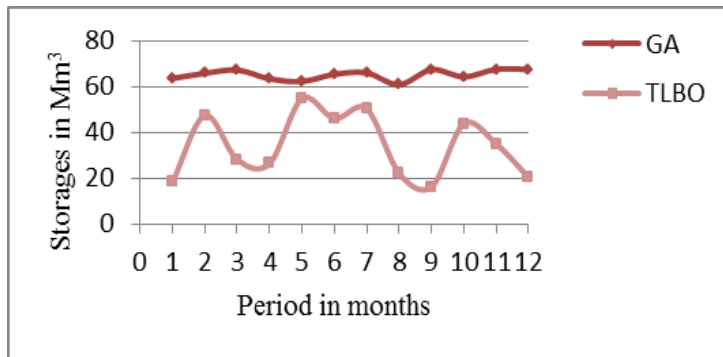
Figure 2(d). Punegaon Reservoir storage from GA and TLBO in time period
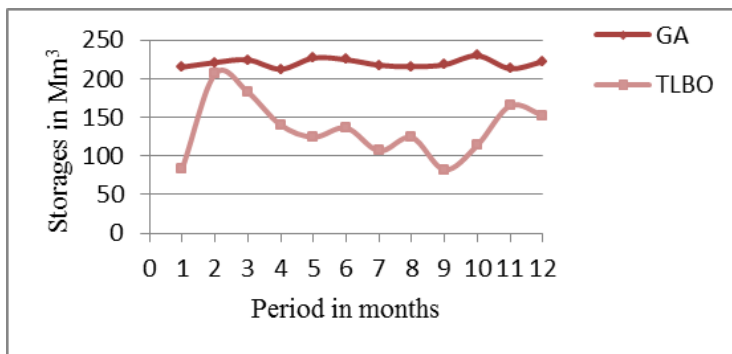
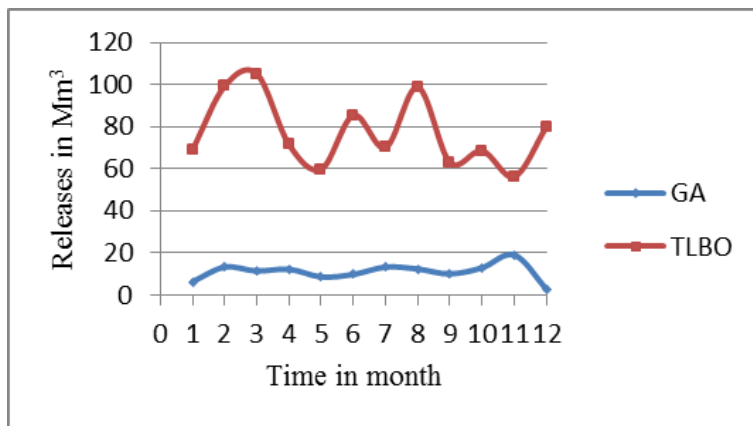*Figure 2(e).* Palkhed Reservoir storage from GA and TLBO in time period

The storages for the five reservoirs for the time period are obtained using GA and TLBO. The TLBO requires a population size of 25 and the number of generations is 40. The results get evaluated twice; therefore, the numbers of evolution functions obtained are 2000. The results of GA are obtained by using the tool box of genetic algorithm in Matlab. In obtaining the solution, the population decided is 5000, creation function is used as the feasible population, and selection function is used as tournament selection. In reproduction, crossover probability is decided as 0.4, mutation is carried out with mutation function as an adaptive feasible, and crossover function is considered to be single point. Furthermore, optimisation is carried out with 50,000 numbers of function evaluations. In both the results of GA and TLBO, storages in all the time periods satisfy the defined constraints.

The releases from all the reservoirs are also determined considering the constraints of releases and irrigation demand. In obtaining the releases from all the reservoirs, they must satisfy the mass balance equations of all the reservoirs. The releases obtained from GA and from TLBO are shown in Figure 3(a), Figure 3(b), Figure 3(c), Figure 3(d) and Figure 3(e).

Figure 3(a). Karanjwan Reservoir releases from GA and TLBO in time period

Figure 3(b). Waghad Reservoir releases from GA and TLBO in time period

Figure 3(c) Punegaon Reservoir releases from GA and TLBO in time period

Figure 3(d). Ozarkhed Reservoir releases from GA and TLBO in time period

Figure 3(e). Palkhed Reservoir releases from GA and TLBO in time period

Operation policy obtained from GA and TLBO is shown in Table I which, for the Karanjwan reservoir releases obtained from GA and TLBO, they showed significantly less difference because they are satisfying the constraints of the irrigation demand and releases. For Waghad reservoir GA showing approximately the same releases in all the months of the water year except the first two months, but TLBO shows the operation policy as per requirement of irrigation and downstream demand. For the Punegaon reservoir, GA shows fluctuation in the releases for the first and last two months, whereas TLBO satisfies the constraints and as per requirement, the release policy is decided. In the Ozarkhed reservoir, releases obtained from GA at the end of the water year and at the starting of the monsoon are less. The Punegaon reservoir releases obtained from GA are much less and it could not satisfy the demand of irrigation and downstream. However, for the same reservoirs, TLBO satisfies the constraints of irrigation demand as well as water balance equation of the reservoir. From the results, it is observed that compared with GA, TLBO gives better results, which satisfies the demand as well as it distributes the water as per the requirements in all the months of the water year.

Table 1
Comparison of the releases for five reservoirs obtained by GA and TLBO

		t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12
R1	GA	10.78	15.87	15.84	15.66	15.21	15.89	15.8	15.04	15.97	15.93	15.49	14.49
	TLBO	8.00	9.98	10.30	13.35	11.51	9.70	8.72	8.41	9.64	10.19	8.10	8.27
R2	GA	36.76	52.01	51.88	51.78	51.70	51.55	51.71	52.21	51.64	52.12	51.90	52.12
	TLBO	30.15	39.47	33.71	32.05	30.38	30.07	30.21	31.60	32.43	31.20	35.66	34.69
R3	GA	1.70	17.74	9.47	11.03	10.00	10.90	10.21	15.90	15.88	13.62	9.41	3.75
	TLBO	10.98	11.16	12.44	11.42	12.17	11.32	12.46	12.67	11.13	11.17	11.21	12.32
R4	GA	2.82	12.16	14.03	15.60	10.60	11.58	10.29	8.44	10.60	12.24	12.75	1.18
	TLBO	19.16	25.84	19.03	11.30	29.23	16.24	41.82	18.07	43.89	21.43	22.13	22.49
R5	GA	6.34	13.26	11.42	12.27	8.64	9.78	13.29	12.37	10.06	12.97	19.03	2.82
	TLBO	68.99	99.31	105.13	71.48	59.76	85.44	70.03	98.84	62.99	68.58	56.28	79.93
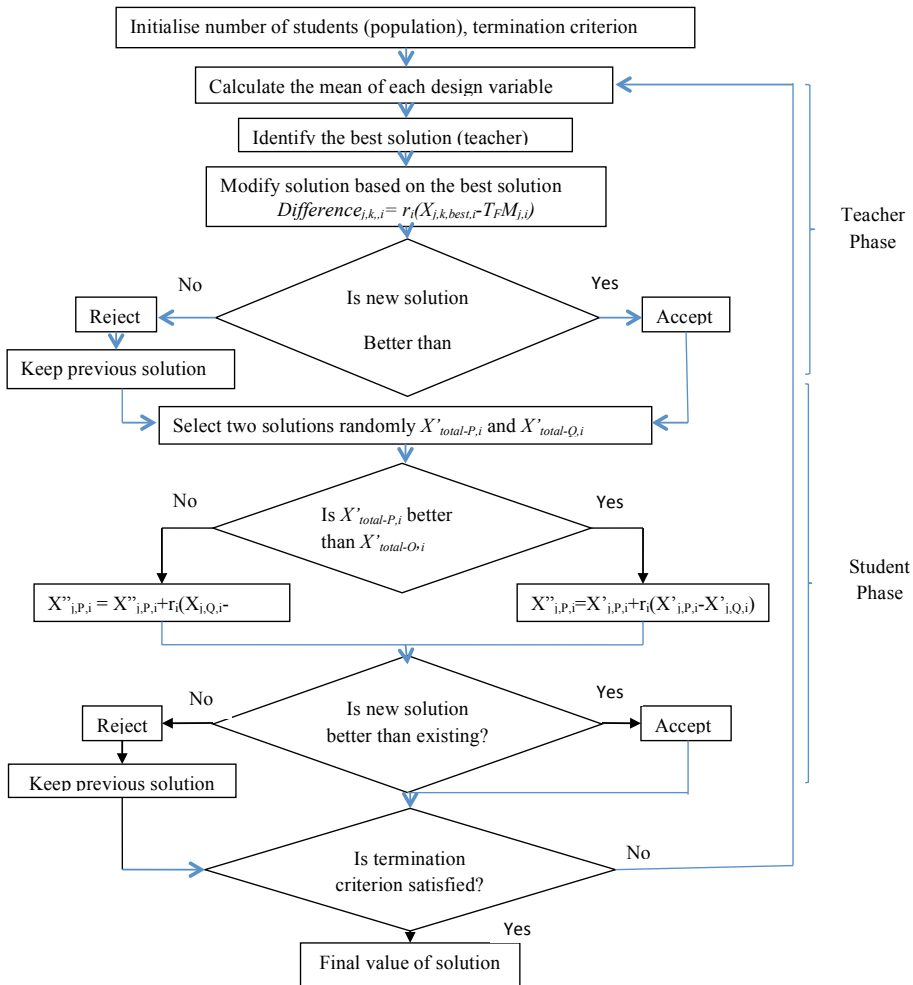
Figure 4. Performance of TLBO algorithm

CONCLUSION

A newly developed teaching-learning based optimisation algorithm is implemented for optimisation of a multireservoir operating policy. The TLBO has the capability to solve problems that contain a large number of decision variables and which are complex in nature. It is used to solve the multireservoir system of the Upper Godavari project. While obtaining the results, TLBO tries for different combinations of population size and number of generations. Because it does not require any algorithm-specific parameters, the time required for the results is also less as compared to GA. TLBO requires less numbers of function evaluations as compared to GA. Therefore, the time is also less. TLBO gives releases as per irrigation and downstream requirements of all reservoirs in each month of water year, whereas for Punegaon, Ozarkhed, and Palkhed, GA shows much less releases and thus does not even satisfy the demand. The GA requires making decisions on selecting the algorithm specific parameters. Therefore, it is concluded the operating policy of TLBO shows practical utility in the field because the results are comparatively reasonable and acceptable.

ACKNOWLEDGMENT

REFERENCES

Ahmad, A., El-Shafie, A., Razali, S. F. M., & Mohamad, Z. S. (2014). Reservoir optimization in water resources: a review. *Water Resources Management, 28*(11), 3391-3405.

Delipetrev, B., Jonoski, A., & Solomatine, D. P. (2017). A novel nested stochastic dynamic programming (nSDP) and nested reinforcement learning (nRL) algorithm for multipurpose reservoir optimization. *Journal of Hydroinformatics, 19*(1), 47-61.

Jothiprakash, V., & Shanthi, G. (2006). Single reservoir operating policies using genetic algorithm. W*ater Resources Management, 20*(6), 917-929.

Labadie, J. W. (2004). Optimal operation of multireservoir systems: state-of-the-art review. *Journal of Water Resources Planning and Management, 130*(2), 93-111.

Laifa, A., & Boudour, M. (2009). Multi-objective particle swarm optimization for FACTS allocation to enhance voltage security. *International Review of Electrical Engineering, 4*(5), 994-1004.

Nayak, M. R., Nayak, C. K., & Rout, P. K. (2012). Application of multi-objective teaching learning-based optimization algorithm to optimal power flow problem. *Procedia Technology, 6*, 255-264.

Rao, R. V., & Kalyankar, V. D. (2013). Parameter optimization of modern machining processes using teaching–learning-based optimization algorithm. *Engineering Applications of Artificial Intelligence, 26*(1), 524-531.

Rao, R. V., & Patel, V. (2013). Multi-objective optimization of heat exchangers using a modified teaching-learning-based optimization algorithm. *Applied Mathematical Modelling, 37*(3), 1147-1162.

Rao, R. V., Savsani, V. J., & Balic, J. (2012). Teaching–learning-based optimization algorithm for unconstrained and constrained real-parameter optimization problems. *Engineering Optimization, 44*(12), 1447-1462.

Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2011). Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems. *Computer-Aided Design, 43*(3), 303-315.

Toğan, V. (2012). Design of planar steel frames using teaching–learning based optimization. *Engineering Structures, 34*, 225-232.

Yeh, W. W. G. (1985). Reservoir management and operations models: A state-of-the-art review. *Water Resources Research, 21*(12), 1797-1818.

Zhang, Z., Jiang, Y., Zhang, S., Geng, S., Wang, H., & Sang, G. (2014). An adaptive particle swarm optimization algorithm for reservoir operation optimization. *Applied Soft Computing, 18*, 167-177.

Zou, F., Wang, L., Hei, X., Chen, D., & Wang, B. (2013). Multi-objective optimization using teaching-learning-based optimization algorithm. *Engineering Applications of Artificial Intelligence, 26*(4), 1291-1300.

Determination of the Optimal Pre-processing Technique for Spectral Data of Oil Palm Leaves with Respect to Nutrient

Helena Anusia James Jayaselan*, Wan Ishak Wan Ismail, Nazmi Mat Nawi and Abdul Rashid Mohamed Shariff

Department of Biological and Agricultural Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Precision agriculture with regard to crop science was introduced to apply only the required and optimal amount of fertiliser, which inspired the present study of nutrient prediction for oil palm using spectroradiometer with wavelengths ranging from 350 to 2500 nm. Partial least square (PLS) method was used to develop a statistical model to interpret spectral data for nutrient deficiency of nitrogen (N), phosphorus (P), potassium (K), magnesium (Mg), calcium (Ca) and boron (B) of oil palm. Prior to the development of the PLS model, pre-processing was conducted to ensure only the smooth and best signals were studied, which includes the multiplicative scatter correction (MSC), first and second derivatives and standard normal variate (SNV), Gaussian filter and Savitzky-Golay smoothing. The MSC technique was the optimal overall pre-treatment method for nutrients in this study, with highest prediction R^2 of 0.91 for N and lowest RMSEP value of 0.00 for P.

Keywords: Nutrients, oil palm, pre-processing, partial least square, spectroradiometer

INTRODUCTION

Oil palm is commonly known as the golden crop as it is the most globally produced and consumed oil (Phin, Stephen, & Markus, 2009). Crude palm oil (CPO) and kernel oil are not only consumed as vegetable oil that is a staple food ingredient for thousands of consumer products, but it also the source of

Current Affiliation:
Nazmi Mat Nawi
Institute of Plantation Studies, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia.

biodiesel renewable energy in fuel industry (Liaghat et al., 2014; Sheil et al., 2009; Carter, Finley, Fry, Jackson, & Willis, 2007; Graboski & McCormick, 1998). It is also widely applied in chemical industry, namely cosmetics, soap and rubber chemicals industry (Keng et al. 2009). A new study from Europe Economic Community show palm oil provides billions of dollars in gross domestic product (GDP), creating millions of sustainable jobs especially for the population in Malaysia (Basiron, 2016; Simeh & Ahmad, 2001).

The scale and reach of palm oil's positive impact creates high demand and the need to boost its productivity (Corley, 2009). Every year, more research is conducted in improving oil palm yield per hectare. Studies have shown that nutrient amendments in inorganic fertilisers lead to changes in the associated CPO constituents, while the quantitative information on the uptake of nutrients in a palm is still very sparse, including the size and growth of palm roots (Hamzah, Salleh, Wong, & Sarmani, 2016; Henson & Chai, 1997). Focusing on the application of fertilisers, study plots were conducted to evaluate the requirement of nutrients in oil palm tree to research on its nutrient intake, demands and losses as well as rate of growth (Dubos, Snoeck, & Flori, 2016; Hashim et al., 2001). Dwelling deep into precision agriculture, many methods of acquiring data were explored, namely Geographic Information System (GIS), Global Positioning System (GPS), and Remote Sensing (RS) that leads to higher potential of increasing yields (Khorramnia, Shariff, Rahim, & Mansor, 2014; Wahid, Xaviar, Tarmizi, & Ibrahim, 2002).

A new and revolutionary method is required to increase the accuracy of acquired data, leading to research on sensors such as the low altitude remote sensing (LARS) multispectral sensors remote temperature sensing and spectroradiometer (Jensen, Apan, Young, & Zeller, 2007; Sugiura, Noguchi, & Ishii, 2005; Jones, 2004). Spectroradiometer was used to obtain reflectance data of oil palm leaflets to determine the level of nutrient (nitrogen (N), phosphorus (P) and potassium (K)) (Jayaselan et al., 2017). The conventional method of detecting nutrient levels in oil palm leaflet was by using destructive methods with application of chemical analysis (Bechlin, Fortunato, Silva, Ferreira, & Neto, 2014). However, it takes a long time to determine the nutrient levels using the conventional methods; thus, immediate in-field results are desirable which would help to make quick judgement on fertiliser applications as part of the study on palm nutrient intake.

In order to obtain robust results using spectral reflectance data, preliminary studies involving the pre-processing method was required prior to its model development, to increase the accuracy of the models and differentiate the nutrients levels in oil palm. The pre-processing methods were divided into four categories of smoothing, derivatives, normalisations and standardisation (Nicolai et al., 2007). For this study, spectral reflectance data from oil palm leaves were acquired for nutrient level detection and classifications. However, prior to this, suitable pre-processing methods are analysed to ensure highly accurate results in this study. The content of interest for this study includes macronutrients nitrogen (N), phosphorus (P), potassium (K), magnesium (Mg) and calcium (Ca) as well as micronutrient boron (B) which are crucial for better productivity and yield (Corley & Mok, 2008).

The purpose of data pre-processing is to ensure a good correlation between spectral data and the concentration values while enhancing spectral features and removing unwanted variation of artefact in data. Hence, it is important to select the most suitable pre-treatment

technique because it can either make or break the data subsequent analysis (Engel et al., 2013). The specified objective was to determine the optimal pre-processing technique for spectral reflectance data of oil palm leaves with respect to nutrients N, P, K, Mg, Ca, and B using partial least square (PLS) model.

MATERIALS AND METHOD

Sample Acquisition

A total of 30 oil palm leaves were taken from frond number 17 (F17) of trees aged 20 years. The choice of f rond number determined was based on results and discussion of previous researchers (Hashim et al., 2001; Rajaratnam, Chan, & Goh, 1977). Plots from an oil palm plantation located at United Plantations, Jenderata, Malaysia were used for this experiment. The leaves from both right and left side of frond 17 were cut at its mid-section as shown in Figure 1. All the leaf samples were cleaned with distilled water before taking reading using a portable contact probe spectroradiometer (Jayaselan et al., 2017). This Analytical spectral device (ASD) called the Fieldspec 4 is a full range spectroradiometer (FRS) that has wavelength ranging from 350 to 2500 nm which provides uniform visible (vis), near infrared (NIR) and short-wave infrared (SWIR) manufactured by Boulder, Colorado, USA. Once the spectral reflectance data of leaf samples were acquired, the lot was sent to the nearest lab to obtain its nutrient level using conventional destructive method of chemical analysis.

Figure 1. Samples were obtained from mid-section of frond 17

The digital numbered (DN) reflectance data acquired using spectrometer was then converted into percentage of reflectance using ASD ViewSpec software. The data selected was in reflectance format rather than absorption or transmittance for better representation of nutrients level (Lelong et al., 2010; Kruse, 1994). In spectroscopy measurements, reliable calibration and prediction models are vital, thus removal of unwanted background noise and smoothing of data is necessary (Nawi, Jensen, & Chen, 2013).

Spectral Pre-Processing Techniques

It is necessary to reduce the signal contamination due to variations of sunlight scatters, other backgrounds reflectance effects, the instrumental noise and dirt on leaves (Nawi, Jensen, &

Chen, 2012; Cen & He, 2007). These disturbances usually act as baseline on the reflectance spectra, which can be a first approximation and considered as linear (Jamshidi, Minaei, Mohajerani, & Ghassemian, 2012). There is no efficient methodology yet on the selection of parameter such as smoothing window size, degree of polynomial fit and derivative order. Therefore, a few sets of combination of these parameters were tested and the optimal solution extracted. In random, six types of pre-processing techniques were explored, which consist of multiplicative scatter correction (MSC), standard normal variate (SNV), Gaussian filter (GF), Savitzky Golay smoothing (SG), gap segment (GS) and normalisation (NM). Among the six types of techniques, each were further explored among the same type, varying in terms of degree of polynomial fit and derivative order to determine the best technique.

The six techniques were further classified into four main methods of smoothing, differentiation, standardisation and normalisation. The MSC technique falls under normalisation method that is popularly used due to its ability to handle both additive and multiplicative effects of light scattering in reflectance spectroscopy. However, a number of similar effects are successfully treated with MSC, such as offset shifts and interference. It uses the mean spectrum as reference for the data set with the assumption that the calculated mean is representing all current and future spectra collected using similar field conditions without affecting the chemical signal in the data (Bi et al., 2016). The SNV technique is another type of normalisation method that removes multiplicative interference of scatter and particle size, baseline shift and curvilinearity of spectral (Rammal et al., 2014; Ding & Xu, 1999). The difference between the SNV and MSC techniques is that SNV is applied to an individual spectrum, while MSC uses a reference spectrum, such as the mean spectrum of the calibration set for correction of the spectral (Moghimi, Aghkhani, Sazgarnia, & Sarmad, 2010).

The SG and GS technique of smoothing method uses information from a localised segment of the spectrum to compute the derivative at a particular wavelength. Usually, this avoids the problem of noise enhancement from the simple difference method and applies smoothing to the data. The GS technique also requires appropriate gap size and smoothing segment size for best results. The SG and GS also fall into the category of derivation, provided the order is more than one. The SG technique was explored as it uses convolution function, and thus the number of segments in the function must be specified to one and two polynomial order to have best representation of local spectrum behaviour and avoid over smoothing (Liaghat et al., 2014). The SG technique that consists of a polynomial order of two followed by a derivative computation for both first and second order is also used to perform baseline correction especially to detect subtleness in vegetation (He, Li, & Ma, 2015; Liaghat et al., 2014). The GS derivatives are commonly applied as correction for baseline effects in spectra for the purpose of removing non-chemical effects and resolving overlapped bands, as well as aiding in emphasising small spectral variations, which are not evident in the raw data, thus creating robust calibration models. The first derivative was used since it is a measure of the slope of the spectral curve at every point, effective for removing purely additive baseline offsets (Jamshidi et al., 2012).

Nevertheless, peaks in raw spectral usually become zero-crossing points in first derivative spectra, which cause difficulty in interpreting them. This is when the second derivative is usually desired to resolve nearby peaks and sharpen spectral features while maintaining the intensities of the original curves. The second derivative is explored as a measure of the change

in the slope of the curve to remove pure additive offset, unaffected by any linear "tilt" that may exist in the data and is therefore a very effective method for removing both the baseline offset and slope from a spectrum. The first and second derivative is effective in increasing the spectral resolution (Cen & He, 2007). The third order derivative was explored with a hope to reveal phenomena, which do not appear clearly when using lower-order derivatives and can be helpful in understanding the spectral data. Gaussian Filter is a form of smothering for spectral signals, more commonly used for pulse de-noising (Wang, Zhang, & Lu, 2016).

NM techniques falls under standardisation method that are transformations computed based on samples. There are many types of normalisation, such as area normalisation, mean normalisation, unit vector normalisation, maximum normalisation, range normalisation and peak normalisation. Area normalisation and mean normalisation (standard method) are similar methods of computing area under the curve and only differ by a constant multiplicative factor, provided the samples are all in the same units (Noviyanto, Abdulla, Yu, & Salcic, 2015). The unit vector normalisation method follows pattern normalisation, usually used for pattern recognition (Kilz & Cazes, 2001). While maximum normalisation divides each row by its maximum absolute value instead of the average like the mean normalisation method, range normalisation divides each row by its range (max value – min value) (Mandic, 2012).

The peak normalisation attempts to correct spectra for indeterminate path length by effectively removing it and normalising the spectrum to the intensity of the peak. Normalisation technique is important method of pre-processing as it performs both scaling and offset correction at once. However, only area normalisation, mean normalisation, maximum normalisation and range normalisation methods were explored in terms of its relevance to the sample features (Liaghat et al., 2014). All the pre-processing techniques were implemented using the Unscrambler 10.4 software of Camo Process As, Oslo, Norway.

Calibration and Validation Models

Effects of several pre-processing techniques on the performance of PLS models were investigated for six different nutrients. The unprocessed raw spectral reflectance data was first analysed and set as benchmark to evaluate the actual gain of pre-processing technique. The analysis of PLS regression generalises and combines features from principal component analysis (PCA) and multiple regression (MR) in order to predict a set of dependent variables from a set of independent variables or predictors. This prediction is achieved by extracting from the predictors a set of orthogonal factors also known as latent variables (LV) which provides best predictive power (Abdi, 2007).

The PLS models were evaluated based on the representation of coefficient of correlation (R^2) and root mean square error (RMSE) values of both calibration and prediction models (Lazim, Nawi, Chen, Jensen, & Rasli, 2017). Before developing the calibration models, the sample data were randomly divided into calibration set (75% of the whole samples) and prediction set (25% of the whole samples) (Nawi et al., 2012). RMSE is termed the root mean square error in calibration (RMSEC) for the calibration set and the root mean square error in prediction (RMSEP) for the external prediction set. Multivariate variables such as the R^2, RMSEC, and RMSEP were computed to evaluate the quality of the PLS models. Models that

gave the lowest possible LV with the highest coefficient of determination (R^2), and lowest root mean square error of prediction (RMSEP) with respect to the lab results obtained from the destructive chemical analysis are selected for each nutrient.

Mostly, the LV used for calibration model was 10. Full cross-validation with leave-four-out option was executed to determine the optimal number of LVs. Exploring the PLS analysis, two types of validation were compared, between the Nipals and Kernal. Generally, Nipals runs with an iterative algorithm that could impute missing values and does well with large set of samples and variables while Kernal performs non-iterative algorithms and is better with large sample group with small set of variables but does not handle missing values (Mevik, Wehrens, & Liland, 2015).

RESULTS AND DISCUSSION

Before starting on pre-processing of spectra, the first and last 50 nm (Jamshidi et al., 2012) of wavelengths were removed because of the existence of considerable noise in these regions. Thus, only 2050 number of wavelengths are considered for pre-processing. A few numbers of pre-processing techniques were tested, alone and in combinations. The Figure 2(a) below shows the original unprocessed, raw leaf spectral reflectance data of 30 samples in percentage, from 400 nm to 2450 nm for F17 samples. Figure 2(b)-(d) below presents the result of some of the pre-processing techniques of the raw spectral. The SG (b) and raw (a) spectra looks alike to the naked eye, while the MSC(d) treated spectra with light scatter corrections has denser, slimmer and sharper spectral compared with the original raw spectral. The GS on the other hand looks very different due to the second order derivative effects. Based on research study by Noviyanto et al. (2015), there is no one best pre-processing technique for a particular field or task, but for every study, suitable spectral pre-treatment is to be determined respectively. Therefore, several pre-processing techniques were explored to find the most suitable method for oil palm leaf.

(a)
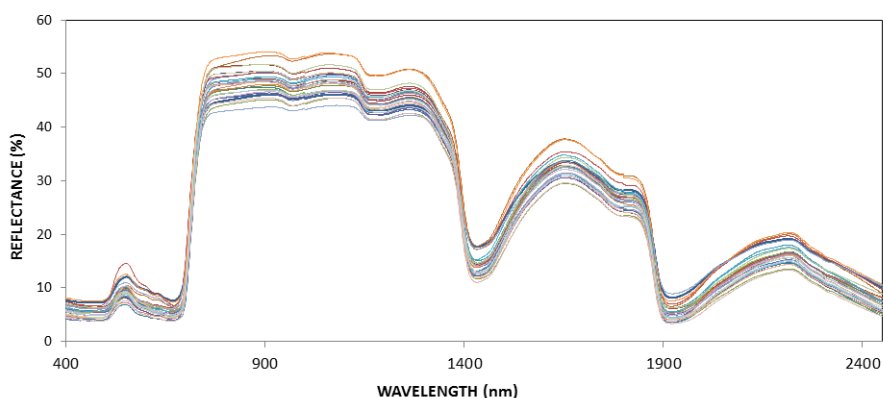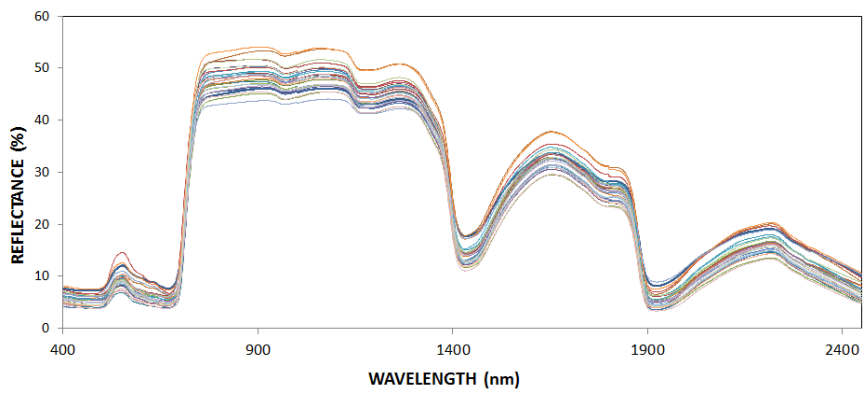
(b)
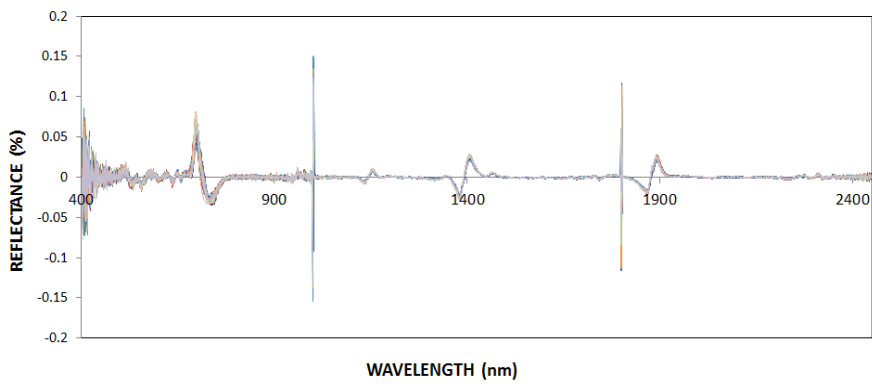
(c)
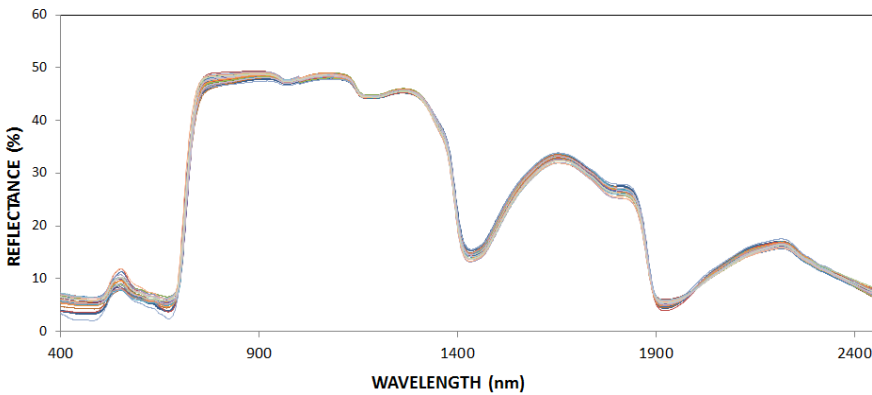
(d)

Figure 2. Reflectance spectral from 30 leaf samples of frond 17 and results of pre-processing techniques; a) Raw spectral, b) Savitzky-Golay (SG), c) Gap-segment 2d (GS) and d) Multiplicative scatter correction (MSC)

All the techniques that were explored had calibration of R^2 value in the range from 50 to 99%, with mostly falling into the mid and upper region. Hence, the prediction model's R^2 value was the determinant factor for the optimal pre-processing technique. The differentiation method used was inclusive of first and second derivative to increase spectral resolution and remove background (Cen & He, 2007). The Savitzky-Golay (SG) technique was explored between polynomial one and two, as it was important to choose the proper width of smoothing window (Cen & He, 2007). The results obtained were mixed, where N, K, Mg and Ca performs better with SG of polynomial one while P and B were better with SG of polynomial two as shown in Table 1.

As illustrated in Table 1, SG techniques with first and second derivative were compared. Between these two, the SG with first derivative clearly shows better representation of prediction R^2 and RMSEP. Lelong et al. (2010) had 94% of overall accuracy for Ganoderma infection level classification based on spectra that was treated with second order derivative of PO 3 of oil palm canopy. Gap Segment (GS) technique of pre-processing was tested for first, second and third order derivatives settings of each with gap size of two and segment size of three. As shown in Table 1, generally the GS with second derivative shows good results for most nutrients except Mg which however had good R^2 representation for Calibration model.

However, only second derivative was found to be the most suitable pre-treatment for protein content determination in wheat flour with R^2 value of 0.98 from calibration model of modified PLS analysis (Huali et al., 2011). The smoothing method that optimises signal to noise ratio and removes random noise includes SG and moving average techniques (Cen & He, 2007; Naes, Isaksson, Fearn, & Davies, 2002). The GF technique of SS5 was considered since its prediction R^2 representation was better for nutrients of K, Mg, Ca and B with considerably low RMSEP in Table 2.

Table 1

Partial Least Square (PLS) models prediction results for N, P, K, Mg, Ca and B for raw spectral, Savitzky-Golay (SG) and Gap segment (GS) of derivative pre-processing method

Pre-treatment	Nitrogen (N)		Phosphorus (P)		Potassium (K)		Magnesium (Mg)		Calcium (Ca)		Boron (B)	
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
Raw	0.05	0.22	0.19	0.01	0.31	0.30	0.14	0.06	-	-	0.57	7.3
S.G 1st D	0.21	0.31	0.06	0.13	0.05	2.09	0.32	0.36	0.11	0.28	0.00	10.74
S.G 2nd D	0.18	0.17	0.18	0.02	-	1.31	-	0.23	0.01	0.31	-	-
G.S 1st D	0.26	0.30	0.05	0.11	0.02	2.01	0.51	0.36	0.17	0.32	-	8.09
G.S 2nd D	0.13	0.32	0.09	0.09	-	0.08	0.20	0.21	0.4	0.23	0.26	14.98
G.S 3rd D	0.05	0.18	0.07	0.01	0.31	0.43	0.40	0.09	0.30	0.16	0.09	9.73

* - indicates no results from the Unscrambler software

Between the Gaussian Filter (GF) of segment size (SS) SS3 and SS5, it was clearly shown in Table 2 that the GF with SS5 gives better representation of R^2 for calibration as well as prediction model. Although some nutrients such as N, P, K and Ca may have higher R^2 representation for calibration, its prediction value was not satisfactory. Standardisation of spectral data was a pre-processing method used by divide the spectrum at every wavelength with its standard deviation at certain wavelength, which includes basic normalisation (NM) techniques. For the pre-processing NM technique, comparison was made among the NM-area, NM-mean, NM-range and NM-Maximum, where the results are close to each other as shown in Table 3. However, NM-range technique performed better for most nutrients of P, K, Mg, Ca and B, although this may not be the best technique available for oil palm leaf reflectance.

Table 2

Partial Least Square (PLS) models prediction results for N, P, K, Mg, Ca and B for raw spectral, Savitzky-Golay (SG) and Gaussian filter (GF) of smoothing pre-processing method

Pre-treatment	Nitrogen (N)		Phosphorus (P)		Potassium (K)		Magnesium (Mg)		Calcium (Ca)		Boron (B)	
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
Raw	0.05	0.22	0.19	0.01	0.31	0.30	0.14	0.06	-	-	0.57	7.3
S.G(PO1)	0.40	1.83	0.09	0.15	0.63	0.08	0.59	0.08	0.46	2.00	0.46	1.98
S.G(PO2)	0.05	0.22	0.19	0.01	0.31	0.30	0.13	0.06	-	-	0.57	7.31
G.F (SS5)	0.66	1.26	0.66	0.18	0.87	0.08	0.92	0.10	0.95	0.57	0.95	1.01
G.F (SS3)	0.05	0.22	0.19	0.01	0.31	0.30	0.07	0.06	-	-	0.57	7.34

* - indicates no results from the Unscrambler software

Table 3

Partial Least Square (PLS) models prediction results for N, P, K, Mg, Ca and B for raw spectral, mean, area, range and maximum (nm) of standardisation pre-processing method

Pre-treatment	Nitrogen (N)		Phosphorus (P)		Potassium (K)		Magnesium (Mg)		Calcium (Ca)		Boron (B)	
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
Raw	0.05	0.22	0.19	0.01	0.31	0.30	0.14	0.06	-	-	0.57	7.3
NM-Mean	0.07	0.23	0.03	0.01	0.46	0.28	0.01	0.09	-	-	0.64	6.90
NM-Area	0.07	0.23	0.05	0.01	0.46	0.28	0.01	0.09	-	0.18	0.64	6.90
NM-Range	0.01	0.26	0.20	0.01	0.10	0.37	0.01	0.09	-	0.16	0.80	5.14
NM-Maxi	0.19	0.25	0.29	0.01	0.50	0.28	0.08	0.09	0.30	0.18	0.64	6.79

* - indicates no results from the Unscrambler software

The normalisation method applies multiplicative effects and compensates for baseline shifts in the spectral data includes, MSC and SNV were as shown in Table 4 (Lazim et al., 2016). The MSC technique was evaluated in terms of PLS-Nipals and PLS Kernal. For most of the nutrients tested, the MSC technique applied with Kernal presents better results in terms of prediction R^2 and RMSEP, as shown in Table 4. This was seen in P and Mg in its R^2 representation for prediction and RMSEP for B. The SNV pre-processing technique performed similar to MSC and SNV, MSC-SG and MSC-SNV-SG combinations with MSC-SNV-SG being slightly better combination among this lot as seen in Table 4.

Table 4

Partial Least Square (PLS) models prediction results for N, P, K, Mg, Ca and B for raw spectral, multiplicative scatter correction (MSC) and standard normal variate (SNV) of normalisation method and mix of pre-processing techniques

Pre-treatment	Nitrogen (N)		Phosphorus (P)		Potassium (K)		Magnesium (Mg)		Calcium (Ca)		Boron (B)	
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
Raw	0.05	0.22	0.19	0.01	0.31	0.30	0.14	0.06	-	-	0.57	7.3
MSC Kernal	0.91	0.34	0.72	0.00	0.90	0.03	0.66	0.05	0.26	0.11	0.74	0.1
MSC Nipals	0.11	0.22	0.03	0.01	0.42	0.27	0.04	0.07	0.37	0.19	0.73	9.19
SNV	0.13	0.23	0.03	0.01	0.44	0.27	0.04	0.07	0.00	0.18	0.73	8.38
MSC-SNV	0.13	0.23	0.03	0.01	0.44	0.27	0.04	0.07	0.00	0.16	0.73	8.38
MSC-S.G	0.11	0.22	0.03	0.01	0.42	0.27	0.04	0.07	0.08	0.16	0.73	9.33
MSC-SNV-S.G	0.13	0.23	0.03	0.01	0.44	0.27	0.04	0.07	-	0.16	0.73	8.32

* - indicates no results from the Unscrambler software

The SG with MSC was found to have best relationship between spectral absorbance and soluble solids content (SSC Brix) with higher R^2 of 0.98 when compared to SG and MSC as well as Moving average method with MSC or SNV (Jamshidi et al., 2012). However, in an overall comparison among the best setting of each type of pre-processing technique, the MSC technique seems to fulfil the aim of high R^2 representation as well as low RMSEP. This was better than a study on grass for N using localised SNV that was found to have higher RMSEP of 0.169 (Bi et al., 2016). The PLS using spectroradiometer as sensor was proven to have good relationship for prediction of nutrients N, P, and K by Jayaselan et al. (2017).

CONCLUSION

In order to obtain highly accurate analysis and modelling results, pre-processing of spectral data is required. After evaluation of four methods and six techniques of spectral pre-processes, the MSC technique was found to be the best choice for this data form, as an overall nutrient predictor. The MSC results for N, P, K, Mg, Ca, and B in terms of R^2 representation were 0.91, 0.72, 0.90, 0.66, and 0.26, and 0.74 while in terms of RMSEP values were 0.34, 0.00, 0.03, 0.05, 0.11, and 0.1. Nevertheless, in the context of representing only one nutrient, it was found that the most suitable pre-processing technique for Mg, Ca and B was Gaussian Filter with SS5 parameter and MSC for N, K, and P nutrients.

REFERENCES

Adbi, H. (2007). Partial least square regression. PLS-regression. *Encyclopedia of Measurement and Statistics, 6(*4), 792-795.

Basiron, Y. (2016). Booster for global economies; New research shows palm oil's impact on revenues worldwide. *Palm Voices Columnist.* The Star (4 October 2016).

Bechlin, M. A., Fortunato, F. M., Silva, R. M. D., Ferreira, E. C., & Neto, J. A. G. (2014). A simple and fast method for assessment of the nitrogen-phosphorus-potassium rating of the fertilisers using high-resolution continuum source atomic and molecular absorption spectrometry. *Spectrochimica Acta Part B: Atomic Spectroscopy, 101*, 240-244.

Bi, Y., Yuan, K., Xiao, W., Wu, J., Shi, C., Xia, J., … & Zhou. (2016). A local pre-processing method for near-infrared spectra combined with spectral segmentation and standard normal variate transformation. *Journal of Analytica Chinica Acta, 909*, 30-40.

Carter, C., Finley, W., Fry, J., Jackson, D., & Willis, L. (2007). Palm oil markets and future supply. *European Journal of Lipid Science and Technology, 109*(4), 307-314.

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology, 18*, 72-83.

Corley, R. H. V. (2009). How much palm oil do we need? *Journal of Environmental Sciences and Policy, 12*(2), 134-139.

Corley, R. H. V., & Mok, C. K. (2008). Effects of nitrogen, phosphorus, potassium and magnesium on growth of the oil palm. *Journal of Experimental Agriculture, 8*(4), 347-353.

Ding, H. B., & Xu, R. J. (1999). Differentiation of beef and kangaroo meat by visible/near infrared reflectance spectroscopy. *Journal of Food Science, 64*(5), 814-817.

Dubos, B., Snoeck, D., & Flori, A. (2017). Excessive use of fertiliser can increase leaching process and modify soil reserves in two Ecuadorian oil palm plantations. *Journal of Experimental Agriculture, 53*(2), 255-268.

Engel, J., Gerretzen, J., Szymanska, E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L. M. C. (2013). Breaking with trends in pre-processing? *Journal of Trends in Analytical Chemistry, 50*, 96-106.

Graboski, M. S., & McCormick, R. L. (1998). Combustion of fat and vegetable oil derived fuels in diesel engines. *Journal of Progress in Energy and Combustion Science, 24*(2), 125-164.

Hamzah, A., Salleh, S. N. M., Wong, K. K., & Sarmani, S. (2016). Nutrient amendments of inorganic fertiliser and oil palm empty fruit bunch and their influence on bacterial species dominance and degradation of the associated crude oil constituents. *Journal of Soil and Sediment Contamination, 25*, 256-265.

Hashim, M., Ibrahim, A. L., Rasib, A. W., Shah, R., Nordin, L., & Haron, K. (2001). Detecting oil palm tree growth variability using a field spectroradiometer. *ASIAN-PACIFIC Remote Sensing and GIS Journal, 14*, 25-32.

He, Z., Li, M., & Ma, Z. (2015). Prediction of dry matter, protein and acidity in corn steep liquor using near infrared spectroscopy. In *7th International Conference on Awareness Science and Technology (iCAST)* (pp. 212 – 215). IEEE Conference Publications.

Henson, I. E., & Chai, S. H. (1997). Palm productivity: Biomass distribution, productivity and turnover of the root system. *Elaeis, 9*(2), 78-92.

Huali, J., Jinshui, W., Lihui, Y., Rui, G., & Li, L. (2011). Establishment of non-destructive testing model of the protein content in wheat flour by NIR spectroscopy. In *New Technology of Agricultural Engineering (ICAE), 2011 International Conferenc*e (pp. 1087-1091). IEEE.

Jamshidi, B., Minaei, S., Mohajerani, E., & Ghassemian, H. (2012). Reflectance Vis/NIR spectroscopy for non-destructive taste characterization of Valencia oranges. *Journal of Computers and Electronics in Agriculture, 85*, 64-69.

Jayaselan, H. A. J., Nawi, N. M., Ismail, W. I. W., Shariff, A. R. M., Rajah, V. J., & Arulandoo, X. (2017). Application of spectroscopic method to predict nutrient content of oil palm from leaf. *Journal of Experimental Agriculture, 15*(3), 1-9.

Jensen, T., Apan, A., Young, F., & Zeller, L. (2007). Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform. *Journal of Computers and Electronics in Agriculture, 59*(1), 66-77.

Jones, G. H. (2004). Application of thermal imaging and infrared sensing in plant physiology and ecophysiology. *Journal of Advances in Botanical Research, 41*, 107-163.

Keng, P. S., Baasri, M., Zakaria, M. R. S., Rahman, M. B. A., Ariff, A. B., Rahman, R. N. Z. A., & Salleh, A. B. (2009). Newly synthesized palm esters for cosmetics industry. *Journal of Industrial Crops and Products, 29*(1), 37-44.

Khorramnia, K., Shariff, A. R. M., Rahim, A. A., & Mansor, S. (2014). Toward malaysian sustainable agriculture in 21st century. In *IOP Conference Series: Earth and Environmental Science* (Vol. 18, No. 1, p. 012142). IOP Publishing.

Kilz, P., & Cazes, J. (2001). *Encyclopaedia of chromatography* (pp. 195-199). Marcel Dekkar Inc.

Kruse, F. A. (1994). Imaging spectrometer analysis: A Tutorial. In *Proceeding of International Symposium on Spectral Sensing Research (ISSSR)* (pp: 10-15). San Diego, CA.

Lazim, S. S. R. M., Nawi, N. M., Chen, G., Jensen, T., & Rasli, A. M. M. (2016). Influence of different pre-processing methods in predicting sugarcane quality from near infrared (NIR) spectral data. *International Food Research Journal, 23*, 231-236.

Lelong, C. C. D., Roger, J. M., Bregand, S., Dubertret, F., Lanore, M., Sitorus, … & Caliman, J. P. (2010). Evaluation of oil palm fungal disease infestation with canopy hyperspectral reflectance data. *Journal of Sensors, 10*, 734-747.

Liaghat, S., Mansor, S., Ehsani, R., Shafri, H. Z. M., Meon. S., & Sankaran, S. (2014). Mid-infrared spectroscopies for early detection of basal stem rot disease in oil palm. *Journal of Computers and Electronics in Agriculture, 101*, 48-54.

Mandic, Z. (2012). *Physio chemical methods in drug discovery and development.* Zagreb, Croatia: IAPC Publishing.

Mevik, B. H., Wehrens, R., & Liland, K. H. (2015). *Package 'PLS'.* Retrieved from http://mevik.net/work/software/pls.html

Moghimi, A., Aghkhani, M. H., Sazgarnia, A., & Sarmad, M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Journal of Biosystem Engineering, 106*(3), 295-302.

Nawi, N. M., Jensen, T., & Chen, G. (2012). The application of spectroscopic methods to predict sugarcane quality based on stalk cross-sectional scanning. *Journal of American Society of Sugar Cane Technologists, 32*, 16-27.

Nawi, N. M., Jensen, T., & Chen, G. (2013). Application of spectroscopic method to predict sugar content of sugarcane internodes. *Journal of Tropical Agriculture and Food Science, 41*(2), 211-220.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*, 99-118.

Noviyanto, A., Abdulla, W., Yu, W., & Salcic, Z. (2015). Research trends on optical spectrum for honey analysis. In *Proceedings of Asia-Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference* (pp. 416-425). IEEE Conference Publications.

Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to multivariate calibration and classification.* Charlton, Chichester, UK: NIR Publications.

Phin, C. K., Stephen, R., & Markus, A. (2009). In vitro antimicrobial activity and fungi toxicity of syringic acid, caffeic acid and 4-hydroxybenzoic acid against Ganoderma Boninense. *Journal of Agriculture Science, 1*(2), 15-20.

Rajaratnam, J. A., Chan, K. W., & Goh, K. H. (1977). The foundation for selecting leaf 17 for nutrient requirements of mature oil palms. In *Proceeding Conference on Classification and Management of Tropical Soils (Joseph, K.T., Ed.)* (pp: 340-348). Malaysian Society of Soil Science, Kuala Lumpur, Malaysia.

Rammal, A., Perrin, E., Vrabie, V., Bertrand, I., Habrant, A., & Chabbert, B. (2014). Optimal pre-processing and FCM clustering of MIR, NIR and combined MIR-NIR spectra for classification of maize roots. In *The Third International Conference on e-Technologies and Networks for Development* (pp: 110–115). IEEE Conference Publications.

Sheil, D., Casson, A., Meijaard, E., Nordwijk, V. M., Gaskell, J., Groves, S. J., … & Kanninen, M. (2009). *The impacts and opportunities of oil palm in Southeast Asia: What do we know and what do we need to know?* (No. 51). CIFOR, Bogor, Indonesia.

Simeh, A., & Ahmad, T. M. A. T. (2001). The case study on Malaysian palm oil. In *Regional Workshop on Commodity Export Diversification and Poverty Reduction in South and South-East Asia* (pp 3-5). Bangkok.

Sugiura, R., Noguchi, N., & Ishii, K. (2005). Remote sensing technology for vegetation monitoring using an unmanned helicopter. *Biosystem Engineering, 90*(4), 369-379.

Wahid, O., Xaviar, A., Tarmizi, A. M., & Ibrahim, S. (2002). Precision agriculture: Fertiliser management MAP: Spatial and correlation analysis of yield and leaf nutrient. *MPOB Information Series*.

Wang, D., Zhang, D., & Lu, G. (2016). A robust signal pre-processing framework for wrist pulse analysis. *Journal of Biomedical Signal Processing and Control, 23*, 62-75.

SCIENCE & TECHNOLOGY

A Novel Entropy Algorithm for State Sequence of Bakis Hidden Markov Model

Jason Chin-Tiong Chan[1]* and Hong Choon Ong[2]

[1]*Ted Rogers School of Management, Ryerson University, 350 Victoria St, Toronto, ON M5B 2K3, Canada*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Gelugor, Penang, Malaysia*

ABSTRACT

Hidden Markov model (HMM) can be categorised as an ergodic model or a left-to-right model. The categorization is subject to its state transition. An ergodic Hidden Markov model has full state transitions but a left-to-right hidden Markov model has partial state transitions. A Bakis Hidden Markov model (BHMM) is a special type of the left-to-right Hidden Markov model. State sequence for a BHMM is invisible but this research is able to track the most likelihood state sequence using Viterbi algorithm. However, while tracking the optimal state sequence for BHMM, the conventional algorithm does not provide a measure of uncertainty which is present in the solution. This issue can be overcome by the proposed novel algorithm, namely, BHMM entropy-based forward algorithm (BHMM-EFA) for computing state entropy of a BHMM. This algorithm is based on a decreasing-ladder trellis structure which provides a clear picture on how the entropy associated with the optimal state sequence is determined. Therefore, the novel algorithm requires $O(TN)$ calculations for tracking the optimal state sequence of a first-order BHMM where T is the length of the observational sequence and N is the number of hidden states.

Keywords: Bakis Hidden Markov model, entropy, forward probability, state transition, uncertainty, Viterbi Algorithm

INTRODUCTION

A left-to-right Hidden Markov Model (LR-HMM) is an important subclass of Hidden Markov Models for modelling time series data (Rabiner, 1989). This subclass model is a single directional structure model and hence named a left-to-right Hidden Markov Model. This subclass of model structure is useful in applications such as speech recognition (Gales & Young, 2008; Nogueries et al., 2001; Juang

& Rabiner, 1991) and DNA sequence analysis (Regad, 2008; Loytynoja & Milinkovitch, 2003). HMM is a statistical sequence model where the observational sequence is generated independently but conditioned on a hidden state Markov chain. The index of the hidden state for LR-HMM either increases or maintains the same state as time progresses. As a result, the transition matrix of LR-HMM is an upper triangular matrix with zeros on the below of the main diagonal entries, but there is at least one non-zero entry on or above the diagonal per column. State transitions between non-adjacent states are permitted in this LR-HMM, for example, a transition from state 2 to state 6. In some applications such as speech recognition (Nogueries et al., 2001), it requires more constraints on the state transition structure so that only transitions among neighbouring states are allowed. This type of special LR-HMM is called as a Bakis Hidden Markov Model (BHMM). There are two ways of state transitions for BHMM: self-state transition (i.e., from state i to state i) and one-step transitions (i.e., from state i to state i + 1). Hence, there is at least one entry for the main diagonal and the first upper diagonal is non-zero per row for the state transition matrix structure in a BHMM. The hidden state sequence of HMM can be tracked from a given observational sequence and the restored state has many applications especially when the hidden state sequence has meaningful interpretations in terms of prediction. For example, Ciriza et al. (2011) determined the optimal printing rate based on the HMM model parameter and an optimal time-out which is based on the restored states. Viterbi algorithm is the most common technique for tracking hidden state sequence (Rabiner, 1989). However, it does not measure the uncertainty present in the solution. Proakis and Salehi (2002) proposed a method for measuring the error of a single state. This method is unable to measure the error of the entire state sequence computed by Viterbi algorithm. Hernando et al. (2005) proposed a method of using entropy for measuring the uncertainty of the state sequence of a first-order HMM tracked from a single observational sequence with a length of T. The method is based on the forward recursion algorithm integrated with entropy. Mann and McCallum (2007) developed an algorithm for computing the subsequent constrained entropy of HMM which is similar to probabilistic model conditional random fields (CRF). Ilic (2011) developed an algorithm based on forward-backward recursion over the entropy semiring, namely Entropy Semiring Forward-backward algorithm (ESRFB) for a first-order HMM with a single observational sequence. ESRFB has lower memory requirement as compared with Mann and McCallum's algorithm for subsequent constrained entropy computation.

This paper is organised as follows. In section 2, a Bakis Hidden Markov model is defined and a modified forward probability variable is introduced in a decreasing-ladder trellis structure diagram. In section 3, an algorithm is proposed for computing state entropy for a Bakis Hidden Markov Model, namely BHMM entropy-based forward algorithm (BHMM-EFA) and a numerical example of computing state entropy is illustrated. We discuss future research in section 4 on this state entropy algorithm.

FIRST-ORDER DISCRETE BAKIS HIDDEN MARKOV MODEL

In this section, definitions and notations for a first-order BHMM is introduced. These are followed by the definition of the modified forward probability variable for a first-order BHMM. These forward probability variables form the basis for the proposed algorithm, namely BHMM

entropy-based forward algorithm (BHMM-EFA). A trellis structure diagram is then defined and a a decreasing-ladder trellis structure diagram is introduced which provides a clear picture on how the entropy associated with the optimal state sequence of BHMM is determined.

Element of BHMM

BHMM involves two stochastic processes, namely hidden state process and observation process. The hidden state process cannot be directly observed. The observation sequence is generated by the observation process incorporated with the hidden state process. For discrete BHMM, it has to satisfy the following conditions:

The hidden state process $\{q_t\}_{t=1}^T$ is a first-order Bakis Markov chain that satisfies

$$P\left(q_t \middle| \{q_l\}_{l<t}\right) = P\left(q_t \middle| q_{t-1}\right) \tag{1}$$

where q_t denotes the hidden state at time t, and $q_t \in S$ where S is the finite set of hidden states. Note that if q_{t-1} is at the state s_i then q_i is either at the state s_i or state s_{i+1} due to self-transition or one-step transition.

The observation process $\{o_t\}_{t=1}^T$ is incorporated into the hidden state process according to the state probability distribution that satisfies

$$P\left(o_t \middle| \{o_l\}_{l<t}, \{q_l\}_{l\le t}\right) = P\left(o_t \middle| q_t\right) \tag{2}$$

where o_t denotes the observation at time t, and $o_t \in V$ where V is the finite set of observation symbols.

The elements for a first-order discrete BHMM are as follows:

- Number of distinct hidden states, N
- Number of distinct observed symbols, M
- Length of observed outputs, T
- Observed output sequence, $O = \{o_t, t = 1,2,...,T\}$
- Hidden state sequences, $Q = \{q_t, t = 1,...,T\}$
- Possible values for each state, $S = \{s_i, i = 1,2,...,N\}$
- Possible symbols per observation, $V = \{v_w, w = 1,2,...,M\}$
- Initial hidden state probability vector, π_i

    π_i is the probability that the model will start from state s_i

    $$\pi_i = P\left(q_1 = s_i\right) , \quad \sum_{i=1}^N \pi_i = 1, \pi_i \ge 0, \ 1 \le i \le N$$

- State transition probability matrix, $A = \{a_{ij}\}$

A is the two-dimensional state transition probability matrix, and a_{ij}, is the probability of a transition to state s_j given that it has had a transition from state s_i where $1 \leq i, j \leq N$.

$$a_{ij} = P(q_t = s_j | q_{t-1} = s_i), \sum_{j=i}^{i+1} a_{ij} = 1, a_{ij} \geq 0$$

- Emission output probability matrix, $B = \{b_i(v_m)\}$

B is the two-dimensional emission output probability matrix, and $b_i(v_m)$ is a probability of observing v_m in state s_i where $1 \leq i \leq N$.

$$b_i(v_m) = P(o_t = v_m | q_t = s_i), \sum_{m=1}^{M} b_i(v_m) = 1, b_i(v_m) \geq 0$$

For a first-order discrete BHMM, parameters by using the components of $\lambda = (\pi, A, B)$ are summarised.

Note that throughout this paper, we will use the following notations.

- $q_{1:t}$ denotes $q_1, q_2, ..., q_t$
- $o_{1:t}$ denotes $o_1, o_2, ..., o_t$

Forward Probability for BHMM

Hernando et al. (2005) proposed an algorithm that incorporated the forward recursion process for computing the entropy of state sequence. Ilic (2011) also developed an algorithm based on forward-backward recursion process for state entropy computation. Both algorithms are formulated for a first-order HMM with a single observation sequence. The conventional forward probability based on the state transition structure of BHMM is modified. The BHMM has two types of state transition, which are self-state and one-step transition. These forward probabilities are computed recursively in the recursion and the termination phase. The algorithms proposed by Hernando et al. (2005) and Ilic (2011) require $O(TN^2)$ calculations for computing state entropy of a generalised first-order HMM. The new algorithm performs state entropy computation that requires $O(TN)$ due to its modified forward probability. The conventional forward probability for a first-order HMM is defined as follows (Rabiner, 1989):

Definition 1. The conventional forward variable $\alpha_t(i)$ for a first-order HMM is a joint probability of the partial observation sequence $o_1, o_2, \cdots, o_t$ and the hidden state s_i of at time t where $1 \leq t \leq T$. It can be denoted as

$$\alpha_t(i) = P(o_1, o_2, \cdots, o_t, q_t = s_i | \lambda) \tag{3}$$

From (3), $t = 1$ and $1 \le i \le N$, the initial forward variable can be expressed as

$$\alpha_1(i) = P\big(o_1, q_1 = s_i \big| \lambda\big) = P\big(q_1 = s_i\big)P\big(o_1 \big| q_1 = s_i\big)$$
$$= \pi_i b_i(o_1)$$

(4)

From (3), (4), the recursive forward variable for BHMM is obtained where $t = 2,...,T$,

$$\alpha_t(j) = P\big(o_1, o_2, \ldots, o_t, q_t = s_j \big| \lambda\big)$$

$$= \sum_{i=j-1}^{j} P\big(o_1, o_2, \ldots, o_t, q_{t-1} = s_i, q_t = s_j \big| \lambda\big)$$

$$= \sum_{i=j-1}^{j} P\big(o_1, o_2, \ldots, o_{t-1}, q_{t-1} = s_i \big| \lambda\big) P\big(q_t = s_j \big| q_{t-1} = s_i\big) P\big(o_t \big| q_t = s_j\big)$$

$$= \sum_{i=j-1}^{j} \alpha_{t-1}(i) a_{ij} b_j(o_t)$$

(5)

Note that the above summation is from j-1 to j at time t due to the state transition structure of BHMM whereas the summation is from 1 to N for the conventional recursive forward probability where N is the number of the hidden states.

The recursive modified forward probability in (5) can be represented in the following form

$$\alpha_t(j) = \begin{cases} \alpha_{t-1}(j) a_{11} b_1(o_t) & if \quad j = 1 \\ \big[\alpha_{t-1}(j-1) a_{(j-1)(j)} + \alpha_{t-1}(j) a_{jj}\big] b_j(o_t) & if \quad 2 \le j \le N \end{cases}$$

(6)

The modified forward variable is normalised that is required as an intermediate variable in our proposed algorithm. The following is the definition for a normalised forward variable which is used for obtaining a normalising modified forward variable.

Definition 2. The normalising modified forward probability variable $\hat{\alpha}_t(i)$ in a first-order BHMM is defined as the probability of the hidden state of s_i at time given the partial observation sequence $o_1, o_2, \cdots, o_t$ where $1 \le t \le T$.

$$\hat{\alpha}_t(i) = P\big(q_t = s_i \big| o_1, o_2, \cdots, o_t\big)$$

(7)

From (4), (7), $t=1$ and $1 \le i \le N$, the initial normalised forward variable is obtained as

$$\hat{\alpha}_1(j) = P\big(q_1 = s_j \big| o_1\big)$$
$$= \frac{P\big(q_1 = s_j, o_1\big)}{P(o_1)}$$
$$= \frac{\pi_j b_j(o_1)}{r_0}$$

(8)

where $r_0 = \sum_{k=1}^{N} \pi_k b_k(o_1)$ (9)

From (5), (7), (8) and $t = 2, \cdots, T$, $1 \le j \le N$, the following is obtained

$$\hat{\alpha}_t(j) = P\left(q_t = s_j \mid o_1, o_2, \ldots, o_t\right)$$

$$= \frac{P\left(q_t = s_j, o_1, o_2, \ldots, o_t\right)}{P\left(o_1, o_2, \ldots, o_t\right)}$$ (10)

$$= \frac{\sum_{i=j-1}^{j} \alpha_{t-1}(i) a_{ij} b_j(o_t)}{r_t}$$

where $r_t = \sum_{l=1}^{N} \sum_{k=l-1}^{l} \alpha_{t-1}(k) a_{kl} b_l(o_t)$ (11)

Note that normalisation factors r_i ensure the modified forward probabilities sum to one.

The normalised initial and recursive forward probabilities given by (8) and (10) are used in the current algorithm for computing the recursive state entropy of a first-order BHMM. The majority of recursive entropy computations are performed in the recursion phases that results in $O(TN)$ operations. It only requires memory space of $2N$ since the memory space is independent of the length of the observational sequence. A numerical illustration is shown in section 3.2 Hernando et al. (2005) use a trellis diagram to show the structure for the recursive computation of the entropy. The time is shown on the horizontal axis. This diagram contains both states and its entropy for each time step. The Figure 1 is an example of a trellis diagram displaying the structure for the recursive computation of the entropy for a generalised first-order HMM with 4 states and a length T of observational sequence.

	o_1		o_2		o_3	$\cdots$	o_T
State 1	$H_1(1)$ $\hat{\alpha}_1(1)$		$H_2(1)$ $\hat{\alpha}_2(1)$		$H_3(1)$ $\hat{\alpha}_3(1)$		$H_T(1)$ $\hat{\alpha}_T(1)$
State 2	$H_1(2)$ $\hat{\alpha}_1(2)$		$H_2(2)$ $\hat{\alpha}_2(2)$		$H_3(2)$ $\hat{\alpha}_3(2)$	$\cdots$	$H_T(2)$ $\hat{\alpha}_T(2)$
State 3	$H_1(3)$ $\hat{\alpha}_1(3)$		$H_2(3)$ $\hat{\alpha}_2(3)$		$H_3(3)$ $\hat{\alpha}_3(3)$		$H_T(3)$ $\hat{\alpha}_T(3)$
State 4	$H_1(4)$ $\hat{\alpha}_1(4)$		$H_2(4)$ $\hat{\alpha}_2(4)$		$H_3(4)$ $\hat{\alpha}_3(4)$		$H_T(4)$ $\hat{\alpha}_T(4)$

Figure 1. The evolution of a trellis structure with 4 states and a length of T observational sequence

Due to the structure of state transition for BHMM, a modified trellis structure diagram is introduced, namely a decreasing-ladder trellis structure diagram which provides a clear picture on how the entropy associated with the optimal state sequence of BHMM is determined. Figure 2 is an example of a decreasing-ladder trellis structure diagram with 4 states and 6 observations.

	o_1	o_2	o_3	o_4	o_5	o_6
State 1	$H_1(1)$ $\hat{\alpha}_1(1)$	$H_2(1)$ $\hat{\alpha}_2(1)$	$H_3(1)$ $\hat{\alpha}_3(1)$			
State 2	$H_1(2)$ $\hat{\alpha}_1(2)$	$H_2(2)$ $\hat{\alpha}_2(2)$	$H_3(2)$ $\hat{\alpha}_3(2)$	$H_4(2)$ $\hat{\alpha}_4(2)$		
State 3	$H_1(3)$ $\hat{\alpha}_1(3)$		$H_3(3)$ $\hat{\alpha}_3(3)$	$H_4(3)$ $\hat{\alpha}_4(3)$	$H_5(3)$ $\hat{\alpha}_5(3)$	
State 4	$H_1(4)$ $\hat{\alpha}_1(4)$				$H_5(4)$ $\hat{\alpha}_5(4)$	$H_6(4)$ $\hat{\alpha}_6(4)$
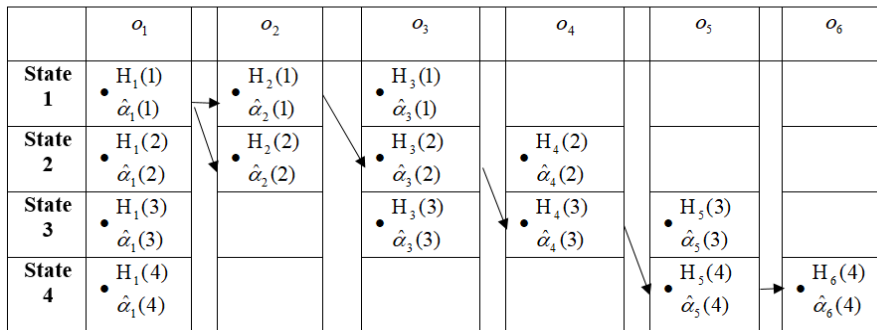
Figure 2. The evolution of a decreasing-ladder trellis structure with 4 states and 6 observations

STATE ENTROPY COMPUTATION FOR A FIRST-ORDER BHMM

The optimal state sequence can be obtained from a given BHMM model's parameters and observational sequence using the conventional Viterbi algorithm. This algorithm provides the solution along with its likelihood. This likelihood probability can be determined as follows.

$$P(q_1, q_2, \cdots, q_T | o_1, o_2, \cdots, o_T) = \frac{P(q_1, q_2, \cdots, q_T, o_1, o_2, \cdots, o_T)}{P(o_1, o_2, \cdots, o_T)}$$

Hence, this probability can be used as a measure of quality of the solution. The higher the probability of the "solution", the better it is. However, this conventional algorithm does not provide a clear process for obtaining the optimal state sequence and its likelihood probability. Alternatively, entropy is proposed for measuring the quality of the state sequence and using a decreasing-ladder trellis structure to illustrate the process of computing the possible state sequence along with the uncertainty.

The entropy H(X) of a random variable of X, is a measure of its uncertainty (Cover & Thomas, 2006). This concept for quantifying the uncertainty of the state sequence tracked by the BHMM's parameter is applied and a given single observational sequence. This entropy can be viewed as a measure of how well the parameters for generating a certain observational sequence. The entropy of the state sequence equals to 0 if there is only one state sequence that could have generated the observation sequence as there is no uncertainty in the solution. The higher this entropy, the higher the uncertainty involved in tracking the hidden state.

Entropy of a discrete random variable is defined as follows (Cover & Thomas, 2006):

Definition 3. The entropy H(X) of a discrete random variable X with a probability mass function $P(X=x)$ is defined as

$$H(X) = -\sum_{x \in X} P(x) \log_2 P(x)$$ (12)

When the log has a base of 2, the unit of the entropy is bits and it is noted that $0 \log 0 = 0$

From (12), the entropy of the distribution for all possible state sequences is defined as follows:

$$H(q_{1:T}|o_1, o_2, \cdots, o_T) = -\sum_S \left[P(q_{1:T} = s_{1:T}|o_1, o_2, \cdots, o_T) \log_2 P(q_{1:T} = s_{1:T}|o_1, o_2, \cdots, o_T) \right]$$ (13)

BHMM Entropy-based Forward Algorithm (BHMM-EFA)

For this algorithm, an intermediate variable that is state entropy, $H_t(s_j)$ is required. The state entropy, $H_t(s_j)$ can be computed recursively using the previous one that is $H_{t-1}(s_j)$ (Hernando et al., 2005).

The state entropy for a first-order BHMM is defined as follows:

Definition 4. The state entropy, $H_t(s_j)$ in a first-order BHMM, is the entropy of all the state sequences that lead to state of at time , given the observations $o_1, o_2, \cdots, o_t$. It can be denoted as

$$H_t(s_j) = H(q_{1:t-1}|q_t = s_j, o_{1:t})$$ (14)

From (14) and $t=1$, the initial state entropy variable is expressed as

$$H_1(s_j) = 0$$ (15)

From (14), and (15), the recursion on the entropies for $t = 2, \ldots, T$, $1 \le i, j \le N$, and $i = j - 1, j$ is obtained as below

$$
\begin{aligned}
H_t(s_j) &= H(q_{1:t-1}|q_t = s_j, o_{1:t}) \\
&= H(q_{1:t-2}, q_{t-1}|q_t = s_j, o_{1:t}) \\
&= H(q_{t-1}|q_t = s_j, o_{1:t}) + H(q_{1:t-2}|q_{t-1}, q_t = s_j, o_{1:t})
\end{aligned}
$$ (16)

where

$$H(q_{t-1}|q_t = s_j, o_{1:t}) = -\sum_{i=j-1}^{j} \left[P(q_{t-1} = s_i|q_t = s_j, o_{1:t}) \log_2 \left(P(q_{t-1} = s_i|q_t = s_j, o_{1:t}) \right) \right]$$

and

$$H\left(q_{1:t-2}\big|q_{t-1},q_t=s_j,o_{1:t}\right)=\sum_{i=j-1}^{j}\left[P\left(q_{t-1}=s_i\big|q_t=s_j,o_{1:t}\right)H\left(q_{1:t-2}\big|q_{t-1}=s_i,q_t=s_j,o_{1:t}\right)\right]$$

$$=\sum_{i=j-1}^{j}\left[P\left(q_{t-1}=s_i\big|q_t=s_j,o_{1:t}\right)H_{t-1}(s_i)\right]$$

The auxiliary probability $P\left(q_{t-1}=s_i\big|q_t=s_j,o_{1:t}\right)$ where $i=j-1,j$ is required for our algorithm. It can be computed as follows:

$$P\left(q_{t-1}=s_i\big|q_t=s_j,o_{1:t}\right)=P\left(q_{t-1}=s_i\big|q_t=s_j,o_t,o_{1:t-1}\right)$$

$$=\frac{P\left(q_{t-1}=s_i,q_t=s_j,o_{1:t-1},o_t\right)}{P\left(q_t=s_j,o_{1:t-1},o_t\right)}$$

$$=\frac{P\left(q_t=s_j,o_t\big|q_{t-1}=s_i,o_{1:t-1}\right)P\left(q_{t-1}=s_i\big|o_{1:t-1}\right)}{P\left(q_t=s_j,o_t\big|o_{1:t-1}\right)}$$

$$=\frac{P\left(o_t\big|q_t=s_j\right)P\left(q_t=s_j\big|q_{t-1}=s_i\right)P\left(q_{t-1}=s_i\big|o_{1:t-1}\right)}{P\left(o_t\big|q_t=s_j\right)P\left(q_t=s_j\big|o_{1:t-1}\right)} \tag{17}$$

$$=\frac{P\left(q_t=s_j\big|q_{t-1}=s_i\right)P\left(q_{t-1}=s_i\big|o_{1:t-1}\right)}{\sum_{k=j-1}^{j}P\left(q_t=s_j\big|q_{t-1}=s_k\right)P\left(q_{t-1}=s_k\big|o_{1:t-1}\right)}$$

$$=\frac{a_{ij}\hat{\alpha}_{t-1}(i)}{\sum_{k=j-1}^{j}a_{kj}\hat{\alpha}_{t-1}(k)}$$

For the final process, $H\left(q_{1:T}\big|o_{1:T}\right)$ is computed which can be expanded as follows:

$$H\left(q_{1:T}\big|o_{1:T}\right)=H\left(q_{1:T-1},q_T\big|o_{1:T}\right)=H\left(q_{1:T-1}\big|q_T,o_{1:T}\right)+H\left(q_T\big|o_{1:T}\right)$$

$$=\sum_{i=1}^{N}H_T(s_i)\hat{\alpha}_T(i)-\sum_{i=1}^{N}\hat{\alpha}_T(i)\log_2\left(\hat{\alpha}_T(i)\right) \tag{18}$$

The basic entropy concept in (12) and the following basic properties of BHMM are used for proving lemma 1. According to a first-order BHMM, state q_{t-r}, $r \geq 2$ and q_t are statistically independent given q_{t-1}. The same applies to q_{t-r}, $r \geq 2$ and o_t are statistically independent given q_{t-1}.

The following proof is due to Hernando et. al. (2005).

Lemma 1: For a first-order BHMM, the entropy of the state sequence up to time t - 2, given the states at time t - 1 and the observations up to time t - 1, is conditionally independent on the state and observation at time t

$$H_{t-1}(s_i)=H\left(q_{1:t-2}\big|q_{t-1}=s_i,q_t=s_j,o_{1:t}\right)$$

Proof:

$$H\left(q_{1:t-2}\middle|q_{t-1}=s_i,q_t=s_j,o_{1:t}\right)$$

$$=H\left(q_{1:t-2}\middle|q_{t-1}=s_i,o_{1:t-1},q_t=s_j,o_t\right)$$

$$=H\left(q_{1:t-2}\middle|q_{t-1}=s_i,o_{1:t-1}\right)$$

$$=H_{t-1}\left(s_i\right)$$

The entropy algorithm for BHMM is based on normalised forward recursion probability variable, state entropy recursion variable and auxiliary probability.

From (8), (10), (15), (16), (17) and (18), the algorithm of entropy computation for a first-order BHMM is constructed as below:

1. Initialisation: for $1 \le j \le N$,

$$H_1\left(s_j\right)=0$$

$$\hat{\alpha}_1(j)=\frac{\pi(j)b_1(o_1)}{\sum_{k=1}^{N}\pi(k)b_k(o_1)}$$

2. Recursion:

    For $t=2:T$

        For $j=1:N-1$

            If $\hat{\alpha}_{t-1}(j)>0$

            $w=j$

                For $j=w:w+1$

$$\hat{\alpha}_t(j)=\frac{\sum_{i=j-1}^{j}\hat{\alpha}_{t-1}(i)a_{ij}b_j(o_t)}{\sum_{l=1}^{N}\sum_{k=l-1}^{l}\hat{\alpha}_{t-1}(k)a_{kl}b_l(o_t)}$$

                    For $i=j-1:j$

$$P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)=\frac{a_{ij}\hat{\alpha}_{t-1}(i)}{\sum_{k=j-1}^{j}a_{kj}\hat{\alpha}_{t-1}(k)}$$

            End

$$H_t(s_j) = \sum_{i=j-1}^{j} H_{t-1}(s_i) P(q_{t-1} = s_i | q_t = s_j, o_{1:t})$$
$$- \sum_{i=j-1}^{j} P(q_{t-1} = s_i | q_t = s_j, o_{1:t}) \log_2 P(q_{t-1} = s_i | q_t = s_j, o_{1:t})$$

End

End if

End

If $\hat{\alpha}_{t-1}(N) > 0$

$$j = N$$

$$\hat{\alpha}_t(j) = \frac{\sum_{i=j-1}^{j} \hat{\alpha}_{t-1}(i) a_{ij} b_j(o_t)}{\sum_{l=1}^{N} \sum_{k=l-1}^{l} \hat{\alpha}_{t-1}(k) a_{kl} b_l(o_t)}$$

For $i = N - 1 : N$

$$P(q_{t-1} = s_i | q_t = s_j, o_{1:t}) = \frac{a_{ij} \hat{\alpha}_{t-1}(i)}{\sum_{k=j-1}^{j} a_{kj} \hat{\alpha}_{t-1}(k)}$$

End

$$H_t(s_j) = \sum_{i=j-1}^{j} H_{t-1}(s_i) P(q_{t-1} = s_i | q_t = s_j, o_{1:t})$$
$$- \sum_{i=j-1}^{j} P(q_{t-1} = s_i | q_t = s_j, o_{1:t}) \log_2 P(q_{t-1} = s_i | q_t = s_j, o_{1:t})$$

End if

End

3. Termination

$$H(q_{1:T} | o_{1:T}) = \sum_{i=1}^{N} H_T(s_i) \hat{\alpha}_T(i) - \sum_{i=1}^{N} \hat{\alpha}_T(i) \log_2(\hat{\alpha}_T(i))$$

Illustration

A BHMM is considered with the following model parameters $\pi = (\lambda, A, B)$ where π is the initial transition probability vector, A is the state transition probability matrix and B is the emission probability matrix.

$$\pi = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \; A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \; B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The algorithm for computing the state entropy is applied based on the observational sequence $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$. The result is shown as follows.

	$o_1 = 1$	$o_2 = 2$	$o_3 = 2$	$o_4 = 3$	$o_5 = 5$	$o_6 = 5$
State 1	$H_1(1) = 0$ $\hat{\alpha}_1(1) = 1$	$H_2(1) = 0$ $\hat{\alpha}_2(1) = 0$				
State 2	$H_1(2) = 0$ $\hat{\alpha}_1(2) = 0$	$H_2(2) = 0$ $\hat{\alpha}_2(2) = 1$	$H_3(2) = 0$ $\hat{\alpha}_3(2) = 1$	$H_4(2) = 0$ $\hat{\alpha}_4(2) = 0.5$	$H_5(2) = 0$ $\hat{\alpha}_5(2) = 0$	
State 3	$H_1(3) = 0$ $\hat{\alpha}_1(3) = 0$		$H_3(3) = 0$ $\hat{\alpha}_3(3) = 0$	$H_4(3) = 0$ $\hat{\alpha}_4(3) = 0.5$	$H_5(3) = 1$ $\hat{\alpha}_5(3) = 0$	
State 4	$H_1(4) = 0$ $\hat{\alpha}_1(4) = 0$				$H_5(4) = 0$ $\hat{\alpha}_5(4) = 1$	$H_6(4) = 0$ $\hat{\alpha}_6(4) = 1$
Entropies	0	0	0	1	0	0

Figure 3. The evolution of the decreasing-ladder trellis structure with the observation sequence $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$

The total entropy after each time step is displayed at the bottom of Figure 3. For example, after receiving the fourth observation, i.e. $o_{1:4} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3)$, it has produced two possible state sequences which are $q_{1:4} = (q_1 = state1, q_2 = state2, q_3 = state2, q_4 = state2)$ and $q_{1:4} = (q_1 = state1, q_2 = state2, q_3 = state2, q_4 = state3)$ as shown by the arrows. Each possible state sequence has a probability of 0.5 and hence the entropy is 1 bit. After receiving the fourth observation, intermediate entropy and normalized forward probability for state 2 and state 3 are computed whereas both values for state 1 and state 4 are zeros due to the structure of the state transition. For the fifth observation, the intermediate entropy and normalized forward probability for state 2, state 3 and state 4 are computed whereas both values for state 1 are zeros due to the structure of the state transition. The following are the required computations after receiving the fourth observation, i.e. $o_{1:4} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3)$ in order to compute the total entropy at $t = 4$. The model has produced only one possible state sequence that is $q_{1:3} = (q_1 = state1, q_2 = state2, q_3 = state2)$ with a probability of 1 after receiving the third

observation. The total entropy is 0 at which $t = 3$ indicates that there is no uncertainty. Due to the structure of the state transition, at time $t = 4$, only two state transitions occurred that are the transition from state 2 at $t = 3$ to state 2 at $t = 4$ and state 2 at $t = 3$ to state 3 at $t = 4$.

From (10) and $j = 2$, we obtain

$$\hat{\alpha}_4(2) = \frac{\hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4)}{\hat{\alpha}_3(1)a_{11}b_1(o_4) + \hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4) + \hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4) + \hat{\alpha}_3(3)a_{34}b_4(o_4) + \hat{\alpha}_3(4)a_{44}b_4(o_4)}$$

$$= \frac{0(0.5)(0.5) + 1(0.5)(0.5)}{0(0.5)(0) + 0(0.5)(0.5) + 1(0.5)(0.5) + 1(0.5)(0.5) + 0(0.5)(0.5) + 0(0.5)(0) + 0(1)(0)} = 0.5$$

From (10) and $j = 3$, we obtain

$$\hat{\alpha}_4(3) = \frac{\hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4)}{\hat{\alpha}_3(1)a_{11}b_1(o_4) + \hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4) + \hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4) + \hat{\alpha}_3(3)a_{34}b_4(o_4) + \hat{\alpha}_3(4)a_{44}b_4(o_4)}$$

$$= \frac{1(0.5)(0.5) + 0(0.5)(0.5)}{0(0.5)(0) + 0(0.5)(0.5) + 1(0.5)(0.5) + 1(0.5)(0.5) + 0(0.5)(0.5) + 0(0.5)(0) + 0(1)(0)} = 0.5$$

From (17) and $i = 1, j = 2$, we obtain

$$P(q_3 = s_1 | q_4 = s_2, o_{1:4}) = \frac{a_{12}\hat{\alpha}_3(1)}{a_{12}\hat{\alpha}_3(1) + a_{22}\hat{\alpha}_3(2)} = \frac{0.5(0)}{0.5(0) + 0.5(1)} = 0,$$

From (17) and $i = 2, j = 2$, we obtain

$$P(q_3 = s_2 | q_4 = s_2, o_{1:4}) = \frac{a_{22}\hat{\alpha}_3(2)}{a_{12}\hat{\alpha}_3(1) + a_{22}\hat{\alpha}_3(2)} = \frac{0.5(1)}{0.5(0) + 0.5(1)} = 1,$$

From (17) and $i = 2, j = 3$, the following is obtained

$$P(q_3 = s_2 | q_4 = s_3, o_{1:4}) = \frac{a_{23}\hat{\alpha}_3(2)}{a_{23}\hat{\alpha}_3(2) + a_{33}\hat{\alpha}_3(3)} = \frac{0.5(1)}{0.5(1) + 0.5(0)} = 1,$$

From (17) and $i = 3, j = 3$ we obtain

$$P(q_3 = s_3 | q_4 = s_3, o_{1:4}) = \frac{a_{33}\hat{\alpha}_3(3)}{a_{23}\hat{\alpha}_3(2) + a_{33}\hat{\alpha}_3(3)} = \frac{0.5(0)}{0.5(1) + 0.5(0)} = 0,$$

From (16), $t = 4$ and $s_j = 2$, the following is obtained

$$H_4(2) = H_3(1)P(q_3 = s_1 | q_4 = s_2, o_{1:4}) + H_3(2)P(q_3 = s_2 | q_4 = s_2, o_{1:4}) -$$

$$P(q_3 = s_1 | q_4 = s_2, o_{1:4}) \log P(q_3 = s_1 | q_4 = s_2, o_{1:4}) - P(q_3 = s_2 | q_4 = s_2, o_{1:4}) \log P(q_3 = s_2 | q_4 = s_2, o_{1:4})$$

$$= 0(0) + 0(1) - 0 \log 0 - 1 \log 1 = 0$$

From (16), and $t = 4$ and $s_j = 3$, we obtain

$$\mathrm{H}_4(3) = \mathrm{H}_3(2)P\big(q_3 = s_2 | q_4 = s_3, o_{1:4}\big) + \mathrm{H}_3(3)P\big(q_3 = s_3 | q_4 = s_3, o_{1:4}\big) -$$

$$P\big(q_3 = s_2 | q_4 = s_3, o_{1:4}\big) \log P\big(q_3 = s_2 | q_4 = s_3, o_{1:4}\big) - P\big(q_3 = s_3 | q_4 = s_3, o_{1:4}\big) \log P\big(q_3 = s_3 | q_4 = s_3, o_{1:4}\big)$$

$$= 0(1) + 0(0) - 1 \log 1 - 0 \log 0 = 0$$

From (18) and $t = 4$, the total entropy is obtained

$$\mathrm{H}\big(q_{1:4} | o_{1:4}\big) = \mathrm{H}_4(s_1)\hat{\alpha}_4(1) + \mathrm{H}_4(s_2)\hat{\alpha}_4(2) + \mathrm{H}_4(s_3)\hat{\alpha}_4(3) + \mathrm{H}_4(s_4)\hat{\alpha}_4(4) - \hat{\alpha}_4(1)\log_2\big(\hat{\alpha}_4(1)\big) -$$

$$\hat{\alpha}_4(2)\log_2\big(\hat{\alpha}_4(2)\big) - \hat{\alpha}_4(3)\log_2\big(\hat{\alpha}_4(3)\big) - \hat{\alpha}_4(4)\log_2\big(\hat{\alpha}_4(4)\big)$$

$$= 0(0) + 0(0.5) + 0(0.5) + 0(0) - 0\log_2 0 - 0.5\log_2 0.5 - 0.5\log_2 0.5 - 0\log_2 0 = 1 \text{ bit.}$$

After receiving the sixth observation, i.e. $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$, the model has produced only one possible state sequence that is $q_{1:6} = (q_1 = state1, q_2 = state2, q_3 = state2, \quad q_4 = state3, q_5 = state4, q_6 = state4)$ with a probability of 1 and hence the total entropy is 0 which indicates that there is no uncertainty.

CONCLUSION AND FUTURE WORK

An algorithm for computing the state entropy for a first-order BHMM was introduced. This algorithm needs to run with Viterbi algorithm in tracking the state sequence as well as the entropy of the distribution of the state sequence. A decreasing-ladder trellis structure is introduced in this paper. The algorithm can be represented in this structure diagram shows a clear process of obtaining the uncertainty present in state sequence. This algorithm requires $\mathrm{O}(TN)$ calculations and performs the state computation linearly with the length of observational sequence and the number of hidden states of BHMM. This research can be also extended for a discrete as well as a continuous high-order BHMM. For any generalised high-order BHMM, the historical state information is well explored for predicting the next state and these models are widely used in speech recognition.

REFERENCES

Ciriza, V., Donini, L., Durand, J., & Girard, S. (2011). *Optimal timeouts for power management under renewal or hidden Markov processes for requests.* Technical Report. Retrieved from http://hal.inria.fr/hal-00412509/en.

Cover, T. M., & Thomas. J. A. (2006). *Elements of Information Theory* (2nd Ed.). New York: Wiley.

Gales, M., & Young, S. (2008). The application of hidden Markov Models in speech recognition. *Foundations and Trends in Signal Processing, 1*(3), 195–304.

Hernando, D., Crespi, V., & Cybenko, G. (2005). Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE Transactions Information Theory, 51*(7), 2681 – 2685.

Ilic, V. M. (2011). *Entropy semiring forward-backward algorithm for HMM entropy computation.* Retrieved from ArXiv:1108.0347.

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics, 33*(3), 251–272.

Löytynoja, A., & Milinkovitch, M. C. (2003). A Hidden Markov model for progressive multiple alignment. *Bioinformatics, 19*(12), 1505–1513.

Mann, G. S., & McCallum, A. (2007). Efficient computation of entropy gradient for semi-supervised conditional random fields. In *Human Language Technologies 2007:* The *Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume. S*hort Papers on XX, NAACL '07 (pp. 109–112). Morristown, NJ, USA. Association for Computational Linguistics.

Nogueiras, A., Moreno, A., Bonafonte, A., & Marino, J. (2001). Speech emotion recognition using hidden markov models. In *Eurospeech 2001, Poster Proceedings* (pp. 2679-2682). Scandinavia.

Proakis, J. G., &Salehi, M. (2002). *Communications System Engineering.* Upper Saddler River, New Jersey: Prentice-Hall.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* (pp. 257–286). IEEE.

Regad, L., Guyon, F., Maupetit, J., Tufféry, P., & Camproux, A. C. (2008). A Hidden Markov Model applied to the protein 3D structure analysis. Computational Statistics and Data Analysis, 52(6), 3198–3207.

A Portable and Low Cost Multi-sensor for Real Time Remote Sensing of Water Quality in Agriculture

Sandeep Bansal[1]* and G. Geetha[2]

[1]*Department of Electronics and Communication Engineering, Lovely Professional University, Jalandhar, 144401, Punjab, India*
[2]*Department of Research and Development, Lovely Professional University, Jalandhar, 144401, Punjab, India*

ABSTRACT

Water is an important natural resource for all living organisms. Due to increase in population, industrial magnification and urbanisation, water gets contaminated these days. The aim of present study is to design a low cost and reliable system for the monitoring of real time water quality. This study includes monitoring of physiochemical parameters such as pH, Temperature, Turbidity and Total Dissolved Solids (TDS). Microcontroller based multi sensor system can measure the said parameters for detecting water contamination and incorporates communication technology for further processing and alerts. Data communication module can transmit the data received from system to intended user for making alerts regarding water quality. User can check the water quality information perpetually even from far away and he or she can take several safety measures to prevent health hazards. Facile design and low cost make this system captivating enough for large scale deployment.

Keywords: Automation, contamination detection, multi sensor system, turbidity sensor, water quality monitoring

INTRODUCTION

With rapid economic growth and industrialisation, there are serious concerns over their impact on the environment. Water pollution is one of the major concerns (CPCB, 2007). Water quality can be verified based upon basic parameters, such as pH, Turbidity, Biological Oxygen Demand (BOD), Temperature, TDS, Chemical Oxygen Demand (COD), nitrate, nitrite, and phosphate etc (Lambrou et al., 2014).

Water quality monitoring is utilised to screen and decide the kind of water contamination, the amount of various water contaminants and the changing tendency (Fengyun, 2010). Level of contamination can be identified in time with such monitoring on water pollution, so that emergency situations can be resolved.

Traditionally water quality monitoring is based on laboratory methods, which includes manual sampling of water and to test these samples in dedicated labs (Das & Jain, 2017).

But these methods are very slow and time consuming due to transportation of samples from far places (Cloete et al., 2014). Also, more manpower is required in these methods, which makes them less suitable for effective monitoring. Hence automatic water quality monitoring is presented for speedy and accurate monitoring of water resources.

Looking at all these limitations, there is an obvious requirement of change in traditional monitoring system. So, our contribution through this research is to develop an intelligent, low cost and reliable system to provide accurate water quality information to intend user at their home using mobile device. Also, in agriculture applications, the requirement of good quality water is always there to enhance quality and productivity.

RELATED WORK

Postolache et al. (2002) developed a multi sensor that can evaluate water quality parameters. Additionally, this sensor has the capability to maintain online database of information. Authors have considered pH, temperature and turbidity of water. This GSM-enabled sensor is advantageous in avoiding manual intervention as well as enhance real time water quality tracking.

Breijo et al. (2002) designed a system to measure temperature, pH, conductivity, dissolved oxygen, turbidity, ORP and diverse ions. Basically, they focused on disadvantages of lab methods that are too slow and that continuous measurement is not possible in laboratory process. They addressed these limitations. Subsidiary et al. (2006) proposed a set of water parameters to be considered in healthcare units as well as methods to measure these parameters. Additionally, they developed computerised solutions for analyses of various water samples used in health sector.

O'Flynn et al. (2007) developed a "Smartcoast" sensor to evaluate various parameters i.e. phosphate, temperature, turbidity, pH, dissolved oxygen, conductivity and water level. They have employed wireless sensor network to incorporate plug and play feature in the sensor system. Jiang et al. (2009) designed a sensor to automate water quality monitoring process. The designed system can measure pH (1-14) and temperature (0-80°C) of water.

Regan et al. (2009) had designed a low power solution to measure various water quality parameters, namely temperature, pH, turbidity, conductivity and dissolved oxygen.

Tie-zhu and Le (2010) had developed a system to first measure the parameters i.e. temperature, pH and dissolved oxygen and then information is sent to monitoring centre using GPRS technology. Finally, proper analysis is performed at monitoring centre. Meng et al. (2010) had developed a hand-held device to measure 17 water parameters based on sensor of

YSI6600. The whole system is battery powered using MSP430F149. Basic advantage of this sensor was its speed, precision and small error.

Liu et al. (2011) had used Principal Component Analysis technique to find out important water parameters that may cause variances in the quality. They began with 13 parameters and came up with3 of them at the end of the process.

Nasirudin et al. (2011) employed economic power solution to evaluate water quality. They used green power source i.e. Solar energy to optimise power consumption and measured four parameters, namely pH, turbidity, temperature and dissolved oxygen. Deqing et al. (2012) provided low cost solution for water quality analysis i.e. they had employed DS18B20 to prepare their own sensor for temperature, pH and turbidity. Chou et al. (2012) proposed a solution to enhance life time of pH and chlorine sensor. Vaddadi (2012) had focused on quality of fish pond water. They had considered temperature, pH, conductivity, pressure and dissolved oxygen.

Damian et al. (2012) had discussed various parameters and their importance in various fields. They had suggested that salinity was the important parameter to differentiate between ocean water and fresh water. Turbidity is the parameter to specify changes in water quality and finally conductivity is inverse of resistance and it is also an important water quality parameter.

Miskam et al. (2013) analysed water quality in paddy fields that may affect level of production. They had considered temperature and dissolved oxygen.

WATER QUALITY PARAMETERS

Selection of WQ parameters is based upon the guidelines for water quality regulation by WHO (Gorchev et al., 2011) and other reputed organisations like USEPA, EU (Hall et al., 2007). These are the esteemed organisations, which set the permissible limits for all these water quality parameters for various purposes like human consumption, agriculture or other needs. Also, they designate which chemical or physical indicator parameters must be considered or targeted on priority basis and tested on daily basis to protect the health of the users and to make sure the water is whole some and safe. Mostly physiochemical parameters have been concentrated for monitoring.

List of some water quality parameters is listed in Table 1 along with the permissible limits where "-" represents unitless quantities.

Table 1
Suggested Water Quality Parameters for monitoring purpose (Lambrou et al., 2014)

Parameters	Units	Permissible limits
pH	-	6.5-8.5
Taste	-	Unobjectionable
TDS	mg/l or ppm	<= 500
Turbidity	NTU	0 – 5
Temperature	°C	Depends on application
Dissolved Oxygen	mg/l	Depends on application
Electrical Conductivity	µS/cm	500 – 1000

Parameters selected for the proposed study were Temperature, pH and Turbidity as these parameters were sufficient for the basic requirements of water quality monitoring for general purpose.

Temperature

Measurement of hotness or coldness is generally referred to as temperature. It can be measured in degree Celsius and Fahrenheit. Temperature is important because it impacts on water chemistry. At higher temperature, the rate of chemical reactions generally increases. Its effects in water are dissolved oxygen, diffusion rate or gases, marine life. The solubility of oxygen decreases as temperature increases and it affects the dissolved oxygen concentration. Rather than cold water, some doctors recommend that room or body temperature water is better to drink. Warm water helps aid in digestion, help your body detox, settle an upset stomach. Cold water tastes better, raises metabolism, lowers fever.

Due to changes in temperature, plants also have an impact on their growth. Some aquatic plants can withstand with cool water temperature, but some plants require warm temperature for growth. Tropical plants show hibiscus, orchids, gardenia show decrease in their growth, if the water temperature is below 21°C.

Aquatic life will also be affected when there are changes in water temperature. Water temperature also influences the movement of fish, their metabolism and their respiration. If the temperature of water is higher (warm water) there will be an increase in metabolism, movement and respiration rate of the fish. If the temperature of the water is less, the fish will become inactive. Temperature below 5°C can be harmful to the species. This is also applicable to other insects and small aquatic organisms.

Causes of changes in water temperature:

- Sunlight/Solar radiation
- Deforestation
- Seasonal changes
- Turbidity

Effects of unusual water temperature:

- Increase or decrease in temperature of water is strongly influenced by the solubility of dissolved oxygen. In warm water, less gas can be dissolved compared to cold water.
- Increase in water temperature may lead to reduction of oxygen amount for aquatic organisms to survive, ultimately leads to threat for marine life.
- Extreme water temperature i.e. too hot or too cold organisms become stressed, reducing their resistant to diseases and pollutants.

Turbidity

It is the measure of cloudiness or haziness in water. Cloudiness is mainly due to presence of air bubbles in water, after a few seconds they clean up. This cloudiness is caused due to small and tiny suspended particles, solids, organic and inorganic matter etc., which are not visible to our naked eye. These suspended particles mainly include clay, sand, silt, microscopic plants, bacteria/germs and chemical precipitates etc. Measuring turbidity is an important factor when we need to determine water quality. Higher levels of turbidity are caused mainly due to higher levels of bacteria and parasite which are sometimes attached to the dirt in water. Higher turbidity in drinking water is an indicator that water may contain disease causing organisms. Increased level of turbidity may lead to gastrointestinal diseases, health issues for new-borns and people with low immune systems such as those with HIV, and those undergoing chemotherapy among others.

Governments have set allowable levels for turbidity in drinking water. According to WHO, maximum permissible level value of turbidity should not be more than 5NTU. It should be ideally below 1NTU. The units of turbidity are NTU which stands for Nephelometric turbidity units and JTU which stands for Jackson turbidity units. Nephelometer is an electronic instrument in which NTU is measured.

Effects of unusual turbidity:

- Cost of water treatment may increase for drinking and food processing.
- High level of turbidity leads to gastrointestinal diseases.
- Damages valves and tapes because the high turbidity water which flows through them will fill them with mud and silt.
- Turbidity indicates the presence of disease causing organisms which causes symptoms, such as nausea, cramps and headache.

pH

pH is defined as $-\log[H+]$ and is the measure of concentration of hydrogen ion in water. Higher concentration of hydrogen ions indicates lower values of pH and lower concentration of hydrogen ions indicate higher values of ph. Pure water pH value is 7. pH values range from 0 to 14, where 0 is most acidic and 14 is most alkaline. pH of 7 represents neutral. According to WHO standards, recommended levels of pH are 6.5 to 8.5. The pH of water regulates the solubility in water.

Higher pH levels (>9.5) and lower pH levels (<4.5) are not suitable for most marine organisms. Aquatic bodies are more sensible and may die at low pH levels below 5. At high pH, toxic ammonia will be formed from ammonia. Also, higher pH levels which are greater than 9 can harm fish. Table 2 shows the effect on pH value due to natural/Man made factors.

Table 2
Factors effecting pH value (CPCB, 2007)

Factors	Effects on pH
Acid rain	If increases, pH level decreases
Temperature	If increases, pH level decreases
Dissolved minerals	If more, pH level increases
Concentration of carbon dioxide in water	If increases, pH level decreases
Releases from industries	°C
Depends on the acids and bases released	mg/l
Electrical Conductivity	µS/cm

Effects of lower pH levels:

- Water becomes acidic and soft.

- Heavy metal effects and hold metal ions such as copper, iron, lead, manganese and zinc.

- It can be a root for health problems like respiratory failure, seizures and shock.

- Have metallic taste or sour taste.

- Low pH levels can stain laundry and stains on drains and stinks (blue-green colour).

- Lower pH level can corrode pipes.

Effects of higher pH levels:

- Water becomes hard and cause aesthetic problems.

- Have alkali taste or bitter taste.

- Scale deposits can be formed on utensils, dishes and laundry basins.

- It can be a root of health problems, such as arrhythmia, coma and if potassium levels decrease, it can lead to problems in kidneys, heart and digestive system.

SYSTEM DESIGN HARDWARE/COMPONENTS

Temperature Sensor

The proposed study had used DS18B20 (Temperature sensor) for water temperature measurement. It measures temperature from -55°C (min) to +125°C(max) and has an accuracy of ±0.5°C, over a range of temperature from -10°C to +85°C. It is water proof temperature sensor. Figure 1 shows the temperature sensor used for this work.

Figure 1. Temperature sensor DS18B20

Turbidity Sensor

Turbidity sensor sen0189 as shown in Figure 2 was used for measuring the turbidity levels in water. Turbidity sensors help to detect water quality in streams, rivers, wastewater measurements, research and laboratory measurements. Sen0189 provides output modes in both analog and digital signals. Its operating Voltage is 5V DC and its operating temperature is 5°C-90°C.
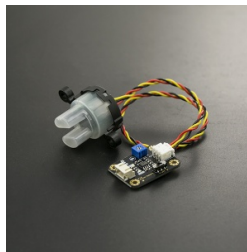
Figure 2. Turbidity Sensor

pH probe

pH electrodes or probes were used to measure the values of pH in proposed work. It measures potential difference (or voltage) of the solution in which it is submerged using Nernst equation and gives output of pH value. Figure 3 shows the pH probe used for this work.

Figure 3. pH probe

Wi-Fi Module

ESP8266 (as shown in Figure 4) was used to upload the sensor data to the web server and stored data can be used for analysis in future. Additionally, it sends the parameter calculated value to user mobile which can be further used to take actions. It operates on 3.3 VDC having 16 GPIO pins. It allows the connection of microcontroller with Wi-Fi network.
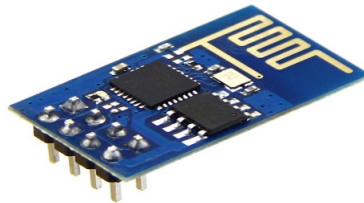
Figure 4. Wi-Fi module for data transmission

PROPOSED METHODOLGY

Following methodology have been adopted for carrying out this work as shown in Figure 5.
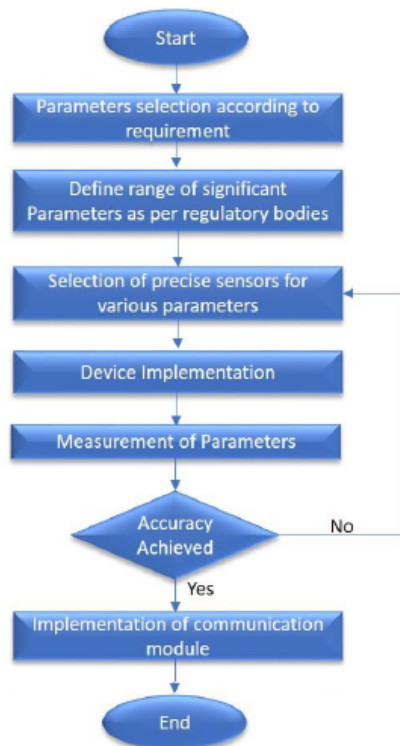
Figure 5. Flowchart of the system

Parameter Selection According to Requirement

The variables selection for any Water Quality (WQ) assessment system should be reflecting functions and issues in water (CPCB, 2007). Also, knowledge about polluting sources is an important part. The main objective of this step is to collect information about variables selection i.e. which WQ parameter to be monitored and for various uses, such as irrigation, drinking or livestock needs.)

Define Range of Significant Parameters as per Regulatory Bodies

Next, it is mandatory to decide the acceptable range of values that are allowed for parameters as per the standards established by organisations. Further, this decision needs to be taken by considering characteristics of water and pollution sources and amount of pollutants inserted in the water (Nasirudin et al., 2011).

Selection of Precise Sensors for Parameters

Based on the decision of selected WQ parameters to be monitored, the selection of sensors and techniques should be done while considering low cost and easy design. The situation is to make efficient use of technologies or sensors specifically to address the water quality issues.

Device Implementation

The implemented device greatly affects the quality of output. Therefore, proper attention is required while implementing the device by embedding sensors and other hardware which will be installed at the inlet and outlet of the source.

Measurement of Parameters

While measuring the parameters from intended source it is important to ensure the normal environmental conditions to make an actual measurement of selected parameters. Also, the other important things to be considered for this purpose are quantitative and qualitative sample.

Accuracy Check

Accuracy or Quality Assurance is one of the most desired thing while monitoring and analysing real time data. If device accuracy is achieved as required (can be mapped to laboratory checked samples) then the next step will proceed, otherwise, improvements in the circuit will be made for that parameter whose accuracy is not up to the mark.

Implementation of Communication Module

This module is responsible for sending parameter values to the mobile device using GSM/Wi-Fi technology. This will ensure the proper monitoring of water resources from any location.

RESULTS AND DISCUSSION

Figure 6 presents the implemented circuit as per proposed methodology. As circuit design was done after considering agriculture applications, we had chosen the sensors and made the hardware to detect WQM for dedicated application. Initially the setup had been installed at actual source and all readings observed by sensors are sent to mobile device (after processing) using Wi-Fi technology as represented in Figure 6. This feature helps update the users about current situation of water quality in the fields. If at any point the measured values go under non-acceptable range, intended user can also take actions accordingly.
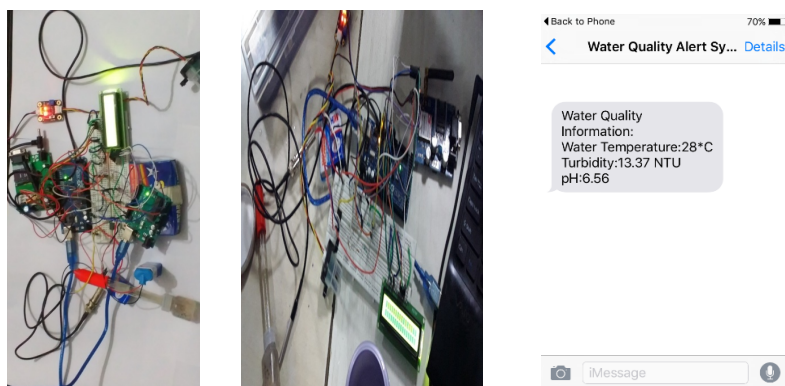
Figure 6. Working model snapshots

Finally, this circuit had been installed at local water body to carry out further experiments. Various samples had been taken from local water body used for agriculture purpose continuously for 15 days in pre-monsoon season. Figures 7, 8, and 9 show variations of pH, Temperature and Turbidity on different days respectively.
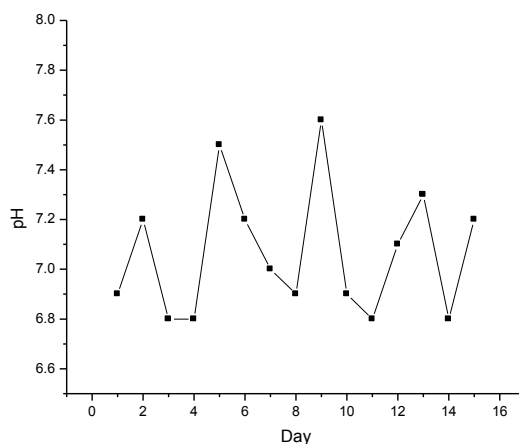
Figure 7. Average concentration of pH value
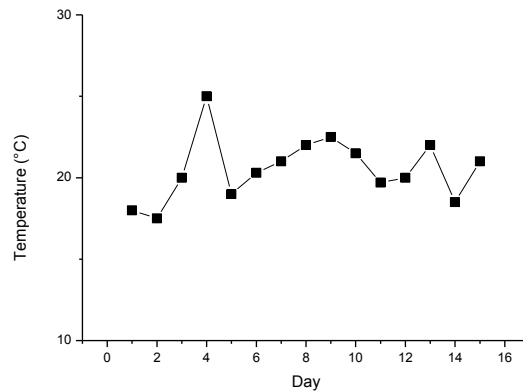
Figure 8. Average concentration of Temperature
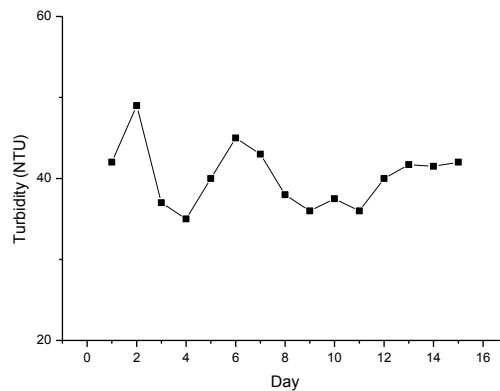
Figure 9. Average concentration of Turbidity

Overall pH value was normal during monitoring. But it could be observed that pH was high during high temperatures. The experimental results were compared with laboratory results of samples taken on field and it had been observed that the designed hardware was giving accuracy in measured results with minor variations of ±3%.

Finally, we had tried to find relation between all these physiochemical parameters as provided in Figure 10. For the same, various samples (S1 through S7) had been taken after a periodical time interval of 24 hours from a local water body under normal environment conditions in month of March. It has been observed from the samples taken on field that water turbidity affects temperature of water body. This happens because suspended particles in a water body absorb sunlight and hence, increases temperature (Miskam et al., 2013).
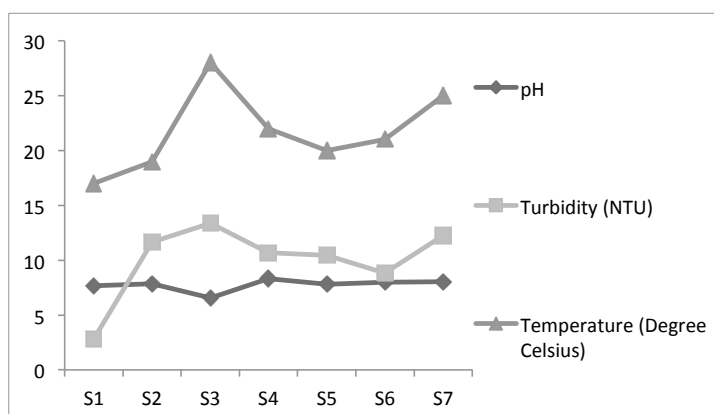
Figure 10. Representation of co-relation between WQM parameters

The results show that change in temperature causes change in ion concentration, ultimately shifting the pH value. As suggested by Le Chatelier's Principle, if a modification to reaction condition in dynamic equilibrium is performed, the equilibrium position moves to restrain the modification (Hall et al., 2007). Hence, if water temperature increases, the equilibrium will shift to decrease the water temperature. For this it will absorb the heat generated. It means the forward reaction will be adopted, and more concentration of hydroxide ions and hydrogen ions will be formed. Hence from the graph, we can see that the pH decreases with increases in temperature.

CONCLUSION

In this study, the pre-designed sensors are used to make a user-friendly hardware to monitor WQ parameters. This is an effective and innovative solution for water quality monitoring constantly. It can measure the various parameters of water quality using different sensors and communicates the results to the user automatically. In this paper, the focus is on the WQ parameters especially agriculture applications are, but the same can be applicable to other applications such as drinking water, water used for industries, domestic uses and all such applications where water quality is a concern. The system is simple and reduces human intervention. The future scope in this area is very vast and depending upon received data, actions for quality improvement can be taken automatically.

REFERENCES

Breijo, E. G., Sanchez, L. G., Civera, J. I., Ferrando, A. T., & Boluda, G. P. (2002). Thick-film multisensor for determining water quality parameters. In *IECON Proceedings (Industrial Electronics Conference)* (Vol. 4, pp. 2791–2796). IEEE. https://doi.org/10.1109/IECON.2002.1182837

Chou, J. C., Chen, C. C., Su, M. W., Chen, C. C., & Yang, S. Y. (2012). Cl - and H + Sensing devices for water quality monitoring system. In *2012 IEEE International Symposium on Circuits and Systems* (pp. 2043–2046). IEEE. https://doi.org/10.1109/ISCAS.2012.6271682

Cloete, N. A., Malekian, R., & Nair, L. (2016). Smart sensors for real-time water quality monitoring. *IEEE Access, 4*(9), 3975–3990. https://doi.org/10.1007/978-3-642-37006-9

CPCB. (2007). *Guidelines for water quality monitoring.* Central Pollution Control Board, Ministry of Environment and Forests, India.

Damian, C., Fosalau, C., Dias Pereira, J. M., Postolache, O., & Girao, P. S. (2012). Sensor network for water quality assesment. In *EPE 2012 - Proceedings of the 2012 International Conference and Exposition on Electrical and Power Engineering,* (Epe) (pp. 849–852). IEEE. https://doi.org/10.1109/ICEPE.2012.6463827

Das, B., & Jain, P. C. (2017, July). Real-time water quality monitoring system using Internet of Things. In *Computer, Communications and Electronics (Comptelix), 2017 International Conference on* (pp. 78-82). IEEE.. doi: 10.1109/COMPTELIX.2017.8003942

Deqing, M., Ying, Z., & Shangsong, C. (2012). Automatic measurement and reporting system of water quality based on GSM. In *Intelligent System Design and Engineering Application (ISDEA), 2012 Second International Conference on* (pp. 1007–1010). IEEE. https://doi.org/10.1109/ISdea.2012.595

Fengyun, M. (2010). Progress in water quality monitoring based on remote sensing and GIS. *Challenges in Environmental Science and Computer Engineering (CESCE), 2010 International Conference on, 2*(66), 209–212. https://doi.org/10.1109/CESCE.2010.246

Gorchev, H. G., & Ozolins, G. (2011). WHO guidelines for drinking-water quality. *WHO Chronicle, 38*(3), 104–108. https://doi.org/10.1016/S1462-0758(00)00006-6

Jiang, P., Xia, H., He, Z., & Wang, Z. (2009). Design of a water environment monitoring system based on wireless sensor networks. *Sensors, 9*(8), 6411–6434. https://doi.org/10.3390/s90806411

Lambrou, T. P., Anastasiou, C. C., Panayiotou, C. G., & Polycarpou, M. M. (2014). A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems. *IEEE Sensors Journal, 14*(8), 2765–2772. https://doi.org/10.1109/JSEN.2014.2316414

Liu, G., Chen, Y., Xu, Y., Wang, Y., Wang, D., Zhao, Y., … et al!!! (2011). Assessment of surface water quality of the Songhuajiang River basin, China. In *2011 International Conference on Remote Sensing, Environment and Transportation Engineering, RSETE* (pp. 5755–5758). IEEE. https://doi.org/10.1109/RSETE.2011.5965661

Meng, X. M. X., Zheng, W. Z. W., Chen, F. C. F., Shen, C. S. C., Sun, G. S. G., & Xing, Z. X. Z. (2010). Hand-held multi-parameter water quality recorder. In *World Automation Congress (WAC), 2010* (pp. 77-81). IEEE.

Miskam, M. A., Rahim, I. A., Sidek, O., Omar, M. Q., & Ishak, M. Z. (2013). Deployment of wireless water-quality monitoring system at titi serong paddy crop field, Malaysia. *Proceedings - 2013 IEEE 3rd International Conference on System Engineering and Technology, ICSET 2013* (pp. 57–60). IEEE. https://doi.org/10.1109/ICSEngT.2013.6650143

Nasirudin, M. A., Za'bah, U. N., & Sidek, O. (2011). Fresh water real-time monitoring system based on wireless sensor network and GSM. In *Open Systems (ICOS), 2011 IEEE Conference on* (pp. 354–357). IEEE. https://doi.org/10.1109/ICOS.2011.6079290

O'Flynn, B., Martinez-Catala, R., Harte, S., O'Mathuna, C., Cleary, J., Slater, C., … & Murphy, H. (2007). SmartCoast: A wireless sensor network for water quality monitoring. In *32nd IEEE Conference on Local Computer Networks* (pp. 815–816). IEEE. https://doi.org/10.1109/LCN.2007.34

Postolache, O., Girao, P., Pereira, M., & Ramos, H. (2002). An Internet and microcontroller-based remote operation multi-sensor system for water quality monitoring. In *Proceedings of IEEE Sensors* (pp. 1532–1536). IEEE. https://doi.org/10.1109/ICSENS.2002.1037350

Regan, F., Lawlor, A., Flynn, B. O., Torres, J., Martinez-Catala, R., O'Mathuna, C., & Wallace, J. (2009). A demonstration of wireless sensing for long term monitoring of water quality. In *2009 IEEE 34th Conference on Local Computer Networks* (pp. 819–825). IEEE. https://doi.org/10.1109/LCN.2009.5355047

Subsidiary, C., Manciu, A., Popa, M., & Mitrea, D. (2006). Parameters monitoring solutions for the quality control of water used in healthcare units. In *Automation, Quality and Testing, Robotics, 2006 IEEE International Conference on* (Vol. 2, pp. 457-462). IEEE.

Tie-zhu, Q., & Le, S. (2010). The design of multi-parameter online monitoring system of water quality based on GPRS. In *Multimedia Technology (ICMT), 2010 International Conference on* (pp. 1-3). IEEE

Vaddadi, S. K. (2012). Development of embedded wireless network and water quality measurement systems for aquaculture. In *Sixth International Conference on Sensing Technology (ICST) Development* (pp. 637–641). IEEE.

SCIENCE & TECHNOLOGY

PERTANIKA
JOURNALS

An Adaptive Mechanism to Optimise Routing Performance in Mobile Adhoc Networks

B. Nithya*, C. Mala and Abhishek Agrawal

Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli 620015, India

ABSTRACT

The primary goal of a MANET routing protocol is to provide a stable and efficient route to exchange messages between source and destination in a timely manner. This paper proposes an adaptive mechanism, namely Optimised Border node based Most Forward within Radius (OBMFR) mechanism to improve the method of choosing forwarding nodes towards the destination. The routing constraints such as region, mobility speed, residual energy and queue length are utilised by fuzzy logic controller to elect the potential forwarding node to improve the stability of routing paths. The performance metrics such as delay, routing overhead and energy consumption are significantly reduced as inferred from NS2 simulation.

Keywords: Fuzzy logic, MANET, NS2, overhead, routing protocol

INTRODUCTION

Mobile Adhoc NETwork (MANET) belongs to the family of wireless adhoc networks which finds application in a wide range of hostile environments, where human access is limited or unfeasible (Conti et al., 2014). A rising interest in MANETs research has

been observed mainly due to its issues in routing, broadcasting, QoS, connectivity, safety and security (Boukerche et al., 2011 and Swain et al., 2017). In the routing protocols design, the issues such as dynamic network topology, recurrent network partition, irregular demography and density are to be considered. Due to these issues, MANET routing protocols are incapable of adopting conventional routing techniques in adhoc-networks (Guptha et al., 2011).

Topology and Position based routing are the two broad categories of routing protocols in MANET (Husain et al., 2011). The topology-based approach utilises information about communication path to transmit data from source to destination. It is further divided into

three types: (i) Proactive (Table Driven) (ii) Reactive (On Demand) and (iii) Hybrid routing protocol. Proactive routing protocol makes routing decision with the help of routing tables which has network connectivity information in the form of tables. Since the routing table is frequently updated, it contains up-to-date connectivity information thus minimizing route discovery time. However, the overhead due to upgrading and maintenance of the routing tables is inevitable, especially in highly dynamic network (Taha et al., 2017).

To minimise such difficulties, the reactive routing protocols trigger the route discovery process when it is required, thereby minimising traffic on the network. Since the route discovery process needs to be started before data can be exchanged between source and destination pairs, it imposes a delay for the first packet to be transmitted. Moreover, if the network topology is frequently changed, then significant amount of routing control packets is to be exchanged (Hong et al., 2002).

Hybrid routing protocols have emerged to address the efficiency and scalability of routing protocols (Royer & Toh, 1999). Proactive routing component is used locally and reactive component is used globally to achieve benefits of both the protocols. Limitation of hybrid protocols is that they are not recommendable for highly mobile nodes, dynamically changing topology and large network size. Position-based routing protocols use periodic beacons to generate position information for routing. As routing is based on the position of the forwarding node and the destination node, unlike other routing protocols, it does not require establishment of the route or maintenance of the route (Mauve et al., 2001).

In this paper, an adaptive position-based routing is proposed to enhance selection of the best forwarding nodes. As multiple routing constraints such as region, mobility speed, residual energy and buffer size, are considered in this proposed work, a fuzzy logic controller is used to facilitate the selection process. The rest of the paper is structured as follows: Section 2 gives the overview of the related protocols found in the literature. In Section 3, the proposed methodology is elaborated by emphasising the next hop selection procedure. Section 4 analyses the performance of the proposed algorithm using NS2 simulation results. Finally, the paper is concluded in Section 5.

RELATED WORK

In this section, some of the existing position-based routing algorithms are discussed. Location Aided Routing (LAR) (De Rango et al., 1998), uses Global Positioning System (GPS) to gather local information to facilitate route discovery process. It uses partial flooding of control packets to enhance the route discovery phase. Three variants of LAR are proposed (Ko & Vaidya, 2000). The aim of the work is to enlarge the size of *request zone* when route discovery fails rather than using flooding. Modified LAR retains the advantages of original LAR with respect to delay and packet loss. In addition, it decreases control overhead.

Probabilistic rebroadcasting scheme in QoS-Aware node Selection Algorithm (QASA) (Mostafaa et al., 2014) utilises node density, distance and transmission range to elect the preferred next hop node. Among these factors, transmission range may be changed based on the application requirement. Similar to QASA, the factors, density and distance between

previous hop and current receiving node are used (Bae et al., 2013) as the fuzzy inputs to calculate rebroadcast degree.

A cross layer approach along with Position based forwarding technique is used (Patil et al., 2009) to improve the performance of Adhoc On-demand Distance Vector (AODV) Routing. Medium Access Control (MAC) layer estimates the signal strength of received packets. If the received packets have lesser signal strength than the threshold value, then MAC layer reports the network layer of the node. The network layer removes such nodes from the routing table. It excels than AODV in terms of latency, throughput and control overheads. The limitations include routing table overhead and the layer message overhead due to cross layer design.

Fuzzy logic approach is suggested (Babu et al., 2012) with fuzzy inputs, such as remaining energy, trust level and distance to determine the priority level. The node which has higher priority level is selected as forwarding node among many worthy nodes. From the simulation results, it has been shown that the lifetime of nodes and hence network lifetime is increased by electing appropriate forwarding node. Energy-Efficient Opportunistic Routing (EEOR) (Mao et al., 2011) selects a forwarder set. The nodes in forwarder set are prioritised using energy saving optimization technique. It mainly focuses on energy efficiency of individual node and did not consider the residual energy of other relaying nodes in the network. To overcome this problem, ENergy Saving via Opportunistic Routing (ENS-OR) algorithm is proposed (Luo et al., 2015). Using optimal energy strategy, the set of forwarding nodes are prioritised to protect the nodes with low residual energy. Based on the distance to the destination and residual energy, the optimal relay node is chosen among the potential forwarding nodes. From the simulation results, it has been shown that ENS-OR reduces energy consumption, thus, maximising network lifetime.

Position Based Multicast Routing Protocol for Ad-hoc Network Using Backpressure Restoration (PBMRP-BR) (Daniel et al., 2010) works for multimedia specific applications in Ad-hoc network. Here, routing is priority based in terms of route length, traffic and bandwidth, hence this information about each node is maintained for routing decisions. Among these parameters, route length has the highest priority whereas traffic load has the lowest priority. If there is more than one optimum path, then the tie is broken based on priority. Backpressure technique is used to deal with deteriorating QoS due to link failure. The PBMRP-BR protocol is better than DSR scheme as it provides higher reliability and lower overhead through Backpressure Restoration. It offers higher bandwidth utilisation and reduces congestion. The PBMRP is suitable for one-to many multicast scenarios.

Beacon-less Routing Algorithm for Vehicular Environments (BRAVE) is developed (Ruiz et al., 2010) based on beacon-less geographic routing. BRAVE executes hop-by-hop data forwarding after opportunistically selecting the next hop node using beaconless geographic routing. It makes use of a reactive scheme in which instead of performing periodic beacons, forwarding decisions are based on the position of the neighbors. Once the current node has forwarded the data packet, the next forwarding node is selected among the nodes which have received the packet correctly. The node which responds first is selected as the next hop node.

Reactive Virtual Cord Protocol (RVCP) (Awad et al., 2011) is a virtual location-based routing protocol that is reactive and uses Distributed Hash Table (DHT) like services for joining nodes and its selection as forwarding nodes. Forwarding packets are sent through a

route and flooding is not employed for route discovery. Adaptive methods are used to offer lower end-to-end delay and minimum power consumption. All the aforementioned protocols demand tedious time-consuming operations which in turn increases energy consumption thereby curtailing network lifetime. The necessary remedial measures are incorporated into the proposed mechanism. In the next subsection, the successors of the proposed mechanism are elaborated along their drawbacks.

PROPOSED OPTIMISED BORDER NODE BASED MOST FORWARD WITHIN RADIUS (OBMFR) MECHANISM

This section discusses about Most Forward within Radius (MFR) protocol, and Border-node based Most Forward within Radius (BMFR) and proposes an Optimised Border node based Most Forward within Radius (OBMFR) mechanism.

Most Forward within Radius (MFR)

Most Forward within Radius (MFR) (Omer et al., 2010), establishes a route in a network by calculating the distance of a neighbour node from the sender node. The next hop chosen for sending the packets further is the one whose progress on the straight line is maximum. Hence, the node chosen to forward the packet is near to the destination compared with the others. It helps in minimising the number of hops between sender and destination (Takagi & Kleinrock, 1984). In order to decide the next hop, unicast forwarding is used by making use of the position of the sender, its neighbour and the destination. Since the node closest to the destination is chosen as next hop, MFR reduces end-to-end delay. But the mobility of node may disturb the network connectivity.

Border-node based Most Forward within Radius (BMFR)

Border-node based Most Forward within Radius (BMFR) (Shringar Raw & Lobiyal, 2010) is an improvement over MFR. It selects the border node as the next hop to forward the packets to the destination. In this method, nodes are classified into three categories: border node, outer node and interior node. Nodes which are in the transmission range of sender are classified as interior nodes, those which are exactly at the maximum transmission range of the sender are classified as border nodes and the nodes lying outside the transmission range of the sender are outer nodes. Neglecting the interior nodes, BMFR selects the border node as the next hop by utilising the position of the nodes. BMFR minimises the number of hops to a maximum extent to reduce end to end delay as compared to MFR.

But, it fails to decide the next hop in the case of conflicting nodes. Moreover, it considered distance as the only parameter to consider the next hop. Before receiving data from the sender, the selected node may go out of the transmission range of the sender. Until the source detects this movement of the selected border node, it keeps sending the data. Hence, the packets are not delivered to the destination, thereby leading to underutilisation of the available resources.

Proposed OBMFR

To mitigate the above-mentioned drawbacks, the proposed OBMFR utilises four significant factors, Radius, Residual Energy, Mobility speed and Queue length, to facilitate next hop election. It yields better performance as shown in Section 4 in terms of throughput, delay, routing overhead, routing reconstruction and energy consumption.

Fuzzy logic approach is applied in control system to improve performance, especially when the input data is insufficient to form the crisp output (Wong & Wong, 2002). The block diagram of the proposed fuzzy logic controller in OBMFR is shown in Figure 1. Fuzzy inference engine applies fuzzy logic rules on the fuzzified variables to grade the node which is then used for forwarding data from source to the destination. The subsequent section discusses the function performed by various blocks in the fuzzy controller.
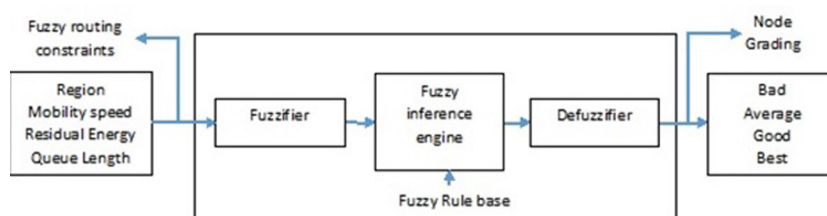
Figure 1. Fuzzy controller in OBMFR

Routing Constraints

The fuzzy system uses the following routing constraints to elect the best candidate node to be used in the routing process.

Region (R_i). The node's region helps to select the node from the appropriate region, which is neither too far nor too close from the sender node. The proposed OBMFR divides the sensing range of a node into four regions, namely R1, R2, R3 and R4 as shown in Figure 2. The snapshot in Figure 3 shows one of the scenarios in the simulation to grade node 1's neighbors. Node 1 is assumed as a source node where the position of all the nodes is generated randomly. All the nodes which are in the sensing range of node 1 are connected via a line with node 1 at the centre. Nodes which are not connected are outside the sensing range of node 1. On applying the fuzzy rules based on the four input parameters, neighbour nodes of node 1 are classified into various grading, i.e., bad, average, good and best. Nodes connected with red, cyan, blue and green colours are bad, average, good and best candidates respectively. The node connected with green is considered as the best candidate and is considered as the next hop for routing and the process continues till the destination is reached.
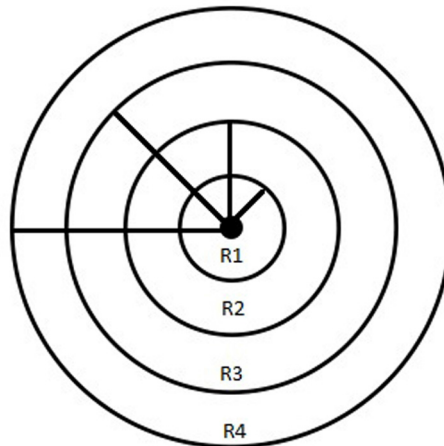
Figure 2. Node with four regions
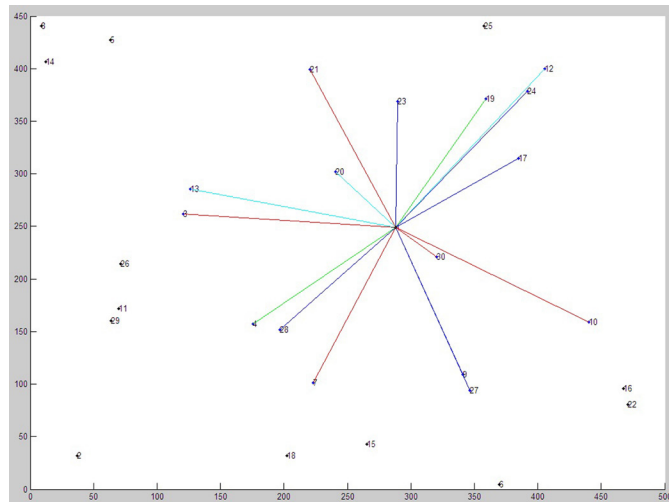
Figure 3. Grading of node 1's neighbours

Mobility Speed (S_i). Mobility speed is an important factor as it indirectly affects connectivity of nodes in the network. Since, the nodes with high speed motion leads to frequent link breakages, it is not considered for further routing process. Nodes with medium and slow speed can lead to better lifetime of a link.

Residual Energy (E_i). As every node is battery powered in a wireless network, residual energy must be used in next hop selection. Only the nodes with adequate energy are selected to avoid node failure resulting from exhausted energy.

Queue Length (Q_i)

The node's queue length can select the optimum node so as to avoid the problem of buffer overflow. Thus, the overloaded node will not be further considered by repeatedly selecting an intermediate node. Instead, alternate nodes with sufficient queue length are elected by the proposed OBMFR.

The membership functions of these fuzzy inputs are shown in Figure 4 (a)-(d) and Figure 5 shows the membership function of output. The minimum and maximum values of inputs and outputs for Fuzzy Logic Controller are shown in Table 1. The input and output domain of fuzzy variables are given in Table 2. Once the membership function is defined, the crisp inputs (R_i, S_i, E_i and Q_i) are fuzzified to map the inputs to the range [0,1]. Based on the fuzzy inputs, fuzzy inference engine determines the grading of node by evaluating the fuzzy rules in the fuzzy rule base. In fuzzy rule base, the set of IF THEN rules are constructed on the basis of human expert's knowledge. Using the fuzzy variable of four parameters, 108 rules can be formed. These rules will then help in deciding the category of the node. A snapshot of fuzzy rules is shown in Figure 6. For example:

Rule 1. From Table 2, if a node lies in the far region with high speed, high energy, and fair queue length, then the node will be considered as a Best candidate.

(a) Region

(b) Speed

(c) Energy

(d) Queue Length
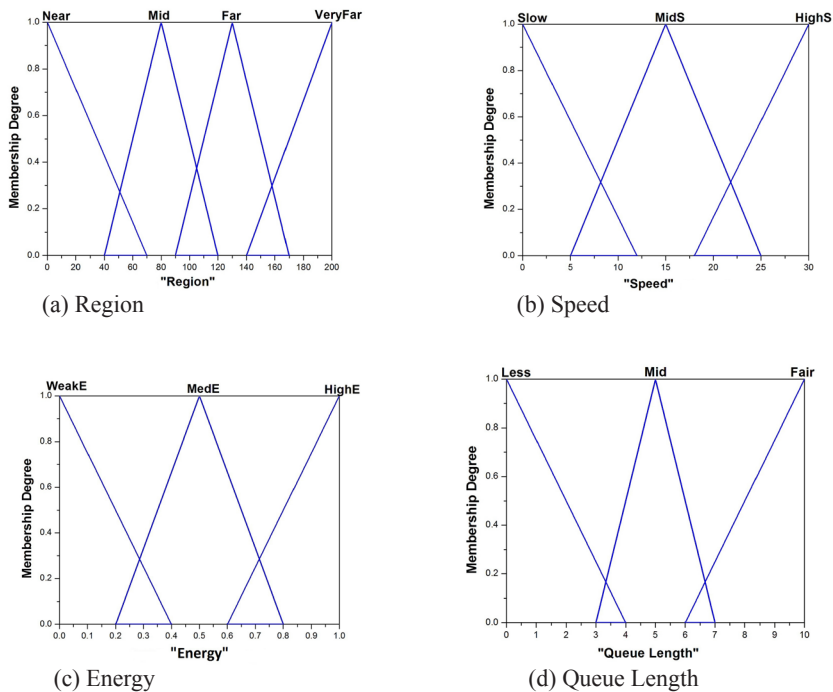
Figure 4(a)-(d). Fuzzy membership function for different Input Parameters
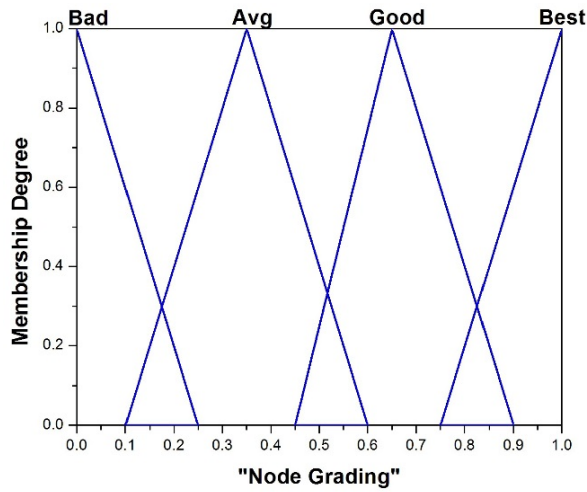
Figure 5. Fuzzy membership function for Output Parameters

Table 1
Universe of Discourse for different input and output parameters

Name	Input /Output	Min. value	Max. Value
Node's Region (Ri)	I	0	200
Node's Energy (E_i)	I	0	1
Node's Speed (S_i)	I	0	30
Node's Queue Len (Q_i)	I	0	10
Node's Grading (G_i)	O	0	1

Table 2
Fuzzy variable for different input parameters with their domain

I. Node's Region		IV. Queue Length	
Input Domain	Fuzzy Variable	Input Domain	Fuzzy Variable
0-80	Near / R1	0-4	Less
40-120	Mid / R2	3-7	Med
90-170	Far / R3	6-10	Fair
140-200	Very Far / R4		
II. Energy Level		Node Grading:	
Input Domain	Fuzzy Variable	Input Domain	Fuzzy Variable
0-.4	Weak	0-.25	Bad
.2-.8	MedE	-1-.6	Avg
.6-1	HighE	.45-.9	Goog
III.Speed		.75-1	Best
Input Domain	Fuzzy Variable		
0-12	Slow		
5-25	MedS		
18-30	HighS		
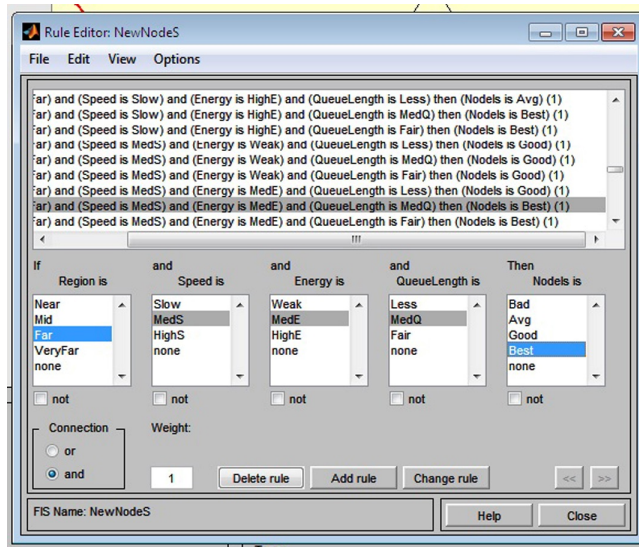
Figure 6. Proposed fuzzy rule base

Rule 2. Similarly, from Table 2, if a node lies in the very far region with high speed, low energy and less queue length, then the node will be considered as bad node, not suited for the routing data to destination. Due to high speed and distance, the node may go out of range before a sender node can send the packet to the underlying node. The surface graphs shown in Figure 7-9 depict the significant impact of fuzzy inputs on node grading. The output of fuzzy inference engine is then aggregated. These aggregated outputs fuzzy set is defuzzified to get a single output that gives the grade level of nodes. With the help of this grading, only the potential nodes are used in the routing to exchange data between source and destination. The efficiency of the proposed algorithm is tested using NS2 simulation and results are discussed in Section 4. Also, its time complexity is analysed and compared with MFR and BMFR in the next subsection.
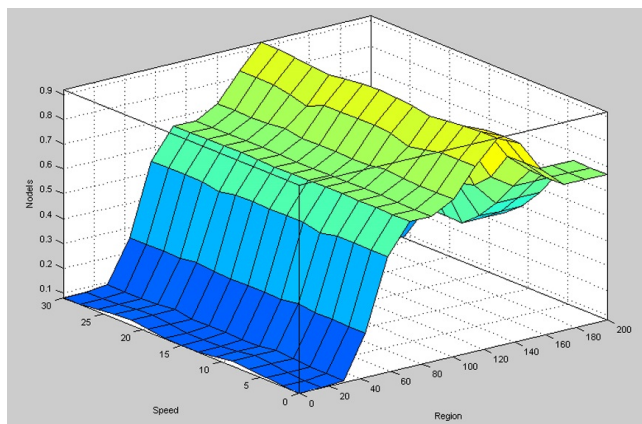
Figure 7. Region and speed vs grading
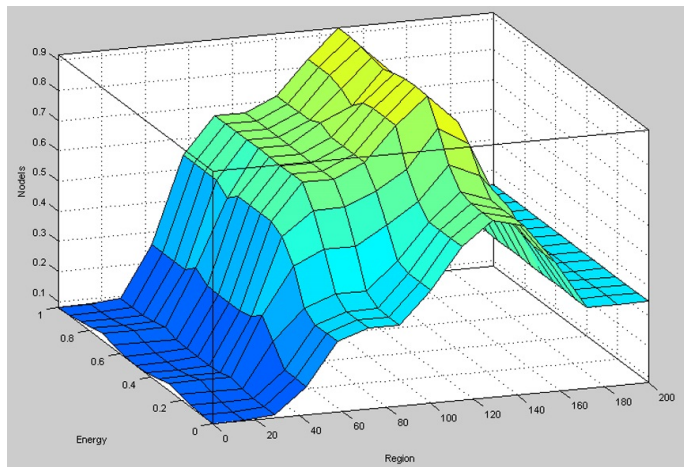
Figure 8. Region and energy vs node grading
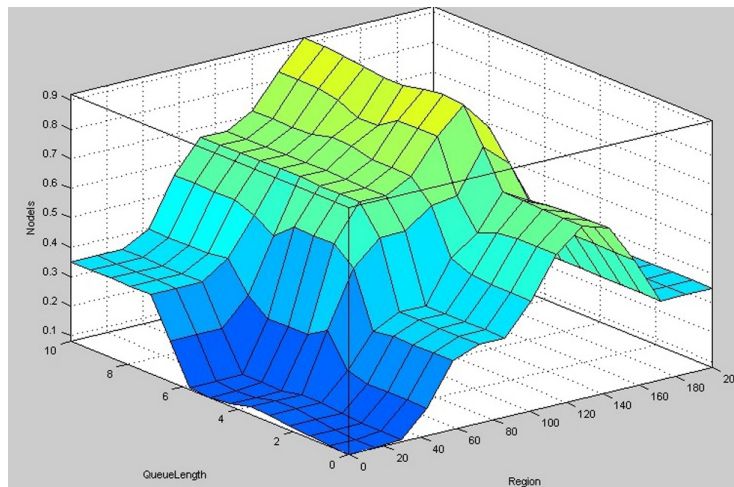
Figure 9. Region and queue length vs node grading

Time Complexity

In this section, the time complexity of MFR, BMFR and the proposed OBMFR are analysed. In MFR, the node which has the greatest projection on the source-destination line is selected as next hop from the source node and this is repeated until the destination node is reached. With n nodes and for one source destination pair, this method takes a worst-case time complexity of $O(n^3)$. When all n nodes become part of the communication, the worst-case time complexity is $O(n^4)$. In BMFR, the Euclidian distance from the source to all of its neighbour nodes are computed and for one node, this takes a worst-case time of $O(n)$. So, for n nodes, it is $O(n^2)$. From this, after examining the neighbouring nodes, the border nodes are selected as candidate nodes and this process takes time of $O(n)$. Hence, the total time taken by BMFR is $O(n^3)$.

The proposed OBMFR algorithm consists of two phases, namely Fuzzy logic-based node selection and data forwarding. In the first phase, the potential next hop neighbours are selected using fuzzy logic. Fuzzy logic computations do not necessarily increase the time complexity in asymptotic sense (Khan et al., 2004), it is calculated for the comparison purpose. As stated by Balázs et al., 2008, the time complexity of Mamdani base fuzzy inference technique is $O(r^*m)$, where r and m denote number of rules and number of input dimension respectively. In the proposed OBMFR, r is 108 and m is 04. In the second phase, forwarding of data from source to destination is done with the aid of selected nodes and worst-case time complexity of this phase is $O(n)$. Hence, the overall time complexity of the proposed OBMFR is $O(r^*m) + O(n)$. Table 3 shows the comparison of time complexity of MFR, BMFR and the proposed OBMFR.

Table 3
Complexity analysis

Algorithm	Time complexity
MFR	$O(n^4)$
BMFR	$O(n^3)$
OBMFR	$O(r^*m)+O(n)$

SIMULATION AND PERFORMANCE ANALYSIS

Simulation of the proposed Optimised Border node based Most Forward within Radius (OBMFR) is performed using NS2. The simulation results obtained from the simulation are compared with BMFR to emphasise the enhancements of OBMFR in terms of throughput, loss, routing overhead and energy consumption. To investigate the scalability of the proposed OBMFR, the number of nodes varies from 10 to 70. The radio propagation range of 250 meters with 2Mbps link capacity is assumed. A free space propagation model with Direct Sequence Spread Spectrum (DSSS) is considered for the simulation. A random way point mobility model is utilised to incorporate node mobility into random topology. The traffic generator that generates 512 bytes TCP data is used. IEEE 802.11 MAC protocol with RTS/CTS access method is used to control channel access. The following subsections analyse the performance of OBMFR, BMFR and MFR under throughput, delay, packet loss, route reconstruction and energy consumption.

Throughput

Figure 10 shows the throughput performance of the proposed OBMFR, BMFR and MFR for varying number of nodes. The proposed OBMFR minimises the frequent network partition thereby improving the connectivity of the network. This enhanced performance is obvious due to the role of the decision parameters in the next hop selection. The low energy node and overloaded node are restricted from the route construction process. OBMFR selects only quality nodes based on the mobility speed, residual energy and location with respect to source node. It leads to greater number of packets to be transmitted through the routing path which is more stable than in BMFR. Thus, the proposed algorithm enhances the success rate of transmission.
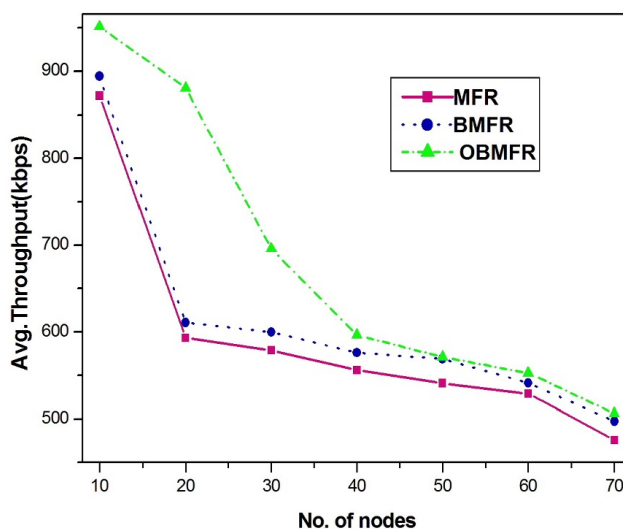
Figure 10. Throughput performance

End-to-End Delay

Figure 11 shows the end-to-end delay performance of proposed OBMFR, BMFR and MFR. Since the probability of selecting potential next hop towards the destination is improved by the proposed OBMFR, the lifetime of the established routing path is increased. The number of disconnected nodes due to the failure (as a result of exhausted energy and buffer overflow) and failures of link (due to mobility) are significantly reduced. Hence the tedious process of route discovery is minimised. By enhancing the connectivity between the nodes in the network, data is exchanged successfully with minimum end to end delay.
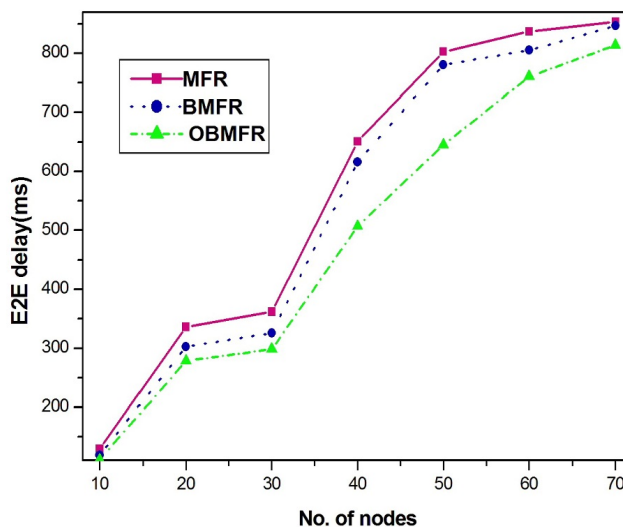
Figure 11. End-to-End delay performance

Packet Loss

Selection of an overloaded node as one of the nodes participating in routing may result in packet loss by frequently dropping the packets. To overcome this problem, queue length is considered as one of the parameters in deciding the next hop. The node with average or fair queue length is preferred to minimise the packet loss as caused by buffer overflow. Figure 12 shows the packet loss incurred by OBMFR, BMFR and MFR. Whereas in BMFR, distance is the only parameter to elect the border node that is closest to the destination. Even though the number of hops between the source and destination is reduced, the border node which lies exactly at the maximum transmission range may move out of the transmission range before receiving the packet from the source. This situation occurs frequently in highly dynamic network, thereby leading to more collision rate as depicted in Figure 13. As a result, more packet losses are inevitable in BMFR and MFR as compared to the proposed OBMFR.
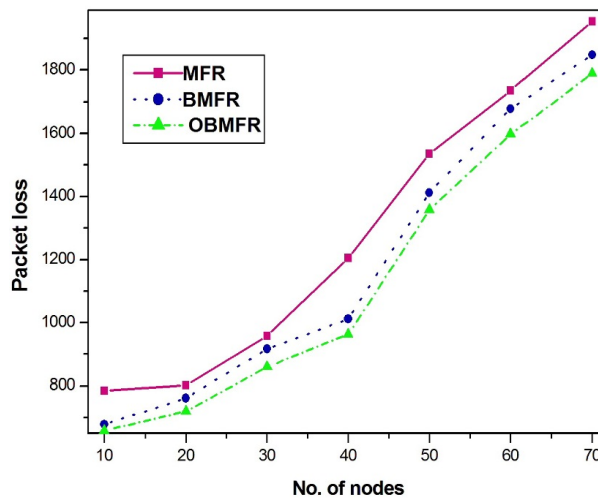
Figure 12. Packet loss performance
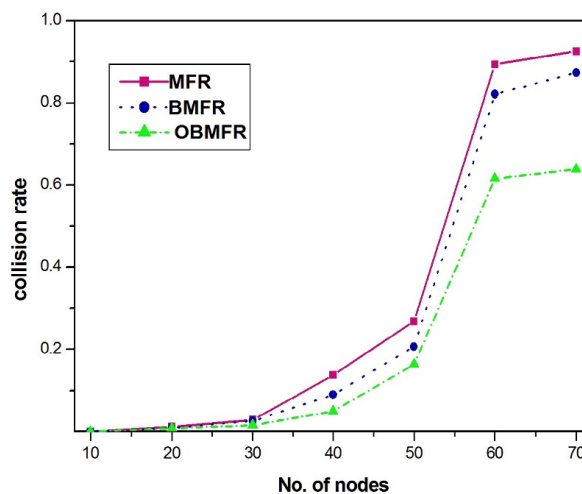
Figure 13. Collision rate performance

Route Reconstruction

Link disruption reinitiates the route discovery process causing too many routing control packets on the fly from source to the destination. In BMFR, communication between the source and the destination is facilitated by the border nodes. Based on the number of border nodes and their mobility, the availability and life time of routing paths may vary affecting the overall network performance. From Figure 14, it is inferred that the proposed OBMFR minimises the occurrences of route reconstruction process by effectively maintaining the connectivity among the nodes. With the aid of significant factors, namely region, residual energy, mobility speed and queue length in the next hop selection, the proposed algorithm improves the stability of routing path. Hence the frequency of route reconstruction is minimised.
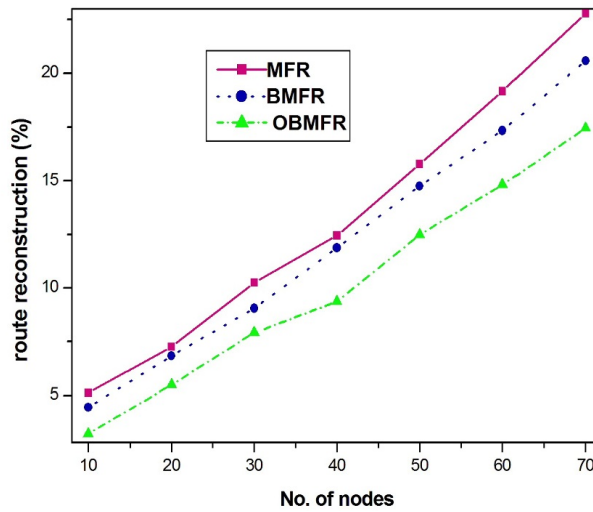
Figure 14. Percentage of route reconstruction performance

Energy Consumption

The lesser the node participates in communication, the lower will be its energy consumption resulting in increased network lifetime. Energy is another driving factor in wireless scenario. The residual energy is considered as one of the parameters in deciding the next hop in the proposed OBMFR. For this reason, a node with moderate or high energy is chosen as next hop using fuzzy rules. The fact is taken into consideration to avoid reselection of a particular node again and again, failing to do so will sooner make the victim node to be dead soon. The proposed OBMFR shows significant decrease in energy consumption as compared to the traditional BFMR as shown in Figure 15. As mentioned in the previous section, the proposed OBMFR significantly reduces the packet loss (thus, minimising the number of retransmission) frequency of route construction process (which curtails the amount of routing control packets). Due to this, the consumed energy is minimised which leads to prolonged network lifetime.
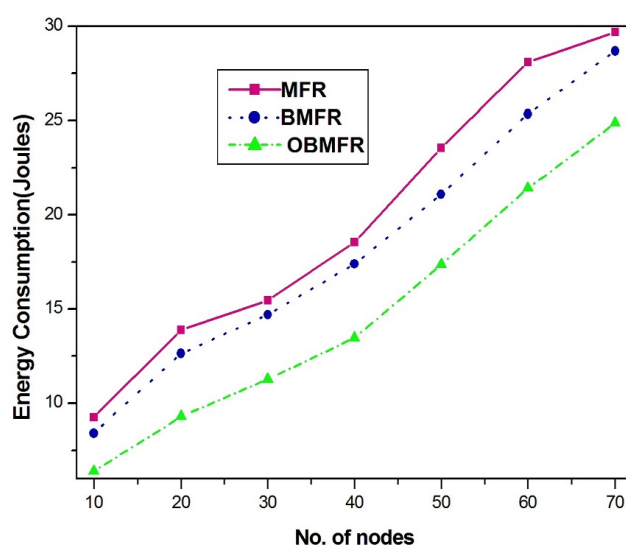
Figure 15. Energy consumption

CONCLUSION

To optimise the routing performance in MANET, an adaptive Optimised Border-node based Most Forward within Radius (OBMFR) mechanism is proposed. Multiple constraints, such as region, residual energy, mobility speed and queue length are fed into fuzzy controller along with the proposed fuzzy rules to select the best candidate node to forward data from source to destination. From the simulation results, it is inferred that the proposed OBMFR provides more stability and reliable routing path by minimising the number of route reconstruction. It further leads to less energy consumption by reducing routing control overheads.

REFERENCES

Awad, A., Thiel, A. M., & Dressler, F. (2011). Reactive virtual position-based routing in wireless sensor networks. In *International Conference on Computer Communications and Networks* (pp. 1-6). IEEE.

Babu, S. S., Raha, S., Naskar, M, Alfandi, K.O., & Hogrefe, D. (2012). Fuzzy logic election of node for routing in WSNs. In *IEEE International Conference on Trust, Security and Privacy in Computing and Communication*s (pp. 1279-1284). IEEE.

Bae I. H., & Olariu, S. (2013). Design and evaluation of Hi-CAST and its variants for safety message dissemination in VANET. In *International Conference on Control, Automation and Systems* (pp. 800-806). IEEE.

Balázs L. K., Kóczy J. T., & Botzheim, J. (2008). Comparison of fuzzy rule-based learning and inference systems. In *International Symposium of Hungarian Researchers on Computational Intelligence and Informatics* (pp. 61-75). CINTI.

Boukerche, A., Turgut, B., Aydin, N., Ahmad, M. Z, Bölöni, L., & Turgut, D. (2011). Routing protocols in ad hoc networks: A survey. *Computer Networks, 55*(13), 3032–3080.

Conti, M., & Giordano, S. (2014). Mobile ad hoc networking: milestones, challenges, and new research directions. *Communications Magazine, 52*(1), 85-96.

Daniel, A. K., Singh, R., & Khan, Z. (2010). Position based multicast routing protocol for AD-hoc wireless network using backpressure restoration. In *International Conference on Computer Engineering and Technology* (pp. 458-462). IEEE.

De Rango, F., Iera, A, Molinaro, A., & Marano, S. (1998). A modified location-aided routing protocol for the reduction of control overhead in Ad-hoc wireless networks. In *International Conference on Telecommunications* (pp. 1033-1037). IEEE.

Gupta A. K., Sadawarti, H., & Verma, A. K. (2011). Review of various routing protocols for MANET. *International Journal of Information and Electronics Engineering, 1*(3), 251-259.

Hong X., Xu, K., & Gerla, M. (2002). Scalable routing protocols for mobile Ad Hoc networks. *IEEE Network Magazine, 16*(4), 11-21.

Husain A., Raw R. S., Kumar, B., & Doegar, A. (2011). Performance comparison of topology and position based routing protocols in vehicular network environments. *International Journal of Wireless and Mobile Networks, 3*, 289-303.

Khan, J. A., & Alnuweiri, H. M. (2004). A fuzzy constraint-based routing algorithm for traffic engineering. In *IEEE Global Telecommunications Conference* (pp. 1366-1372). IEEE.

Ko, Y. B., & Vaidya. N. H. (2000). Location Aided Routing (LAR) in mobile ad hoc networks. *Wireless Networks, 6*(4), 307-321.

Luo, J., Hu, J., Wu, D., & Li, R. (2015). Opportunistic routing algorithm for relay node selection in wireless sensor networks. *IEEE Transactions on Industrial Informatics, 11*(1), 112-121.

Mao, X., Tang, S., Xu., X, Li, X., & Ma, H. (2011). Energy efficient opportunistic routing in wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems, 22*, 1934–1942.

Mauve, M., Widmer, J., & Hartenstein, H. (2001). A survey on position-based routing in mobile ad hoc networks. *IEEE Network Magazine, 15*(6), 30-39.

Mostafaa, A., Vegnib, A. M., & Bandaranayakea, A.D. P. (2014). QoS-aware node selection algorithm for routing protocols in VANETs. *Procedia Computer Science, 40*, 66-73.

Omer, K. A. A. (2010). Analytical study of MFR routing algorithm for mobile Ad hoc networks. *Journal of King Saud University - Computer and Information Sciences, 22*, 29-35.

Patil, R., & Damodaram, A. D. R. (2009). Cross layer AODV with position based forwarding arouting for mobile adhoc network. In I*EEE Conference on Wireless Communication and Sensor Networks* (pp. 1-6). IEEE.

Raw, S., & Lobiyal, D. K. (2010). B-FMR routing protocol for vehicular Adhoc networks. In *International conference on Networking and Information Technology* (pp. 420-423).

Royer, E. M., & Toh, C. K. (1999). A review of current routing protocols for ad hoc mobile wireless networks. *Personal Communications, 6*(2), 46-55.

Ruiz, P. M., Cabrera, V., Martinez, J. A., & Ros, F.J. (2010). BRAVE: Beacon-less routing algorithm for vehicular environments. In *IEEE International Conference on Mobile Adhoc and Sensor Systems* (pp. 709-714). IEEE.

Swain, J., Pattanayak, B. K., & Pati, B. (2017). Study and analysis of routing issues in MANET. In *International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 506-509). IEEE.

Taha, A., Alsaqour, R., Uddin, M., Abdelhaq, M., & Saba, T. (2017). Energy efficient multipath routing protocol for mobile Ad-Hoc network using the fitness function. *IEEE Access, 5*, 10369-10381.

Takagi, H., & Kleinrock, L. (1984). Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transaction on Communication, 32*(3), 246-257.

Wong, Y., & Wong, W. (2002). A fuzzy decision based routing protocol for MANET. In I*EEE International Conference on Networks* (pp. 317-322). IEEE.

SCIENCE & TECHNOLOGY

Design of the Side Sensitive Group Runs Chart with Estimated Parameters Based on Expected Average Run Length

You Huay Woon

Pusat PERMATApintar Negara,Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ABSTRACT

The assumption when constructing a control chart is that the process parameters, i.e. mean and standard deviation, are known. Nevertheless, this assumption is not realistic in practical situations. In most of the application of a control chart, the mean and standard deviation are unknown and are estimated from an in-control Phase-I samples. When the process parameters are estimated, the control chart performs differently compared with the corresponding chart with known process parameters because of the variability of estimators. The usual practice to evaluate the performance of a control chart is to use the average run length (ARL). The ARL is the average number of samples plotted on a control chart before an out-of-control signal is detected. In addition, the expected ARL (EARL) is used as a performance measure for the random process mean shift. In this article, the performance of the side sensitive group runs (SSGR) chart with known and estimated process parameters are studied and examined in terms of ARL and EARL.

Keywords: Estimated parameters, expected average run length, side sensitive group runs

INTRODUCTION

The control chart was introduced by Walter A. Shewhart in 1924 to monitor and determine whether a process is in statistical control (Montgomery, 2012) and if it is, the process is seen as conforming. This indirectly ensures the quality of output. Control charts are widely used in various fields, for example, manufacturing and the service industry. Therefore, the control chart has been recognised as one of the seven magnificent tools in Statistical Process Control (SPC).

The Shewhart chart has been widely used to monitor the process. However, the major shortcoming of the Shewhart chart is lack of sensitivity towards small and moderate process mean shifts (Sanusi, Abujiya, Riaz, & Abbas, 2017). In view of this, the Shewhart chart has been studied extensively to enhance

the sensitivity of the control chart towards small and moderate process mean shift. For example, synthetic chart (Wu & Spedding, 2000a), group run (GR) chart (Gadre & Rattihalli, 2004) and side sensitive group run (SSGR) chart (Gadre & Rattihalli, 2007).

An indispensable assumption when designing a control chart is that process parameters, such as the mean and the standard deviation, are assumed known (Chen, Birch, & Woodall, 2016). However, in many applications of the control chart, the process parameters are unknown. Thus, the process parameters are usually estimated from an in-control Phase-I samples. Woodall and Montgomery (1999, 2014) pointed out that when the process parameters are estimated, the performance of the control chart will differ from the known process parameters case due to estimation error. Therefore, Psarakis, Vyniou and Castagliola (2014) emphasised that it is essential to study the performance of a control chart when process parameters are unknown. Jensen, Jones-Farmer, Champ and Woodall (2006), Saleh, Mahmoud, Jones-Farmer, Zwetsloot & Woodall (2015), You, Khoo, Castagliola and Ou, (2015) and Shepherd, Champ and Rigdon (2016), to name a few, examined the performance of control charts with unknown process parameters.

The performance of a control chart is an important characteristic to consider because it influences the decision on the use of the control chart. A well-known and common performance measure is average run length (ARL) (Chakraborti, 2007). The ARL indicates, on the average, how many samples need to be plotted before an out-of-control signal is detected. The computation of the ARL is based on the particular mean shift provided.

Nevertheless, there are situations when the practitioner is unable to identify the shift of a process (Teoh, Chong, Khoo, Castagliola, & Yeong, 2017). Moreover, the practitioners may not have historical knowledge or experience on the process to determine the shift of a process (Castagliola, Celano, & Psarakis, 2011). Determining the particular mean shift will result in decision inaccuracy of the process if a different mean shift occurs in the process. In view of this, expected average run length (EARL) is introduced in this paper to evaluate and design the SSGR chart. The EARL is obtained by integrating over the density function of the shift size.

In this paper, the SSGR chart with estimated process parameters will be investigated when the process mean shift is random. Moreover, the performance of the control chart will be compared with the particular process mean shift for the known and estimated process parameters case. This approach is to convince the practitioners to implement EARL when the shift size may not be known in advance.

METHODS

The SSGR chart developed by Gadre and Rattihalli (2007) comprises a Shewhart sub-chart and an extended version of the conforming run length (CRL) sub-chart. The Shewhart sub-chart is designed with two control limits, the *LCL* and *UCL*. Meanwhile, the CRL sub-chart has a single control limit, *L*, which is a specified positive integer.

The operation of the SSGR chart is explained as follows:

(1) When a sample falls below the *LCL* or above the *UCL*, the SSGR chart indicates this is a nonconforming sample.

(2) Further investigation using the CRL sub-chart is necessary to determine the state of the process.

(3) The rth CRL value will be computed, i.e. CRL_r, for $r = 1, 2, \ldots$ is denoted as the number of conforming samples plotted on the Shewhart sub-chart between the rth and $(r-1)^{th}$ nonconforming samples.

(4) The SSGR chart declares an out-of-control if

   (i) $CRL_1 \leq L$ or

   (ii) $CRL_r \leq L$ and $CRL_{r+1} \leq L$, for $r = 2, 3, \ldots$ and that both CRL_r and CRL_{r+1} fall on the same side of the Shewhart sub-chart.

It should be emphasised that when a sample falls outside the control limits of the Shewhart sub-chart, the SSGR chart does not signal an out-of-control status immediately. Instead, it just indicates a nonconforming sample. Further investigation using the CRL sub-chart is required before an out-of-control status is signalled.

When the process parameters are assumed known, the *LCL* and *UCL* are

$$LCL = \mu_0 - \frac{H}{\sqrt{n}} \sigma_0 \qquad [1]$$

and

$$UCL = \mu_0 + \frac{H}{\sqrt{n}} \sigma_0, \qquad [2]$$

where H is the design constant. The probability of a conforming sample on the Shewhart sub-chart is $A = \Pr\left(\bar{X} \in [LCL, UCL]\right)$, i.e.

$$A = \Pr\left(\mu_0 - \frac{H}{\sqrt{n}} \sigma_0 \leq \bar{X} \leq \mu_0 + \frac{H}{\sqrt{n}} \sigma_0 \right). \qquad [3]$$

Following simplification, A reduces to

$$A = \Phi\left(H - \delta\sqrt{n}\right) - \Phi\left(-H - \delta\sqrt{n}\right). \qquad [4]$$

Here, $\Phi(.)$ is the standard normal cumulative distribution function (cdf).

Let B denote the probability that a sample on the Shewhart sub-chart is nonconforming, i.e.

$$B = \Pr\left(\bar{X} \notin [LCL, UCL]\right) = 1 - A. \qquad [5]$$

In addition, the probability of an event $CRL_r \leq L$ is

$$C = 1 - (1 - B)^L. \qquad [6]$$

Also, the conditional probability $k = k = \Pr\left(\bar{X} > UCL \,\middle|\, \bar{X} \notin [LCL, UCL]\right)$ for taking into account the side sensitivity aspect is

$$k = \frac{\Pr\left(\bar{X} > UCL\right)}{B} = \frac{1 - \Phi\left(H - \delta\sqrt{n}\right)}{B}.$$ [7]

Finally, the ARL formula for the SSGR chart is (Gadre & Rattihalli, 2007)

$$\mathrm{ARL} = \frac{1 - k(1-k)C^2}{BC^2[1 + k(1-k)(C-2)]}.$$ [8]

The evaluation using ARL requires the shift size to be determined in advance. Unfortunately, this is not a practical situation. Therefore, the EARL can be employed in place of the ARL. The EARL is computed as

$$\mathrm{EARL} = \int_{\delta_{\min}}^{\delta_{\max}} f_\delta\left(\delta\right) \mathrm{ARL}\, d\delta.$$ [9]

It is the expected value of the ARL integrated over the density function $f_\delta(\delta)$ of the mean shift size in the process, i.e. δ. In this paper, two interval combinations of random shift sizes are set: $(\delta_{\min}, \delta_{\max}) = (0.1, 1.0)$ and $(\delta_{\min}, \delta_{\max}) = (1.0, 2.0)$, to evaluate the performance of the SSGR chart.

In a real situation, the process parameters are rarely known. Hence, the process parameters are usually estimated from m in-control Phase-I samples with each of size n. An estimator of μ_0 is

$$\hat{\mu}_0 = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} Y_{i,j}$$ [10]

and an estimator of σ_0 is

$$\hat{\sigma}_0 = \sqrt{\frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(Y_{i,j} - \bar{Y}_i\right)^2},$$ [11]

where the sample mean is $\bar{Y}_i = \dfrac{1}{n} \sum_{j=1}^{n} Y_{i,j}$.

Based on the estimated process parameters, i.e. $\hat{\mu}_0$ and $\hat{\sigma}_0$ from the Phase-I parameter estimation, the LCL and UCL are

$$LCL = \hat{\mu}_0 - \frac{H'}{\sqrt{n}} \hat{\sigma}_0$$ [12]

and

$$UCL = \hat{\mu}_0 + \frac{H'}{\sqrt{n}} \hat{\sigma}_0,$$ [13]

respectively, with H' being the design constant for the Shewhart sub-chart with estimated process parameters. Let $\hat{A}$ represent the probability that a sample is conforming on the Shewhart sub-chart and is computed as

$$\hat{A} = \Pr\left(\hat{\mu}_0 - \frac{H'}{\sqrt{n}} \hat{\sigma}_0 \leq \bar{X} \leq \hat{\mu}_0 + \frac{H'}{\sqrt{n}} \hat{\sigma}_0 \right).$$ [14]

By defining $V = \left(\hat{\mu}_0 - \mu_0 \right) \dfrac{\sqrt{n}}{\sigma_0}$ and $W = \dfrac{\hat{\sigma}_0 \sqrt{n}}{\sigma_0}$, the Equation [14] simplifies to (You, Khoo, Castagliola, & Ou, 2015)

$$\hat{A} = \Phi\left(V + \frac{H'}{\sqrt{n}} W - \delta\sqrt{n} \right) - \Phi\left(V - \frac{H'}{\sqrt{n}} W - \delta\sqrt{n} \right).$$ [15]

Because $\hat{\mu}_0 \sim N\left(\mu_0, \dfrac{\sigma_0^2}{mn} \right)$, it can be deduced that $V \sim N\left(0, \dfrac{1}{m} \right)$. Then, the probability density function (pdf) of V is

$$f_V\left(v \middle| m \right) = f_N\left(v \middle| 0, \frac{1}{m} \right),$$ [16]

where f_N is the pdf of the normal distribution with mean 0 and variance $\dfrac{1}{m}$. According to Zhang, Castagliola, Wu and Khoo (2011), it can be shown that $W^2 = \dfrac{\hat{\sigma}_0^2 n}{\sigma_0^2} \sim \gamma\left(\dfrac{m(n-1)}{2}, \dfrac{2n}{m(n-1)} \right)$. Hence, the pdf of W is

$$f_W\left(w \middle| m, n \right) = 2w f_\gamma\left(w^2 \middle| \frac{m(n-1)}{2}, \frac{2n}{m(n-1)} \right),$$ [17]

where f_γ is the pdf of the gamma distribution with parameters $\dfrac{m(n-1)}{2}$ and $\dfrac{2n}{m(n-1)}$.

The probability that a sample is nonconforming on the Shewhart sub-chart is

$$\hat{B} = 1 - \hat{A}.$$ [18]

In a similar manner, the probability of an event $\mathrm{CRL}_r \leq L'$ and the conditional probability, $\Pr\left(\bar{X} > UCL \middle| \bar{X} \notin \left[LCL, UCL \right] \right)$ are

$$\hat{C} = \Pr\left(\mathrm{CRL}_r \leq L' \right) = 1 - \left(1 - \hat{B} \right)^{L'}$$ [19]

and

$$\hat{k} = \frac{1 - \Phi\left(V + \frac{H'}{\sqrt{n}}W - \delta\sqrt{n}\right)}{\hat{B}}, \tag{20}$$

respectively, with L' being the control limit of the CRL sub-chart with estimated process parameters.

Finally, the ARL of the SSGR chart with estimated process parameters is

$$\text{ARL} = \int_{-\infty}^{+\infty}\int_{0}^{+\infty} \frac{1 - \hat{k}\left(1 - \hat{k}\right)\hat{C}^2}{\hat{B}\hat{C}^2\left(1 + \hat{k}\left(1 - \hat{k}\right)\left(\hat{C} - 2\right)\right)} f_V\left(v|m\right) f_w\left(w|m, n\right) dwdv. \tag{21}$$

Furthermore, the EARL of the SSGR chart with estimated process parameters is

$$\text{EARL} = \int_{\delta_{min}}^{\delta_{max}}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} f_\delta\left(\delta\right) \text{ARL} f_V\left(v|m\right) f_W\left(w|m, n\right) dwdvd\delta. \tag{22}$$

where ARL can be obtained from Equation (8) by replacing B, C and k with $\hat{B}$, $\hat{C}$ and $\hat{k}$, respectively.

RESULTS AND DISCUSSION

Table 1 presents the ARL performance for the SSGR chart. Let ARL_0 and ARL_1 represent the in-control ARL and out-of-control ARL respectively. For comparison purposes, we include computations of ARL_1s for the known process parameters chart and denote it as $m = +\infty$. Meanwhile, $m = \{30, 50, 80, 200, 500\}$ represents the cases when process parameters are estimated.

Different combinations of number of in-control Phase-I samples, $m \in \{30, 50, 80, 200, 500, +\infty\}$, sample sizes, $n \in \{3, 5, 7, 9\}$ and process mean shift sizes, $\delta = \{0.1, 0.3, 0.7, 1.1, 1.5, 2.0\}$ are considered in Table 1. The results of in columns 4 – 9 are evaluated using the optimal pair (H, L) of the chart with known process parameters, i.e. the optimal pair (H, L) displayed in column 3. The optimal pair (H, L) will give an intended $ARL_0 = 370.4$ when $\delta = 0$ for $m = +\infty$.

For example, when n = 5, $\delta = 0.3$, the optimal pair for minimising ARL_1 is $(H, L) = (2.2515, 22)$, for the SSGR chart with known process parameters ($m = +\infty$). The corresponding smallest $ARL_1 = 32.13$ is obtained while achieving the desired $ARL_0 = 370.4$ when $m = +\infty$. This optimal pair yields $ARL_1 \in \{60.59, 45.72, 39.65, 34.82, 33.16\}$ for $m \in \{30, 50, 80, 200, 500\}$. Note that when the process parameters are estimated from a small Phase-I samples, i.e. $m = 30$ with each of size $n = 5$, the corresponding ARL_1 (=60.59) differs from the ARL_1 (=32.13) when process parameters are assumed to be known. This shows that the performance of the SSGR chart deteriorates significantly when the process parameters are estimated due to the effect of parameter estimations. Nevertheless, when $m = 200$, the corresponding ARL_1 (=34.82)

is almost similar to the ARL_1 when process parameters are known. These results reveal that more than 80 samples are required for the SSGR chart with estimated process parameters to perform satisfactorily with the same chart with known process parameters.

Table 1
ARL_1s for n = {3, 5, 7, 9} with different combinations of (m, δ), based on the optimal pair (H, L) corresponding to the known process parameters case when ARL_0 = 370.4

n	δ	(K, L)	m					
			30	50	80	200	500	+ ∞
3	0.1	(2.4913, 63)	1425.24	635.25	439.44	319.14	283.85	263.24
	0.3	(2.3250, 30)	228.44	119.80	89.01	68.97	62.95	59.41
	0.7	(1.9948, 8)	8.72	7.46	6.92	6.48	6.32	6.22
	1.1	(1.8025, 4)	2.20	2.12	2.08	2.05	2.03	2.02
	1.5	(1.5953, 2)	1.28	1.27	1.26	1.25	1.25	1.25
	2.0	(1.5953, 2)	1.04	1.04	1.04	1.03	1.03	1.03
5	0.1	(2.4696, 57)	439.47	327.96	281.04	241.28	226.77	217.43
	0.3	(2.2515, 22)	60.59	45.72	39.65	34.82	33.16	32.13
	0.7	(1.8660, 5)	3.57	3.39	3.30	3.21	3.18	3.16
	1.1	(1.7185, 3)	1.36	1.35	1.34	1.33	1.33	1.33
	1.5	(1.5953, 2)	1.05	1.05	1.05	1.04	1.04	1.04
	2.0	(1.5953, 2)	1.00	1.00	1.00	1.00	1.00	1.00
7	0.1	(2.4537, 53)	299.81	248.96	223.56	199.29	189.72	183.38
	0.3	(2.1886, 17)	31.74	26.22	23.78	21.72	20.99	20.52
	0.7	(1.8025, 4)	2.29	2.23	2.20	2.17	2.15	2.15
	1.1	(1.5953, 2)	1.14	1.13	1.13	1.13	1.12	1.12
	1.5	(1.5953, 2)	1.01	1.01	1.01	1.01	1.01	1.01
	2.0	(1.5953, 2)	1.00	1.00	1.00	1.00	1.00	1.00
9	0.1	(2.4364, 49)	238.61	205.82	187.80	169.55	162.16	157.23
	0.3	(2.1401, 14)	20.21	17.47	16.21	15.11	14.72	14.46
	0.7	(1.7185, 3)	1.76	1.73	1.71	1.69	1.69	1.68
	1.1	(1.5953, 2)	1.06	1.05	1.05	1.05	1.05	1.05
	1.5	(1.5953, 2)	1.00	1.00	1.00	1.00	1.00	1.00
	2.0	(1.5953, 2)	1.00	1.00	1.00	1.00	1.00	1.00

In a real application, practitioners may not know the shift size in advance. Hence, if a practitioner considers a particular shift size, $δ$ and employs the corresponding optimal pair, the performance of the SSGR chart will be significantly affected if a different shift size actually occurs in the process. In view of this, EARL is crucial to adopt in place of ARL for designing and evaluating the SSGR chart.

The $EARL_0$ and $EARL_1$ denote the in-control and out-of-control EARL respectively. Table 2 displays the $EARL_1$ computed using the same (m, n) combinations considered in Table 1. The shift interval $(δ_{min}, δ_{max}) = (0.1, 1.0)$ and $(δ_{min}, δ_{max}) = (1.0, 2.0)$ are considered here, so that

it includes the exact shifts in the Table 1. For instance, the shift interval $(\delta_{min}, \delta_{max}) = (0.1, 1.0)$ includes the shifts $\delta = \{0.1, 0.3, 0.7\}$. Similarly, $(\delta_{min}, \delta_{max}) = (1.0, 2.0)$ in Table 2 is considered to include $\delta = \{1.1, 1.5, 2.0\}$.

Table 2
EARL₁s for n = {3, 5, 7, 9} with Different Combinations of (m, δ_{min}, δ_{max}), based on the optimal pair (H, L) corresponding to the known process parameters case when EARL₀ = 370.4

n	δ_{min}	δ_{max}	(K, L)	m					
				30	50	80	200	500	$+\infty$
3	0.1	1.0	(2.2821, 25)	143.27	81.67	62.56	49.33	45.17	42.69
	1.0	2.0	(1.7185, 3)	1.47	1.44	1.43	1.42	1.41	1.41
5	0.1	1.0	(2.2515, 22)	50.03	39.14	34.28	30.09	28.58	27.62
	1.0	2.0	(1.5953, 2)	1.12	1.12	1.11	1.11	1.11	1.11
7	0.1	1.0	(2.2403, 21)	31.56	26.43	23.89	21.57	20.70	20.13
	1.0	2.0	(1.5953, 2)	1.04	1.04	1.04	1.04	1.04	1.04
9	0.1	1.0	(2.2159, 19)	23.01	19.83	18.18	16.63	16.03	15.65
	1.0	2.0	(1.5953, 2)	1.02	1.01	1.01	1.01	1.01	1.01

For illustration, when $n = 3$, $\delta_{min} = 0.1$ and $\delta_{max} = 1.0$, the optimal pair $(H, L) = (2.2821, 25)$ yields the smallest $EARL_1 = 42.69$ when the process parameters are known. Meanwhile, this optimal pair (H, L) produces the intended $EARL_0 = 370.4$. However, the $EARL_1$ are 143.27, 81.67, 62.56, 49.33 and 45.17 for $m = 30, 50, 80, 200$ and 500, respectively. Comparing these with $EARL_1$ when process parameters are known (i.e. $EARL_1 = 42.96$), the $EARL_1$ values for the SSGR chart with estimated process parameters are significantly larger than the corresponding value of the same chart with known process parameters, especially when m is small. This provides clear insight that the $EARL_1$ value of the SSGR chart with estimated process parameters converge to the known process parameters case when m increases. This phenomenon of the SSGR chart is similar to that presented in Table 1 when ARL_1 is employed as a performance measure.

Furthermore, it is observed that the ARL_1 values computed using the optimal pair (H, L) for minimising $EARL_1$ in Table 2 are quite similar to the ARL_1 values computed using the optimal pair (H, L) for minimising ARL_1 in Table 1, as long as $\delta \in (\delta_{min}, \delta_{max})$.

For instance, when $n = 5$, $\delta_{min} = 0.1$ and $\delta_{max} = 1.0$ the optimal pair $(H, L) = (2.2515, 22)$ and the corresponding $EARL_1 = 27.62$ when $m = +\infty$. By considering $\delta = 0.7$ (i.e. $\delta \in (\delta_{min}, \delta_{max})$), the ARL_1 values are computed as 4.43, 4.27, 4.19, 4.12, 4.09 and 4.08 for $m = 30, 50, 80, 200, 500$ and $+\infty$ using $(H, L) = (2.2515, 22)$. These ARL_1 values are quite similar to those in Table 1 when $n = 5$ and $\delta = 0.7$, although the optimal pair (H, L) are different. In Table 1, when $n = 5$ and $\delta = 0.7$, the optimal pair is $(H, L) = (1.8660, 5)$. This suggests that the optimal pair (H, L) obtained by minimising $EARL_1$ is reliable to compute the ARL_1, as long as $\delta \in (\delta_{min}, \delta_{max})$. Note that the application of $EARL_1$ is more reasonable because practitioners may not have knowledge or experience to determine the exact process shift. Moreover, practitioners may not know the particular process shift that will occur in the process.

CONCLUSION

In summary, measures like ARL require practitioners to determine the exact shift size. In practical application, the particular shift size is usually unknown in advance. Hence, in this paper, the SSGR chart is designed to minimise $EARL_1$ when process parameters are known. The results reveal the optimal pair (H, L) obtained by minimising $EARL_1$ when process parameters are known can be used to compute ARL_1, as long as $\delta \in (\delta_{\min}, \delta_{\max})$.

Moreover, the study results indicate that a large number of Phase-I samples are required for the SSGR chart with estimated process parameters to perform favourably compared with the same chart with known process parameters case. This study can be extended to propose the optimal pair by minimising $EARL_1$ for the SSGR chart when process parameters are estimated. Furthermore, the expected value of the summary measure can be examined, for example, the expected value for percentile, when the exact shift size is unknown in advance.

ACKNOWLEDGEMENTS

REFERENCES

Castagliola, P., Celano, G., & Psarakis, S. (2011). Monitoring the coefficient of variation using EWMA charts. *Journal of Quality Technology, 4*3(3), 249-265.

Chakraborti, S. (2007). Run length distribution and percentiles: The Shewhart chart with unknown parameters. *Quality Engineering, 19*(2), 119-127.

Chen, Y., Birch, J. B., & Woodall, W. H. (2016). Effect of Phase I estimation on Phase II control chart performance with profile data. *Quality and Reliability Engineering International, 32*(1), 79-87.

Gadre, M. P., & Rattihalli, R. N. (2004). A group runs control chart for detecting shifts in the process mean. *Economic Quality Control, 19*(1), 29-43.

Gadre, M. P., & Rattihalli, R. N. (2007). A side sensitive group runs control chart for detecting shifts in the process mean. *Statistical Methods and Applications, 16*(1), 27-37.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology, 38*(4), 349-364.

Montgomery, D. C. (2012). *Introduction to Statistical Quality Control* (7th Ed.) New York, NY: Wiley.

Psarakis, S., Vyniou, A. K., & Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International, 30*(8), 1113-1129.

Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I., & Woodall, W. H. (2015). Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology, 47*(4), 363-382.

Sanusi, R. A., Abujiya, M. R., Riaz, M., & Abbas, N. (2017). Combined Shewhart CUSUM charts using auxiliary variable. *Computers and Industrial Engineering, 105*, 329-337.

Shepherd, D. K., Champ, C. W., & Rigdon, S. E. (2016). Properties of the Markov-Dependent attribute control chart with estimated parameters. *Quality and Reliability Engineering International, 32*(2), 485-498.

Teoh, W. L., Chong, J. K., Khoo, M. B. C., Castagliola, P., & Yeong, W. C. (2017). Optimal designs of the variable sample size chart based on median run length and expected median run length. *Quality and Reliability Engineering International, 33*(1), 121-134.

Woodall, W. H., & Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology, 31*(4), 376-386.

Woodall, W. H., & Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology, 46*(1), 78-94.

Wu, Z., & Spedding, T. A. (2000a). A synthetic control chart for detecting small shifts in the process mean. *Journal of Quality Technology, 32*(1), 32-38.

You, H. W., Khoo, M. B. C., Castagliola, P., & Ou, Y. (2015). Side sensitive group runs charts with estimated process parameters. *Computational Statistics, 30*(4), 1245-1275.

Zhang, Y., Castagliola, P., Wu, Z., & Khoo, M. B. C. (2011). The synthetic chart with estimated parameters. *IIIE Transactions, 43*(9), 676-687.

SCIENCE & TECHNOLOGY

PERTANIKA
JOURNALS

DSSBD: An intelligent Decision Support System for Residual Life Estimation of PN Junction Diode

Shivani and Cherry Bhargva*

Department of Electronics and Communication Engineering, Lovely Professional University, Phagwara, India

ABSTRACT

High reliability, high speed and low cost are the prime factors account for the complexity of electronic systems. Reliability and failure prediction are the major constraints to estimate the residual life of the component to anticipate the costly failures or system unavailability. Reliability prediction of passive components, especially PN junction diode, is of great concern as it is a critical element of bipolar junction transistors and other semiconductor devices, so the chances of failure as well as damage are increased as every component has its own characteristics and operating conditions. In this paper, artificial Intelligence techniques are employed on PN junction diode which embrace knowledge of failure mechanism of a component and predict the residual life of the component and a preventive action to be taken before serious breakdown occurs. The residual life calculated from experimental method is compared with artificial intelligence techniques, namely. ANN, fuzzy logic and ANFIS. The ANFIS has been proved as the most accurate system to predict remaining useful lifetime with an accuracy of 99.03%. A Graphical user interface is also designed based on fuzzy inference system, which indicates the remaining useful lifetime of PN junction diode.

Keywords: Accelerated life testing, ANFIS, artificial intelligence (AI), GUI, residual life

INTRODUCTION

The residual life prediction depends on the average life of the electronic components and on the operating conditions of a component. As many components are integrated on a chip, the chances of failure are increased, as each component has its own characteristics and operating conditions (Coldren, Corzine, & Mashanovitch, 2012). So, Intelligence techniques are implemented which increase

the performance of an individual component or complete system and conserve the working condition of a component such that if it drifts from actual output, a suitable action to be taken before any failure occurs in each electronic component such as diode (Al-Zubaidi, Ghani, & Haron, 2013). One of the major concerns for electronic engineers regarding diode is to predict its remaining useful life to protect it from costly failures and system unavailability. The remaining useful life of diode is determined by various environmental and electrical circumstances, such as temperature, humidity, atmospheric pressure and vibration. Electrical factors are operating voltage and current. Ambient temperature is also effective to ensure the longevity of diode. On the contrary, factors such as vibration, shock and humidity have little effect on the life of the diode(Kang, Kim, Choi, Kim, & Kwon, 2009). In this paper, diode deterioration under various operating conditions is discussed.

PN JUNCTION DIODE

Figure 1 shows an image of a PN junction diode. Temperature rise in PN junction diode due to internal current decreases the performance of diode.

Figure 1. PN junction diode

Operating conditions, such as temperature, voltage and current have a strong impact on the performance and utility of PN junction diode (Diehl, 2003). It is also well known the application of stress can change the energy gap of a semiconductor, the resulting change depending on the magnitude of the stress, as well as the direction with respect to the crystal orientation (Qin, Chung, Lin, & Hui, 2008). The expert system is designed using MATLAB GUI in which the fuzzy expert model is designed first and used to implement the final expert system user interface. This system can be designed using neuro-fuzzy technique to obtain more accurate results (Zhao, Chen, Guo, & Li, 2009).

METHODS AND MATERIALS

The residual life of pn junction diode was predicted using different techniques and methods as shown in Figure 2.
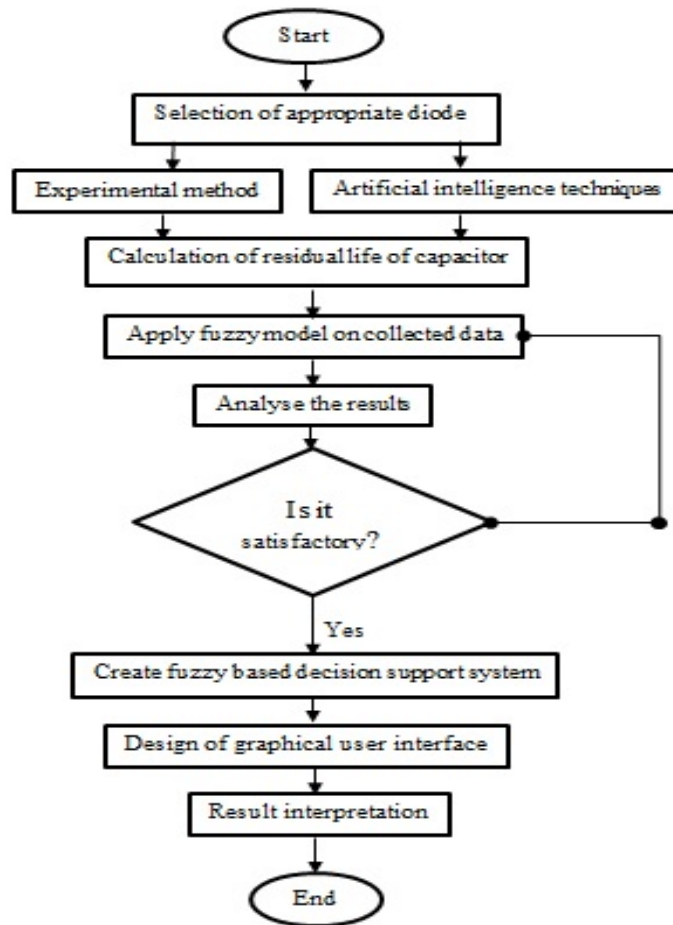
Figure 2. Flowchart of designing fuzzy-based decision support system

Figure. 2 shows the step wise procedure for calculating residual lifetime as well as designing of graphical user interface. The failure time was calculated using experimental method as well as artificial intelligence methods and then fuzzy logic was applied on the collected data. The ANN, fuzzy and ANFIS methods were employed as artificial intelligence techniques. The most accurate method was calculated based on experimental data.

Critical Parameters

The lifespan of diode is dependent on environmental and electrical factors; the environmental factors are temperature, humidity, atmospheric pressure and vibration and electrical factors are operating voltage and current (Silberberg et al., 1984). Temperature (ambient temperature) is the most important to increase the lifespan of diode while vibration, shock and humidity have little impact.

Temperature factor. As the ambient temperature of a diode increases its performance, the degradation begins and after a specified temperature, the diode fails and may harm the complete circuit. So, the higher temperature changes affect the lifetime of a diode (Dwyer, Franklin, & Campbell, 1990).

Voltage factor. The life of diode is affected less by applied voltage. When voltages above the rated one is applied, the internal current of diodes starts rising and more heat dissipation takes place through the diode and if the current rises, it will degrade the performance of diode completely and sometimes leads to failure of diode (Renge & Suryawanshi, 2008).

Life estimation of diode using experimental method

This method is mainly used by big manufacturing units where large number of samples or electrical units are subjected under tests (Fard & Li, 2009). These tests can be environmental, electrical or thermal. One example of such testing is Accelerated Life Testing (ALT) (Ahmad, Islam, & Salam, 2006). In accelerated life testing technique, a component is subjected to high temperature for limited hours, which ensures its remaining lifetime (Naikan & Rathore, 2016). The quality of a component during the life test is observed by regular checking of the value of the component using digital multimetre. Here, the testing was performed on diode based on the following steps (Huang et al., 2017):

a)   In the first step, 20 diodes were placed on the hotplate, the value of each diode was measured and the desired temperature level from 25°C to 160°C was set in the hot plate. Waited till the temperature achieves the maximum rate.

b)   The trial went on for 420 hours. This time length was chosen according to various temperature ranges. Time interval was shorter at higher temperature as chances of failure of components were greater compared with lower temperature limit.

c)   The value of every diode was measured and noted after two hours of using the multimetre. The number of diodes failed was checked after few hours and calculation of the output life was done.

Figure 3. Experimental setup

Life estimation of diode using artificial intelligence techniques

The remaining useful life of diode was also estimated using artificial intelligence techniques such as ANN, Fuzzy and ANFIS. Artificial intelligence techniques are the modern way of estimation as they show better results and estimation is done in an intelligent way like a human brain (Bhargava, Banga, & Singh, 2017).

Life estimation using ANN. Artificial Neural Network is an analogous system of human neural network which tries to mimic the functioning of actual brain. Input data along with target data has been fed to the network. The system will train itself and reduces the error after every epoch and hence, after specific number of epoch the best result is shown (Dylis & Priore, 2001). The number of neurons in the input layer consists of input parameters, such as temperature and time which are used to obtain the output life of electronic component (Yan, Koc, & Lee, 2004).
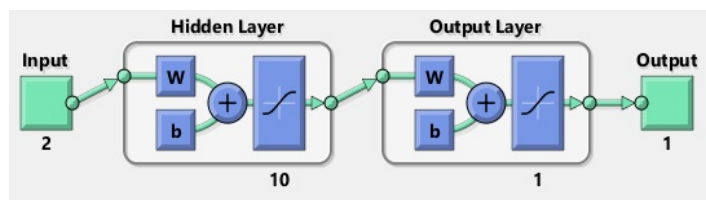
Figure 4. ANN structure

Life estimation of PN junction diode using Fuzzy Logic. Fuzzy Inference System is a soft-computing technique to design intelligent model with an advantage that it is user-friendly as it involves linguistic variable. For example, previously "True" or "False" or "Yes" or "No" or "0" or "1" were used. That showed only extreme value but in practical world we can have many other values like water problems. Water can't be just "Hot" or "Cold", it can be "Slightly cold", "Slightly Hot", "Slightly Moderate" among others. Hence this kind of Interpretation makes it easy for user to understand the response of the system (Kirby & Chen, 2007). That's Fuzzy and is known for handling ambiguity and uncertainty. Figure 5 shows the relation of membership function with the input parameters (Brevern, El-Tayeb, & Vengkatesh, 2009).
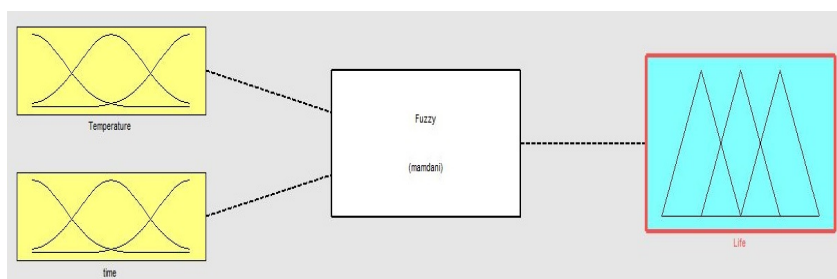
Figure 5. Fuzzy model for diode
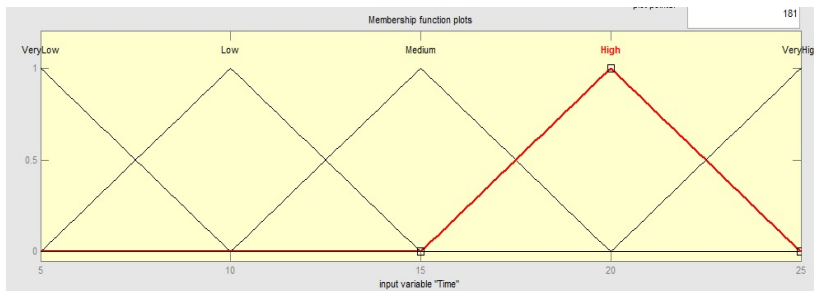
Figure 6. Triangular membership function for time
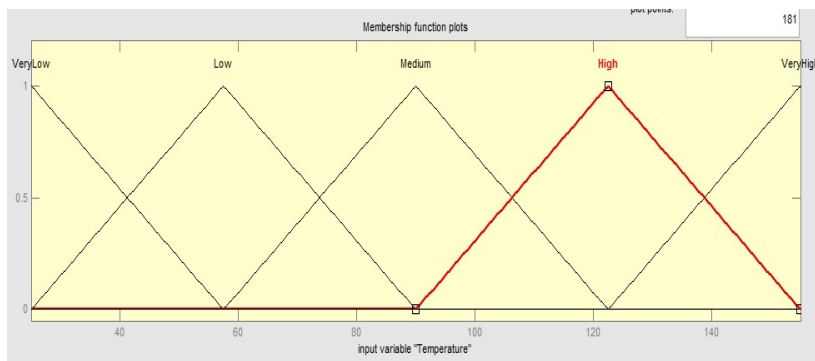
Figure 7. Triangular membership function for temperature
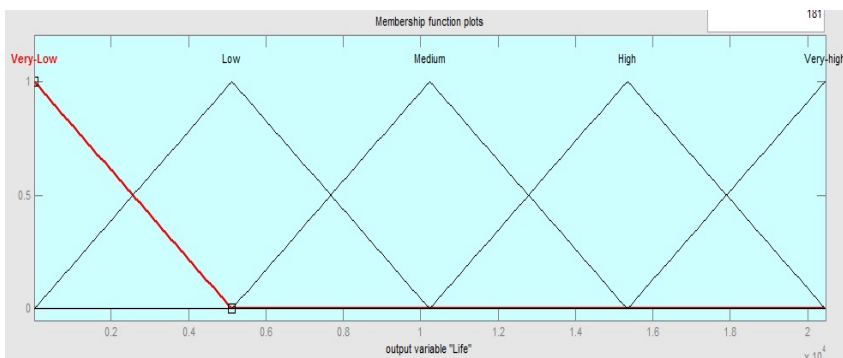
Figure 8. Triangular membership function for life

Life estimation of diode using ANFIS (Artificial neural fuzzy inference system). The ANFIS is Adaptive neuro fuzzy inference system and its architecture is shown in Figure 9. It is a Sugeno fuzzy model where the final fuzzy values are optimised using the artificial neural network training. The main difference between Mamdani and Sugeno system is that the Sugeno output membership functions are either linear or constant (Dragomir, Gouriveau, & Zerhouni, 2008). Therefore, the Sugeno method is computationally more efficient and thus, more likely to have a faster response. In this work, the Sugeno fuzzy inference system is selected. Here,

the applied temperature and maximum temperature are used as input parameters and life is used as the output parameter (Dragomir et al., 2008). The linguistic variables Very low (VL), low (L), medium (M) and high (H), very high (VH) were used for the inputs as well as for the output. Twenty rules were formed and the failure factor such as temperature of diode was kept into this system and rules were designed. The rules are shown in Figure 10. The triangular membership function was chosen for input variable which gave the least error when compared with the other membership functions (Chen, Zhang, & Vachtsevanos, 2012).
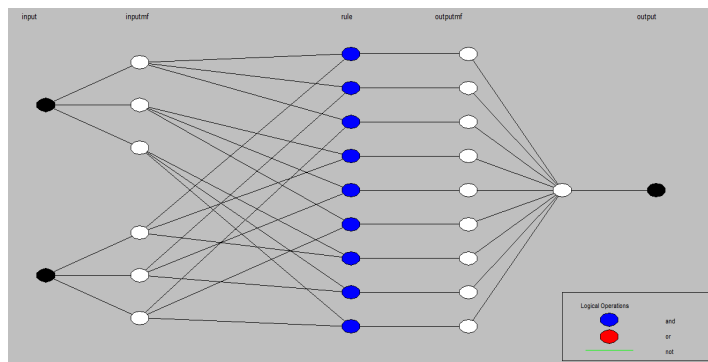
Figure 9. ANFIS structure
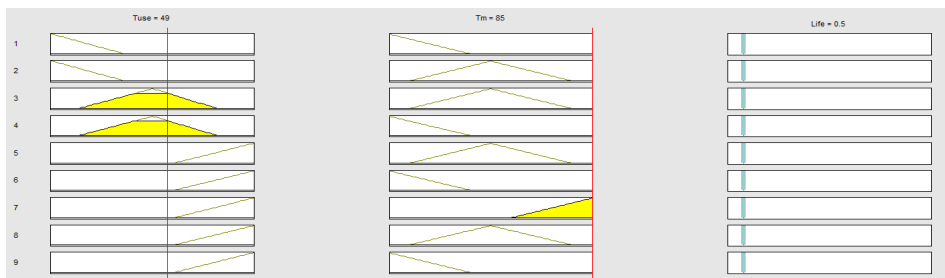
Figure 10. ANFIS Rule viewer (Yu, Zhou, Zhu, & Xu, 2012)(Yu, Zhou, Zhu, & Xu, 2012) (Yu, Zhou, Zhu, & Xu, 2012)

Design of GUI for Diode

Designing of graphical user interface (GUI) is the last phase of this method (Lefkowitz, 2000). The GUI user can interact with the expert system to check the operating condition of diode. The failure parameters of diode are given to this expert system as an input and the expert system gives the output as the remaining useful life. The GUI was created using the MATLAB-R2013a (Simulink & Natick, 1993). The entire database including rules was designed using the fuzzy logic. The following steps explain how a designed fuzzy fuzzy based decision support system works:

• In first step, data is collected using various techniques such as using experimental method, artificial intelligence modelling and analytical method.

- After data collection, numeric data consisting of input variables and output variable in the form of linguistic variables such as high, low or very low are formulated.

- Then, create a new fuzzy model using MATLAB and choose the number of input variables according to actual number of inputs and output variable in the same way.

- Now, choose a membership function such as triangular as per the range of data and fit the complete data in each variable using that membership function.

- Once all the input variables are set in membership functions in linguistic variables, design a rule base to check the output and if the output life is same as actual measured life then save and export the fuzzy file.

- In the next step, click on open new graphical user interface in MATLAB and save that file and after the file is saved, t two files are created one is .fig file and other is .m file. all the editing process is done in .fig file. Choose the static text boxes, edit text boxes and push buttons according to the number inputs and outputs.

- Name all the static boxes as the input variables and output variables and edit boxes are used to enter the input by the user and name the push button as calculate button.

- After designing the complete structure of expert system, right click on calculate and then click callback and view callback.

- In the callback, program is written using various commands in order to set and get all the input and output values and attach the saved fuzzy file in the program to evaluate and get the desired output. The file is saved after that.

- Verify the output by running the .m file from workspace window in MATLAB.
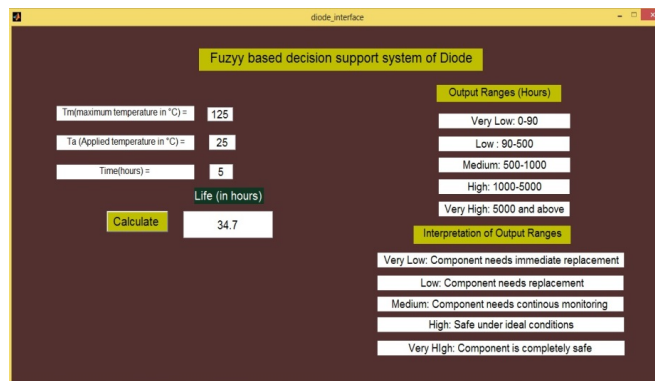
Figure 11. Decision support system for life estimation of PN junction diode

RESULTS AND DISCUSSION

This paper focuses on life estimation of pn junction diode. Various methods have been explored to design critical parameters. The diode residual lifetime is calculated using experimental technique and an intelligent system is created which predicts residual life of pn junction diode. The mean life is obtained using various methods as shown in Table 1 which shows the residual

meantime using accelerated life testing (experimental method) and artificial intelligence techniques. The accuracy of AI techniques has been decided by considering experimental data as reference. The ANFIS is proved as the most accurate method compared with ANN and Fuzzy logic.

Table 1
Comparison of residual life obtained using various methods

Techniques Used	Experimental Method	Artificial Intelligence Techniques		
	Accelerated Life Testing (ALT)	Artificial Neural Networks (ANN)	Fuzzy logic (FL)	Adaptive neuro-fuzzy inference system (ANFIS)
FIT (Failure In Time)	541917.3	573000.2	556096.2	547291.5
MTBF (Mean Time Between Failure) (hours)	1845.3	1745.2	1798.25	1827.18
Accuracy of AI techniques (%)	-	94.57%	97.45%	99.03%

It is found that an Artificial Intelligence technique i.e. adaptive neuro fuzzy inference system ANFIS provides highest accuracy to predict the useful life of diode i.e. 99.03%.

CONCLUSION

The residual life of diode is determined by experimental and artificial intelligence models to determine the relative effectiveness of each of the developed model. The diode is used in many of the electronic devices. The failure parameters of diode must be identified, and these parameters are necessary to ensure the reliable design of diode with an estimated lifetime. The residual lifetime of diode is estimated using experimental and artificial intelligence methods. The accuracy of life estimated using artificial intelligence techniques such as using ANN is 94.57%, while fuzzy shows an accuracy of 97.45%. It has been observed that the adaptive neuro fuzzy (ANFIS) method provides the highest rate of accuracy at 99.03%.

REFERENCES

Ahmad, N., Islam, A., & Salam, A. (2006). Analysis of optimal accelerated life test plans for periodic inspection: The case of exponentiated Weibull failure model. *International Journal of Quality and Reliability Management, 23*(8), 1019-1046.

Al-Zubaidi, S., Ghani, J. A., & Haron, C. H. C. (2013). Prediction of tool life in end milling of Ti-6Al-4V alloy using artificial neural network and multiple regression models. *Sains Malaysiana, 42*(12), 1735-1741.

Bhargava, C., Banga, V. K., & Singh, Y. (2017). Reliability comparison of a fabricated humidity sensor using various artificial intelligence techniques. *International Journal of Performability Engineering, 13*(5), 577.

Brevern, P., El-Tayeb, N., & Vengkatesh, V. (2009). Mamdani fuzzy inference system modeling to predict surface roughness in laser machining. *International Journal of Intelligent Information Technology Application, 2*(1), 12-18.

Chen, C., Zhang, B., & Vachtsevanos, G. (2012). Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms. *IEEE Transactions on Instrumentation and Measurement, 61*(2), 297-306.

Coldren, L. A., Corzine, S. W., & Mashanovitch, M. L. (2012). *Diode lasers and photonic integrated circuits* (Vol. 218). New Jersey: John Wiley & Sons.

Diehl, R. (2003). *High-power diode lasers: fundamentals, technology, applications* (Vol. 78). Germany: Springer Science and Business Media.

Dragomir, O., Gouriveau, R., & Zerhouni, N. (2008). *Adaptive Neuro-Fuzzy inference system for mid term prognostic error stabilization.* Paper presented at the International Conference on Computers, Communications and Control, ICCCC'08.

Dwyer, V., Franklin, A., & Campbell, D. (1990). Thermal failure in semiconductor devices. *Solid-State Electronics, 33*(5), 553-560.

Dylis, D. D., & Priore, M. G. (2001, January 22-25). *A comprehensive reliability assessment tool for electronic systems.* Paper presented at the Reliability and Maintainability Symposium (pp. 308-313). Philadelphia, USA.

Fard, N., & Li, C. (2009). Optimal simple step stress accelerated life test design for reliability prediction. *Journal of Statistical Planning and Inference, 139*(5), 1799-1808.

Huang, X., Denprasert, P. M., Zhou, L., Vest, A. N., Kohan, S., & Loeb, G. E. (2017). Accelerated life-test methods and results for implantable electronic devices with adhesive encapsulation. *Biomedical Microdevices, 19*(3), 46-57.

Kang, J. M., Kim, J. W., Choi, J. H., Kim, D. H., & Kwon, H. K. (2009). Life-time estimation of high-power blue light-emitting diode chips. *Microelectronics Reliability, 49*(9), 1231-1235.

Kirby, E. D., & Chen, J. C. (2007). Development of a fuzzy-nets-based surface roughness prediction system in turning operations. *Computers and Industrial Engineering, 53*(1), 30-42.

Lefkowitz, H. M. (2000). Graphical user interface. U.S. Patent No. 6,091,417. Washington, DC: U.S. Patent and Trademark Office.

Naikan, V., & Rathore, A. (2016). Accelerated temperature and voltage life tests on aluminium electrolytic capacitors: A DOE approach. *International Journal of Quality and Reliability Management, 33*(1), 120-139.

Qin, Y., Chung, H. S., Lin, D., & Hui, S. (2008). *Current source ballast for high power lighting emitting diodes without electrolytic capacitor.* Paper presented at the Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE.

Renge, M. M., & Suryawanshi, H. M. (2008). Five-Level Diode Clamped Inverter to EliminateCommon Mode Voltage and Reduce $ dv/dt $ inMedium Voltage Rating Induction Motor Drives. *IEEE Transactions on Power Electronics, 23*(4), 1598-1607.

Silberberg, Y., Miller, D., Gossard, A., Wiegmann, W., Smith, P., & Eilenberger, D. (1984). Passive mode locking of a semiconductor diode laser. *Optics Letters, 9*(11), 507-509.

Simulink, M., & Natick, M. (1993). *The mathworks*. In: Inc, USA.

Yan, J., Koc, M., & Lee, J. (2004). A prognostic algorithm for machine performance assessment and its application. *Production Planning and Control, 15*(8), 796-801.

Yu, Y., Zhou, T., Zhu, M., & Xu, D. (2012). *Fault diagnosis and life prediction of dc-link aluminum electrolytic capacitors used in three-phase ac/dc/ac converters*. Paper presented at the Second International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC) (pp. 825-830). Harbin, China.

Zhao, F., Chen, J., Guo, L., & Li, X. (2009). Neuro-fuzzy based condition prediction of bearing health. *Journal of Vibration and Control, 15*(7), 1079-1091.

PERTANIKA
JOURNALS

Theoretical Development of Biaxial Fabric Prestressed Composites under Tension-Tension Fatigue Loading

Nawras H. Mostafa[1,2]*, Z. N. Ismarrubie[1], S. M. Sapuan[1,3] and M. T .H. Sultan[4]

[1]*Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Department of Mechanical Engineering, Faculty of Engineering, University of Babylon, Babylon Province, Iraq*
[3]*Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[4]*Aerospace Manufacturing Research Centre, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

The improvement of the composite material against fatigue loading is of a great interest. In this study, the classical lamination theory of laminated composite was developed in order to include the effect of fibre prestressing on the composite's fatigue life when it was subjected to tension-tension fatigue loading. The biaxial fabric prestress term of the plain-weave composite (E-glass/polyester) was included in the theory and simplified. The overall tensile stress within the composite lamina was reduced by inducing compressive residual stress imparted from releasing the fibre pretension load. The fatigue life of the prestressed E-glass/polyester composite lamina was prolonged 36 times compared to the non-prestressed counterparts when the fabric was biaxially prestressed with 100 MPa.

Keywords: Biaxial fabric prestress, development of classical lamination theory, plain-weave fabric, residual stresses, tension-tension fatigue

INTRODUCTION

There is a need for high strength and lightweight materials in applications, such as in the aerospace, civil, automotive and sporting goods industries. Therefore, fibre–reinforced composite materials have sparked interest due to their high strength-to-weight ratio in comparison with most metals. The cost of composites, for instance, the cost of fibreglass–reinforced polymer is

approximately 60% materials and 40% fabrication (Ashby & Jones, 2012). Therefore, the focus on improving the fabrication techniques is to enhance the structural behaviour of the composites. Unfortunately, residual stresses are generated within the composites during manufacturing process (Safarabadi & Shokrieh, 2014). These stresses could diversely affect the structural behaviour of the fibre- reinforced composite. Residual stresses can happen in polymeric matrix composite due to the chemical shrinkage of matrix, the difference in the thermal expansion coefficients of the constituents, moisture content, and fibre prestress (Krishnamurthy, 2006). These stresses can develop within the matrix as a result of changing of the polymeric resin from liquid to solid phase. The mismatch of the thermal expansion coefficient between the reinforcement phase and the matrix phase can create undesired residual stresses in the composite structures when they are cooled down from their curing temperatures (Shokrieh, 2014; Krishnamurthy, 2006). Several methods have been used to minimise the detrimental effects of the induced residual stresses within the fibre reinforced composite such as optimisation of dwell cure cycle (White & Hahn, 1993b), curing the composite at low temperatures (Gopal et al., 2000) using electron beam curing (gamma irradiation) (Korenev, 2001), using expanding monomers (Fu et al., 2014), inserting shape memory alloy wires (Naghashian et al., 2014), and using fibre pretension method (Motahhari & Cameron, 1997).

Resin shrinkage and the difference in the thermal expansion coefficients of the composite's constituents throughout the polymerisation and curing processes, can be considered as the main sources behind badly by–product residual stresses in the polymeric matrix composites (Metehri et al., 2009). The state of internal stresses of a laminated composite can impact its fatigue behaviour critically (Naik, 2003). Residual stresses in the aluminium layer of carbon fibre/epoxy–aluminium alloy and vinylon fibre/epoxy–aluminium alloy laminates have positive effects on their fatigue behaviours (Lin et al., 1991; Sui et al., 1996). Prestrain was applied in the laminates by the post-cured process to change the aluminium internal stress state from tensile to compressive before testing the laminate under fatigue loading. This process prolongs the laminate fatigue cycles-to-failure. Fibre pretension could be one of the available options for enhancing the mechanical properties of polymeric matrix composites (PMCs) without increasing their section dimensions or mass (Graczykowski et al., 2016). The pretension was applied to the fibre before and during the matrix curing. After the matrix was cured, the fibre pretension was released. Tuttle (1988) stated that fibre pretension could prolong the fatigue life of the fibre prestressed laminated composite compared with the non–prestressed counterparts. Krishnamurthy (2006) applied different levels of fibre prestressing in unidirectional E–glass/epoxy composite. The prolong in the tension–tension fatigue life increased to about 75% when applying a fibre prestressing level of 51 MPa and a normalised peak stress (a ratio of maximum varying tensile stress to the ultimate tensile strength of the composite) equalled 0.4. The reduction in the tensile residual stress within the matrix due to the fibre prestressing was the main reason behind this improvement of fatigue life. Harris (1977) identified that fatigue damage of fibre–reinforced plastics depended on matrix and interface strengths. Therefore, any process that can increase the matrix strength against crack propagation or fibre–matrix bonding might enhance the fatigue life of the composite as well. Talreja (2016) showed that the fatigue

damage in the low stress fatigue region consisted of matrix cracking and fibre–matrix interfacial debonding. Fatigue life extension in the low fatigue stress region was due to toughening of the matrix against cracking as a result of compressive residual stresses within it (Krishnamurthy, 2006). Previous studies confirmed the advantages of fibre prestressed composites for plain-weave composites (Mostafa et al., 2016a, 2016b, 2016c, 2017a, 2017b). Fibre pretension (prestressing) during matrix cure has generally a positive effect on the composite mechanical properties as it induces compressive residual stresses in the matrix. This study was inspired by the fact pretension on the composite's strength for tension-tension fatigue loading has not been considered.

Therefore, the main objective of this study was to investigate the effect of biaxial fibre prestress on the fatigue life of the plain weave composite subjected to tension-tension cyclic loadings. In order to perform this task, the macro-mechanics lamination theory of the composite was developed in order to include the fibre prestressing effect and the resulting residual stresses within the composite's constituents.

MATERIALS AND METHOD

The composite used in this study was E-glass plain weave fabric/polyester resin system. Table 1 lists the mechanical properties of the composite's constituents.

Table 1
Mechanical properties of the composite's constituent materials (Mostafa et al., 2016a, 2016c)

Property	E-glass fabric	Polyester resin
Elastic modulus (E), GPa	70	2.77
Tensile strength (σ_{UTS}), MPa	2200	61
Poisson's ratio (v)	0.23	0.25
Thermal expansion coefficient (α), 10^{-6} m/(m°C)	120	5.4
Warp density × Fill density of the fabric (ends/m)	285 × 245	-

The analysis of the fibre reinforced composite was based on the classical lamination theory (CLT). The stress–strain ($\sigma - \varepsilon$) matrix was expressed by Jones (1999) as :

$$\begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{22} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \gamma_{12} \end{Bmatrix} \qquad [1]$$

The subscription 1 denotes the direction of the warp fibres and 2 as the direction of fill fibres.

The reduced stiffnesses Q_{11}, Q_{22}, Q_{12} and Q_{66} was:

$$\left.\begin{array}{l} Q_{11} = \dfrac{E_1}{1-\nu_{12}\nu_{21}} \\[4mm] Q_{22} = \dfrac{E_2}{1-\nu_{12}\nu_{21}} \\[4mm] Q_{12} = \dfrac{\nu_{21}E_1}{1-\nu_{12}\nu_{21}} = \dfrac{\nu_{12}E_2}{1-\nu_{12}\nu_{21}} \\[4mm] Q_{66} = G_{12} \end{array}\right\} \qquad [2]$$

where

G_{12} : Shear modulus of a composite.

E_1 : Elastic modules of a composite along the direction of fibres

E_2 : Elastic modules of a composite transverse to the direction of fibres

ν_{12} : The strain in the transverse direction over the strain in the direction of fibre when the stress is applied along the fibre's direction

ν_{21} : The strain in the direction of fibre over the strain in the transverse direction when the stress is applied along the transverse direction

Residual stress induced in the composite due to chemical shrinkage of the resin is relatively less compared to that of thermal variation (White & Hahn, 1993a). Therefore, the developed residual stress–strain relation of a laminated composite due to thermal variation and fibre prestress could be rewritten as:

$$\begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{Bmatrix}_{res} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{22} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \gamma_{12} \end{Bmatrix}_{res} \qquad [3]$$

$$\begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \gamma_{12} \end{Bmatrix}_{res} \begin{Bmatrix} \varepsilon_1^{pre} - \varepsilon_1^{ther} \\ \varepsilon_2^{pre} - \varepsilon_2^{ther} \\ 0 \end{Bmatrix} \qquad [4]$$

When curing the composite at a temperature different from the ambient temperature, thermal strain was generated within the composite's constituents. This strain was expressed by:

$$\left(\varepsilon_{1,2}^{ther}\right)_{res} = \alpha_{1,2}\,\Delta T \qquad [5]$$

where

$\alpha_{1,2}$: Coefficients of thermal deformation in the 1 and 2 directions, respectively

ΔT : Difference between the cure and cool-down temperatures

In the case of pretensioning the plain–weave fabric in the biaxial directions (i.e. the warp and the fill yarns), the induced residual strains ε_1 and ε_2 due to releasing the fibre pretension load after the matrix was cured can be expressed as:

$$\left(\varepsilon_1^{pre}\right)_{res} = \frac{(\sigma^{pre}V_f)_{1-dir}}{\bar{E}_1} - \frac{\bar{v}_{12}(\sigma^{pre}V_f)_{2-dir}}{\bar{E}_1} \tag{6}$$

$$\left(\varepsilon_2^{pre}\right)_{res} = \frac{(\sigma^{pre}V_f)_{2-dir}}{\bar{E}_2} - \frac{\bar{v}_{21}(\sigma^{pre}V_f)_{1-dir}}{\bar{E}_2} \tag{7}$$

where

$\bar{E}_1, \bar{E}_2$: Effective elastic modulus in warp and fill directions, respectively

$\bar{v}_{12}, \bar{v}_{21}$: Effective Poisson's ratio

V_f : Fibre volume fraction

The fibre and matrix are denoted by the subscriptions f and m respectively. At the micro–mechanical level, the strain in both the fibre and the matrix was equal to the lamina strain due to compatibility of strain (Mostafa et al., 2017a). Residual stresses in the fibre (σ_f^{res}) and the matrix (σ_m^{res}) at a ply level can be calculated using equations 8 and 9:

$$\sigma_f^{res} = E_f \left(\varepsilon_{1,2}^{pre} - \varepsilon_{1,2}^{ther}\right)_{res} + \sigma^{pre} \tag{8}$$

$$\sigma_m^{res} = E_m \left(\varepsilon_{1,2}^{pre} - \varepsilon_{1,2}^{ther}\right)_{res} \tag{9}$$

According to Gay's approach (Gay, 2015), the equivalent elastic properties were equal to:

$$\left. \begin{aligned} \bar{E}_1 &\approx \beta E_1 + (1-\beta)E_2 \\ \bar{E}_2 &\approx (1-\beta)E_1 + \beta E_2 \\ \bar{G}_{12} &= G_{12} \\ \bar{v}_{12} &\approx \frac{v_{12}}{\left(\beta+(1-\beta)\frac{E_1}{E_2}\right)} \\ \bar{v}_{21} &\approx \bar{v}_{12}\frac{E_2}{E_1} \end{aligned} \right\} \tag{10}$$

where

$$\beta = \frac{n_1}{n_1+n_2} \tag{11}$$

Here n_1 and n_2 denote the number of yarns per meter along the warp and fill directions that are equal to 285 and 245 respectively (see Table 1).

The effective composite thermoelastic properties of a unidirectional composite according to the concentric cylinder approach are equal to (Mostafa et al., 2017a):

$$
\left.
\begin{aligned}
E_1 &= E_f V_f + E_m V_m + \frac{4(\nu_f - \nu_m)^2 V_f V_m}{V_m/k^F + V_f/k^M + 1/G_m} \\
E_2 &= \frac{2(1 - \nu_f V_f - \nu_m V_m)\bar{k}E_1}{(E_1 + 4\bar{k}\nu_{12}^2)} \\
G_{12} &= G_m\left(\frac{G_m V_m + G_f(1+V_f)}{G_f V_m + G_m(1+V_f)}\right) \\
G_{m,f} &= \frac{E_{m,f}}{2(1+\nu_{m,f})}
\end{aligned}
\right\} \tag{12a}
$$

and

$$
\left.
\begin{aligned}
\nu_{12} &= \nu_f V_f + \nu_m V_m + \frac{(\nu_f - \nu_m)(1/k^M - 1/k^F)V_m V_f}{V_m/k^F + V_f/k^M + 1/G_m} \\
\nu_{21} &= \nu_{12}\left(\frac{E_2}{E_1}\right) \\
\alpha_1 &= \frac{\alpha_f E_f V_f + \alpha_m E_m V_m}{E_f V_f + E_m V_m} \\
\alpha_2 &= \alpha_m V_m + \alpha_f V_f + \left(\frac{\nu_f E_m - \nu_m E_f}{\frac{E_m}{V_f} + \frac{E_f}{V_m}}\right)(\alpha_f - \alpha_m)
\end{aligned}
\right\} \tag{12b}
$$

The axial residual stress t induced in the lamina due to the biaxial fabric prestress and thermal residual stresses is calculated from equation 3:

$$
\sigma_1{}^{res} = Q_{11}\,\varepsilon_1{}^{res} + Q_{12}\,\varepsilon_2{}^{res} \tag{13}
$$

where Q_{11} and Q_{12} in the above equation were determined from equation 2 that related the elastic properties of the composite's constituents with the fibre orientation angle; however, the $\varepsilon_1{}^{res}$ and $\varepsilon_2{}^{res}$ are the residual strains obtained from equations (3 and 4). Now, if the composite lamina is subjected to external axial tensile stress σ_{axi} in the direction of specimen length, the total axial stress σ_{axi}^{total} within the composite lamina at that direction would become:

$$
(\sigma_{axi})_{total} = \sigma_{axi} + \sigma_1{}^{res} \tag{14}
$$

The axial tensile stress in the composite resulted from applying external axial load without considering residual stresses is equal to:

$$
\sigma_{axi} = \frac{P_{axi}}{A_{total}} \tag{15}
$$

where σ_{axi}, P_{axi} and A_{total} represent the axial stress, axial load and the cross-sectional area of the composite material respectively. Consequently, by substituting equation 15 into equation 14. one can express the total internal axial stress in the prestressed composite as:

$$(\sigma_{axi})_{total} = \frac{P_{axi}}{A_{total}} + \sigma_1{}^{res} \qquad [16]$$

The relationship of stress–cycle (σ-N) fatigue data is usually defined by the power law equation according to Wöhler diagram. The stress-cycle relationship is expressed by:

$$\sigma_{max} = a\, N_f^b \qquad [17]$$

where σ_{max} is the maximum applied cyclic stress, N_f is the number of cycles to failure. The symbols a and b are constants. Equation 17 is developed to include the effect of the fibre prestress, such that:

$$(\sigma_{max})_{total} = \sigma_{max} + \sigma_1{}^{res} = a\, N_f^b \qquad [18]$$

or

$$(\sigma_{max})_{total} = a\, N_f^b - \sigma_1{}^{res} \qquad [19]$$

The constant a is equal to the ultimate tensile strength of the composite lamina (σ_{UTS}).

$$\sigma_{max} = \sigma_{UTS}\, N_f^b \qquad [20]$$

RESULTS AND DISCUSSION

The tensile residual stresses in the fabric when it was prestressed with different levels and released after matrix cure are listed in Table 2. It is clear that increasing the biaxial fibre prestressing level could increase both the tensile residual stress in the fibre and the compressive residual stress within the matrix. On the other hand, the induced residual stresses within the matrix at warp and fill directions of the fabric were not equal due to using unequal yarn number per meter along the principal directions of the used E-glass fabric.

Table 2
Tensile residual stresses in the fibre after matrix cure due to applying different biaxial fabric prestressing levels

Biaxial fabric prestressing (MPa)	Tensile residual stress in the fibre (MPa)	Residual stress in the matrix (warp direction) (MPa)	Residual stress in the matrix (fill direction) (MPa)
25	5.841	-1.14	-1.26
50	11.68	-2.27	-2.53
75	17.52	-3.39	-3.75
100	23.30	-4.52	-5.00

The axial stress in the lamina results from applying axial tensile stress that was developed in equation 14 is shown in Figure 1. The fibre volume fraction of the composite lamina was equal to 11% with cross-sectional area of 3 mm (thickness) × 25 mm (width). This value of fibre volume fraction was considered in previous studies(Mostafa et al., 2016a, 2016b, 2016c, 2017a, 2017b). The effect of the fibre prestressing was very clear in reducing the tensile stresses in the prestressed lamina when subjected to external axial tensile stress.
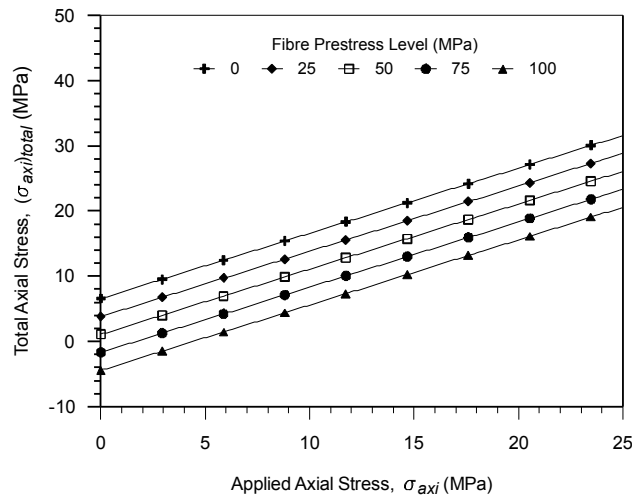
Figure 1. Total axial stress versus applied external stress in the composite lamina (E-glass/polyester) cured at 50°C and cooled down to 25°C and prestressed with different levels

In general, the tensile stress in the prestressed composite lamina could be reduced by increasing the biaxial fabric prestress (Mostafa et al., 2016a, 2016b). This reduction continues as prestressing level increases, but this behaviour does not always result in improving the mechanical properties of the composite as seen in many experimental studies (Krishnamurthy, 2006). The main reason behind this is the existence of interfacial shearing stress between the fibre and matrix (Mostafa et al., 2017b), which is not considered in the classical lamination theory.

Table 3 shows the results obtained from using equation 14 and the experimental data of the unidirectional prestressed E-glass/epoxy composite obtained by Hadi & Ashton, (1998). Elastic moduli of the used composite's constituents were 72.5 GPa and 3.45 GPa for the E-glass fibre and epoxy resin, respectively. The improvement in the tensile strength of the fibre prestressed composite due to applying different fibre prestressing levels is also reported by Hadi and Ashton (1998) for the composite fabricated with two different fibre volume factions (i.e. 35% and 60%). Table 4 shows the experimental data obtained by Mostafa et al. (2016a) and the current study results for a plain-weave E-glass/polyester composite system with two different equi-biaxial fabric prestressing levels. The ultimate tensile strength and the critical stress of the non-prestressed composite samples were equal to 66.5 MPa and 38.21 MPa (Mostafa et al., 2016a) respectively. Here, the critical stress represents the magnitude of stress at which the composite sample exhibits the first obvious sign of matrix fracture (onset

of failure). The findings of this study are supported by Mostafa et al. (2016a). The results of the developed fatigue equation were also compared with the available fatigue data obtained by previous studies. Fatigue data (cycles to failure) obtained by Krishnamurthy (2006) of the unidirectional E-glass/epoxy composite system with fibre volume fraction equals to 60% are listed in Table 5 along with the results of the developed theoretical fatigue equation. According Krishnamurthy (2006), the ultimate tensile strength (σ_{UTS}) of the used composite system and the fibre prestressing level was 1311 MPa and 51 MPa respectively. Different normalised peak stresses are used to validate the current results with Krishnamurthy's experimental data. The comparison revealed an acceptable convergence between the theoretical results and the experimental data with a maximum absolute error at 11.5%.

Table 3

Experimental tensile strength versus current theoretical results of a unidirectional e-glass/epoxy composite with different fibre prestressing levels

Fibre prestressing level (MPa)	Theoretical (current study) (MPa)		Experimental (Hadi & Ashton, 1998) (MPa)	
	V$_f$%			
	35	60	35	60
25	401	655	404±8.2	674±11.3
50	408	670	419±6.1	662±13.6
75	416	685	422±7.2	675±9.8
100	425	695	451±8.3	707±12.9
200	460	760	468±8.7	747±16.7

Table 4

Experimental tensile strength and critical stress versus current theoretical results of a plain-weave e-glass/polyester composite with different equi-biaxial fabric prestressing levels

Fibre prestressing level (MPa)	Theoretical (current study)		Experimental (Mostafa et al., 2016a)	
	Property in MPa			
	Ultimate tensile strength	Critical stress	Ultimate tensile strength	Critical stress
25	70.5	42.21	71.31	42.77
50	74.5	46.21	76.48	46.04

Table 5

Comparison of unidirectional e-glass/epoxy composite fatigue cycles to failure with different normalised peak stresses

Normalised peak stress [$(\sigma_{max})_{total}/\sigma_{UTS}$]	Cycles to failure	
	Theoretical (current study)	Experimental (Krishnamurthy, 2006)
0.40	1,260,716	1,130,147
0.45	436,434	455,483
0.50	109,544	108,532

The developed fatigue equation of the prestressed composite (equation 19) is plotted in Figure 2. For a given applied maximum fatigue stresses (σ_{max}), the fatigue cycles to failure (N_f) were investigated. For the case of comparing the results of the prestressed composites prestressed in the presence of different levels of biaxial fabric prestress, (a) and (b) could be assumed constants for the same composite material. As an example, (a) and (b) were assumed equal to 35 MPa (Mostafa et al., 2016a) and -0.2, respectively. The maximum applied fatigue stress is taken to be 15 MPa. The fatigue cycles to failure were increased by 36 times than the non-prestressed counterpart as the prestressing level was increased to 100 MPa. The prolong in the fatigue cycles to failure is caused by reducing the tensile residual stress induced in the composite lamina by thermal cooling down from 50°C to 25°C during the manufacturing process.
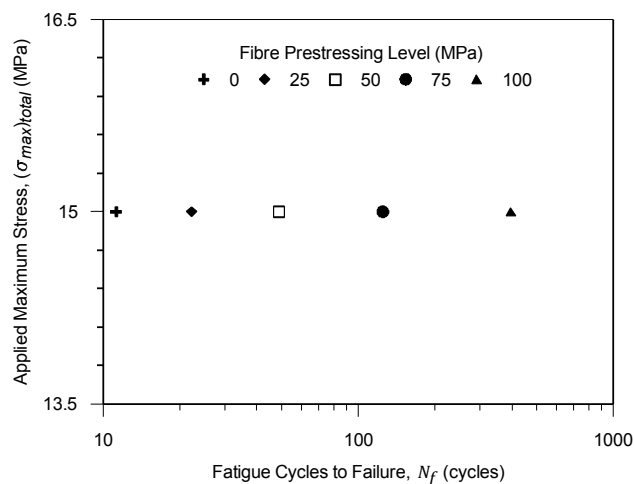
Figure 2. Applied maximum fatigue stress versus fatigue cycles to failure in the composite lamina (E-glass/polyester) cured at 50°C and cooled down to 25°C and prestressed with different levels

In general, the fatigue cycles to failure of the prestressed composite lamina is improved by increasing the biaxial fabric prestress. This improvement continues as the prestressing level increases. This behaviour is mainly due to increasing the compressive residual stress within the matrix; therefore, more cycles are required in order to start the damage in the composite.

CONCLUSION

The biaxial fibre prestress of the composite reinforced with plain weave fabric reduced the tensile residual stresses within the matrix of the composite lamina that resulted from different thermo–mechanical properties of the composite's constitutions. The tensile stress in the fibre prestressed composite lamina reduced by increasing the level of the biaxial fibre prestress as a result of compressive residual stresses within the lamina. This could improve the internal residual stress state of the composite subjected to tension-tension fatigue loading.

REFERENCES

Ashby, M. F., & Jones, D. R. H. (2012). *Engineering materials 1: An introduction to properties, applications, and design. engineering materials 1* (4th Ed.). Oxford: Elsevier.

Fu, J., Liu, W., Liu, X., Tuladhar, S. L., Wan, Q., & Wang, H. (2014). Properties of a new dental photocurable resin based on the expanding monomer and three-component photoinitiator system. *Journal of Wuhan University of Technology- Materials Science Edition, 29*(2), 384–390.

Gay, D. (2015). *Composite materials: Design and applications* (3rd Ed.). 6000 Broken Sound Parkway NW, Suit 300, Boca Raton, FL, USA: CRC Press (Taylor & Francis Group, LLC).

Gopal, A. K., Adali, S., & Verijenko, V. E. (2000). Optimal temperature profiles for minimum residual stress in the cure process of polymer composites. *Composite Structures, 48*(1–3), 99–106.

Graczykowski, C., Orlowska, A., & Holnicki-Szulc, J. (2016). Prestressed composite structures - modeling, manufacturing, design. *Composite Structures, 151*, 172–182.

Hadi, A. S., & Ashton, J. N. (1998). On the influence of pre-stress on the mechanical properties of a unidirectional GRE composite. *Composite Structures, 40*(3–4), 305–311.

Harris, B. (1977). Fatigue and accumulation of damage in reinforced plastics. *Composites, 8*(4), 214–220.

Jones, R. M. (1999). *Mechanics of composite materials. CRC Press* (2nd Ed.). Philadelphia, PA: Taylor & Francis.

Korenev, S. (2001). Electron beam curing of composites. *Vacuum, 62*(2–3), 233–236.

Krishnamurthy, S. (2006). *Prestressed advanced fibre reinforced composites: fabrication and mechanical performance.* (PhD thesis). Engineering System Department, Defence College of Management and Technology, Cranfield University, UK.

Lin, C. T., Kao, P. W., & Yang, F. S. (1991). Fatigue behaviour of carbon fibre-reinforced aluminium laminates. *Composites, 22*(2), 135–141.

Metehri, A., Serier, B., Bachir bouiadjra, B., Belhouari, M., & Mecirdi, M. A. (2009). Numerical analysis of the residual stresses in polymer matrix composites. *Materials and Design, 30*(7), 2332–2338.

Mostafa, N. H., Ismarrubie, Z. N., Sapuan, S. M., & Sultan, M. T. H. (2016a). Effect of equi-biaxially fabric prestressing on the tensile performance of woven E-glass/polyester reinforced composites. *Journal of Reinforced Plastics and Composites, 35*(14), 1093-1103.

Mostafa, N. H., Ismarrubie, Z. N., Sapuan, S. M., & Sultan, M. T. H. (2016b). The influence of equi-biaxially fabric prestressing on the flexural performance of woven E-glass/polyester-reinforced composites. *Journal of Composite Materials, 50*(24), 3385–3393.

Mostafa, N. H., Ismarrubie, Z. N., Sapuan, S. M., & Sultan, M. T. H. (2016c). Effect of fabric biaxial prestress on the fatigue of woven E-glass/polyester composites. *Materials and Design, 92*, 579–589.

Mostafa, N. H., Ismarrubie, Z. N., Sapuan, S. M., & Sultan, M. T. H. (2017a). Fibre prestressed composites: Theoretical and numerical modelling of unidirectional and plain-weave fibre reinforcement forms. *Composite Structures, 159*, 410–423.

Mostafa, N. H., Ismarrubie, Z. N., Sapuan, S. M., & Sultan, M. T. H. (2017b). Fibre prestressed polymer-matrix composites: A review. *Journal of Composite Materials, 51*(1), 39–66.

Motahhari, S., & Cameron, J. (1997). Measurement of micro-residual stresses in fiber-prestressed composites. *Journal of Composite Materials, 16*(12), 1129–1137.

Naghashian, S., Fox, B. L., & Barnett, M. R. (2014). Actuation curvature limits for a composite beam with embedded shape memory alloy wires. *Smart Materials and Structures, 23*(6), 1–10.

Naik, N. K. (2003). Woven-fibre thermoset composites. In B. Harris (Ed.), *Fatigue in composites: Science and technology of the fatigue response of fibre-reinforced plastics* (1ˢᵗ Ed., pp. 296–313). Cambridge, UK: Woodhead Publishing Ltd and CRC Press LLC.

Safarabadi, M., & Shokrieh, M. M. (2014). *Residual stresses in composite materials*. Cambridge: Woodhead publishing.

Shokrieh, M. (2014). *Residual stresses in composite materials. Residual stresses in composite materials.* Cambridge: Woodhead Publishing Limited.

Sui, G. X., Yao, G., & Zhou, B. L. (1996). Effects of prestrain on the fatigue of VIRALL laminate properties. *Composite Science and Technology, 56*(8), 929–932.

Talreja, R. (2016). Fatigue damge mechanisms. In R. Talreja & J. Varna (Eds.), *Modeling damge, fatigue and failure of composite materials* (pp. 25–39). Cambridge, UK: Woodhead Publishing.

Tuttle, M. E. (1988). A mechanical/thermal analysis of prestressed composite laminates. *Journal of Composite Materials, 22*(8), 780–792.

White, S. R., & Hahn, H. T. (1993a). Cure cycle optimization for the reduction of processing-induced residual stresses in composite materials. *Journal of Composite Materials, 27*(14), 1352–1378.

White, S. R., & Hahn, H. T. (1993b). Cure cycle optimization for the reduction of processing-induced residual stresses in composite materials. *Journal of Composite Materials, 27*(14), 1352–1378.

Drying Characteristics of *Curcuma longa* Using Solar Dryer

Fhelix August Soebiantoro¹, Elieser Tarigan²,³, Lie Hwa¹, Violita Putri Halim¹ and Lanny Sapei¹*

¹*Department of Chemical Engineering, Faculty of Engineering, University of Surabaya, Raya Kalirungkut, Surabaya 60293 East Java, Indonesia*
²*Department of Electrical Engineering, Faculty of Engineering, University of Surabaya, Raya Kalirungkut, Surabaya 60293 East Java, Indonesia*
³*Center for Renewable Energy Studies, University of Surabaya, Raya Kalirungkut, Surabaya 60293 East Java, Indonesia*

ABSTRACT

Curcuma longa (turmeric) is a rhizomatous herbaceous perennial plant which is widely cultivated in tropical regions, such as Indonesia. It has been intensively used for medical purposes as an ingredient of traditional medicine for a long time. In order to extend its shelf-life, turmeric is generally dried under the sun prior to use. This method generally takes a longer time and is less controlled, thus yielding less qualified products. This experiment was carried out using a manually designed solar dryer for improving the drying process of turmeric. The drying process using solar dryer is shorter and also protects the samples from dirt, insects, as well as direct contact with UV radiation which may deteriorate its bio-active compounds. The drying was done until moisture of about 10% content was achieved. In general, it took about 450 minutes in the solar dyer in comparison to that of 480 minutes using sun drying to reach the equilibrium moisture content. This was caused by a higher temperature profile distributed inside the solar dryer accelerating the drying process. The use of solar dryer has the potential to be further developed to replace the conventional sun drying method of herbs.

Keywords: Curcuma longa, drying process, solar dryer, sun drying, turmeric

INTRODUCTION

Curcuma longa (turmeric) is known as *kunyit* by Indonesians. This herb is very popular in Indonesia because of its benefits. Turmeric has been used as a spice for traditional Indonesian cuisines, as a traditional drink and the most popular as the main ingredient for traditional medicines because of its high antioxidant contents. Turmeric has also been reported to

have a role in preventing diseases such as cancer due to its powerful antioxidant properties and anti-carcinogenic action. It is also said to prevent cardiovascular diseases (Prathapan, Lukhman, Arumughan, Sundaresan, & Raghu, 2009).

Herbs including turmeric must be initially well-dried before being used as the ingredients of traditional medicines. This helps reduce the moisture content, thus also functions as a disinfection, microbial decontamination and long-term preservation in order to prolong its shelf-life (Schweiggert, Carle, & Schieber, 2007). In Indonesia, herbs are simply dried under the sun immediately after they are harvested. The direct contact of sunlight could destroy the bioactive compounds retained in the samples. Besides that, the drying process is uneasy to maintain since temperature and humidity tend to fluctuate. Furthermore, there has been a greater risk for being contaminated due to insects, birds, or other animals, and dust (Weiss & Buchiner, 2001). Solar dryer has become one of the drying alternatives harnessing the sun energy but in a more controlled way and could eliminate the contaminants. Heat from the sun is collected and transferred indirectly to the herbs, thus preventing the deterioration of their bioactive compounds and improving the overall product quality. Solar dryer has been categorised into 2 types - either active or passive, depending on the presence of fan / blower facilitating the air flow inside the solar dryer.

The main purpose of this work is to study the drying characteristics of turmeric using a passive type solar dryer. The solar dryer was self-designed by Suyanto and Antoro (2016) and had been continually modified to improve its performance for drying herbs in a more efficient and controlled way.

MATERIALS AND METHODS

Turmeric (*Curcuma longa*) was purchased from Jagir Local Market, Surabaya, Indonesia. The turmeric was separated from dust and dirt. It was then sliced horizontally with a thickness of about 3 mm and weighed about 7 grams as a sample using a balance (Mettler, Toledo) before being dried. It was dried under the sun and in the solar dryer until reaching a constant weight and the final moisture content was below 10% (Ministry of Health Republic of Indonesia [KEMKES], 1994).

The solar dryer was self-designed using aluminum for the body, plywood board for the door, glass for the heat collector, and black painted aluminum for absorbing heat from the sunlight. There were four trays arranged vertically as sample containers. The details of the solar dryer design are shown in Figure 1.
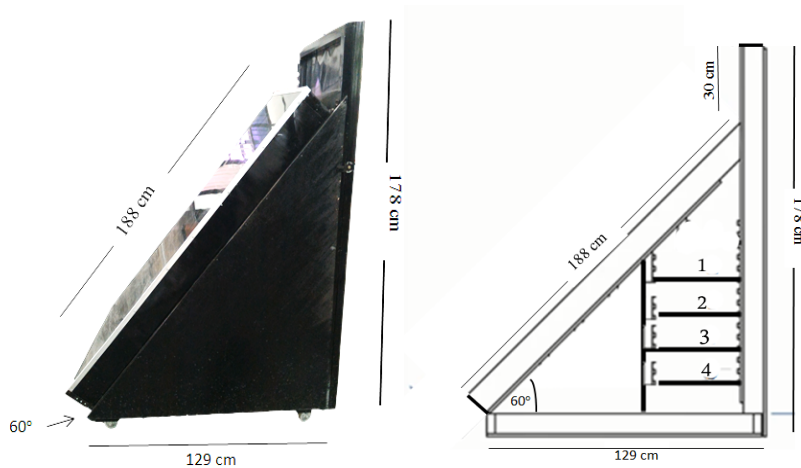
Figure 1. Self-designed solar dryer

Approximately 2 kg turmeric samples were distributed on all trays (tray 1, 2, 3, 4). There was a boundary circle area in the middle of each tray containing of three to four slices for samplings. The temperature of each tray was recorded using thermocouple every five minutes and the sample was weighed every 30 minutes. All trays were rotated downwards every two hours (Figure 2) in order to equalise heat distribution absorbed in each tray since the first tray generally received maximum sun radiation. The solar dryer was moved from time to time following the sun direction.
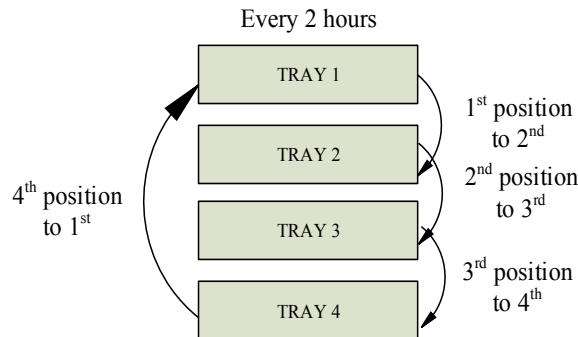
Figure 2. The sequence of tray rotation

The sampling process was conducted in two days. On the first day, sampling was conducted from 9.00 a.m. until 3.00 p.m. The data was simply continued by the sampling taken on the following day starting from 9.00 a.m. until it reached equilibrium weight. Sun drying was also conducted for the comparison of the drying process conducted in the solar dryer. The turmeric

slices were evenly distributed on a tray and placed under the sun until the drying process was completed. During the samplings, two slices of turmeric samples in the middle of the tray were weighed every 30 minutes. The air temperatures, both dry bulb and wet bulb temperatures were measured using alcohol thermometers and the corresponding RHs were calculated using the psychrometric chart.

Results were presented as drying curves whereby free moisture content (X) was monitored versus the drying time (Geankoplis, 2003). Free moisture content was obtained by deducting the equilibrium moisture content (X^*) from the moisture content at certain time (X_t). The calculation of X_t and X can be seen in equation (1) and (2).

$$X_t = \frac{W_t - W_d}{W_d} \qquad [1]$$

where X_t = moisture content at certain time (kg H_2O/kg dry weight); W_t= sample weight at certain time (kg); W_d= sample dry weight (kg) obtained after sample was dried at 120°C for about two hours.

$$X = X_t - X^* \qquad [2]$$

where X = free moisture content (kg H_2O/kg dry weight); X_t = moisture content at certain time (kg H_2O/kg dry weight); X^*= equilibrium moisture content (kg H_2O/kg dry weight). The drying rates were also analysed using equation (3) (Geankoplis, 2003).

$$R = R_c.A = -L_s \frac{dX}{dt} \qquad [3]$$

where R_c= constant drying rate (kg H_2O/ minutes), Ls= sample dry weight (kg), dX/dt= rate of free moisture changes per time (kg H_2O/ (kg dry weight. minutes)). The data was processed using Microsoft Excel 2013 and Curve Expert Professional 2.3.0.

RESULTS AND DISCUSSION

Sun Drying

Drying process under the sun was dependent on the ambient air temperatures and RH. The temperatures and RH profiles during the experiment are presented in Figure 3. Dry air temperatures ranged from 35°C to 41°C with mean temperature of about 38.5°C, and RHs were in the range of about 51% to 65%.
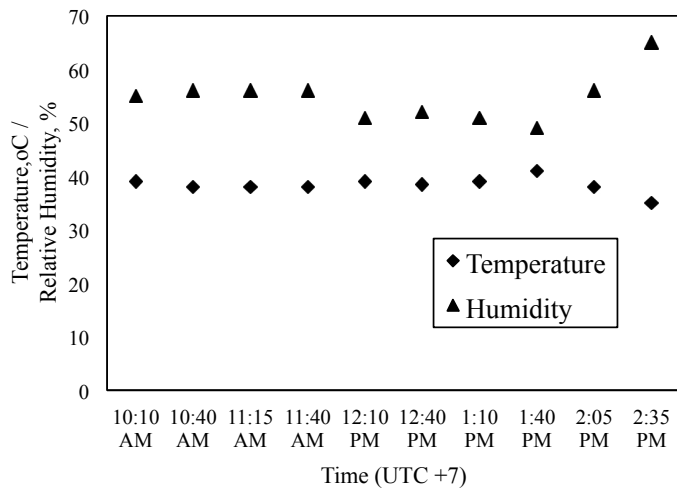
Figure 3. Temperature and relative humidity profiles during sun drying

The drying characteristics of turmeric using sun drying mode can be seen in Figure 4. The free moisture contents continuously decreased until it became constant. Drying began with the constant rate (from 0 minutes until 150 minutes) followed by the falling rate (from 150 minutes until it reached constant moisture content). Drying of turmeric under the sun took about eight hours (480 minutes) to reach the required moisture content of 10% or free moisture content of about 11.1%, according to Indonesian government regulations.
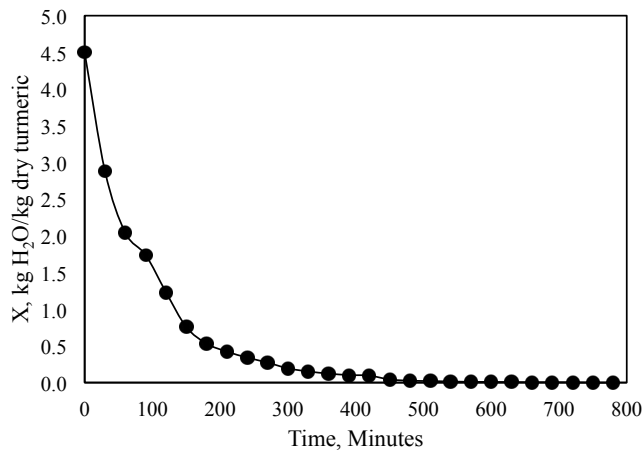
Figure 4. Free moisture content versus time during sun drying

Solar Drying

During the experiment using the solar dryer, the measured temperatures on each tray fluctuated depending on heat intensity. The highest temperature achieved during drying was about 55°C and occurred between 12:24 p.m. to 1.00 p.m. The temperature profiles on each tray inside the solar dryer are shown in Figure 5.
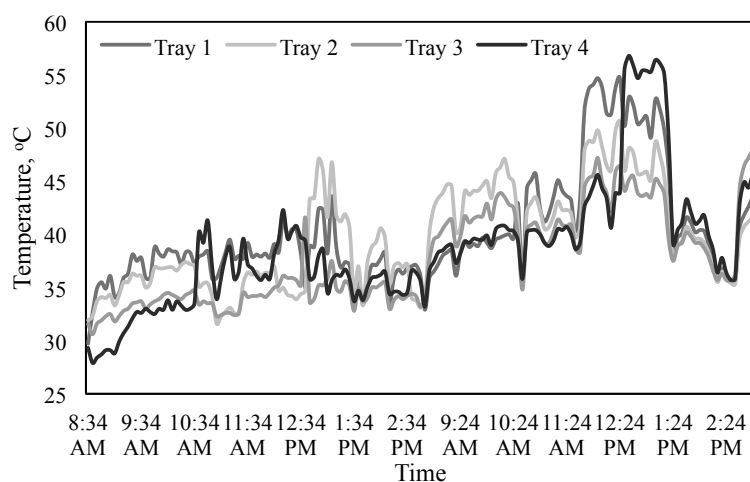
Figure 5. Temperature profiles on each tray inside the solar dryer

The trays which were placed on the top received the highest heat radiation from the sun than the others, thus would reach the highest temperature leading to sample overheating compared to the samples placed on the lower trays. Therefore, trays were rotated during the experiment as shown in Figure 2, so that the temperatures were more evenly distributed amongst the trays. It was evident that the temperature profiles in all trays did not differ very much. The average temperature on each tray was about 37°C to 40°C and the average was about 39°C, which was slightly higher than the average air temperature during sun drying. The temperature inside the solar dryer generally was higher compared to the ambient temperature during sun drying (Weiss & Buchiner, 2001). This was due to heat accumulation in the sunlight collector which was then transferred to the air drying up the samples.

The turmeric drying characteristics in each tray can be seen in Figure 6. Free water content of the turmeric decreased over time until it became constant. Furthermore, after 360 minutes, the free water content of turmeric dried in all four trays approached zero, even though the initial water contents were different and all reached the equilibrium water contents after about 450 minutes of drying. This indicates that homogeneous heat transferred on each rotated tray. Borah, Hazarika and Khayer (2015) previously also dried turmeric using a passive solar dryer. It took about 11 hours to dry the samples until reaching the constant weight. This indicates comparable performance of this self-designed solar drying with other existing passive solar dryers.
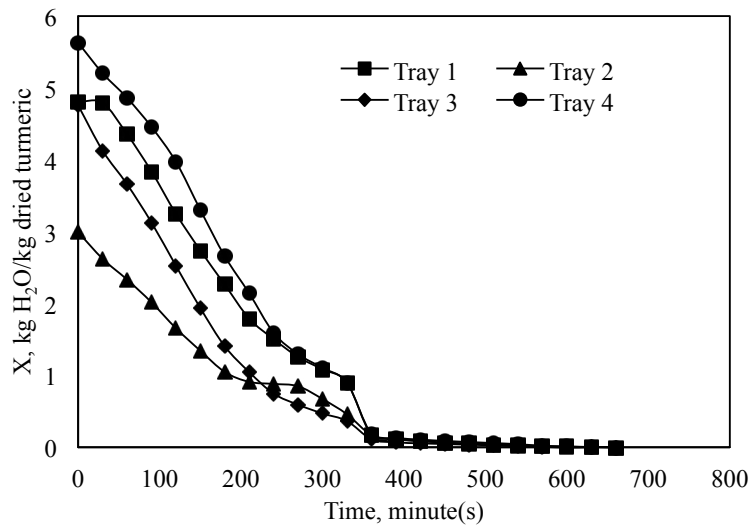
Figure 6. Free moisture content versus time during solar drying

Drying of turmeric at constant rates was indicated by the abrupt linear decrease at the beginning of the drying process. The slope corresponded with the constant drying rate. The steeper the slope the faster the rate of drying was. The differences in these constant drying rates could be due to unsteady temperature profiles (Figure 5) during the rotation of the trays. The calculated constant drying rates of turmeric dried in both solar dryer as well as under the sun are shown in Table 1.

Table 1
Constant drying rate of turmeric

Drying	Tray	Constant drying rate (R_c) [kg H_2O/minutes]
Solar Drying	1	0.01869
	2	0.01863
	3	0.01749
	4	0.01988
Sun Drying	-	0.02442

The constant drying rate of turmeric dried under the sun was a little higher than that dried in the solar dryer. This could be due to more dynamic air flow during sun drying accelerating the drying process during the first period. However, it took 480 minutes for drying turmeric under the sun compared to the time required in the solar dryer of about 450 minutes, until reaching the equilibrium moisture content. This was plausible that the falling rate of turmeric in solar dryer was higher than that of sun drying. Falling rate occurred after the constant rate when rate of water diffusion from the sample surfaced to the air and was higher than the rate of water diffusion from the pores inside the samples to the surface. During the overall drying process,

the average temperatures in the solar drying were slightly higher than the air temperature during sun drying. This would then increase the water mass transfer to the air due to the increase gradient between the saturated vapor pressure inside the sample and partial vapor pressure in the air (Geankoplis, 2003). The colours of the samples dried inside the solar dryer were more preserved compared to those dried under the sun based on the visual observation (Figure 7). This result shows the potential use of the solar dryer for drying herbs in order to prolong their shelf life as well as to ensure their quality. To be able to be used in an active mode, the modification of solar dryer should be further developed in order to improve its performance.
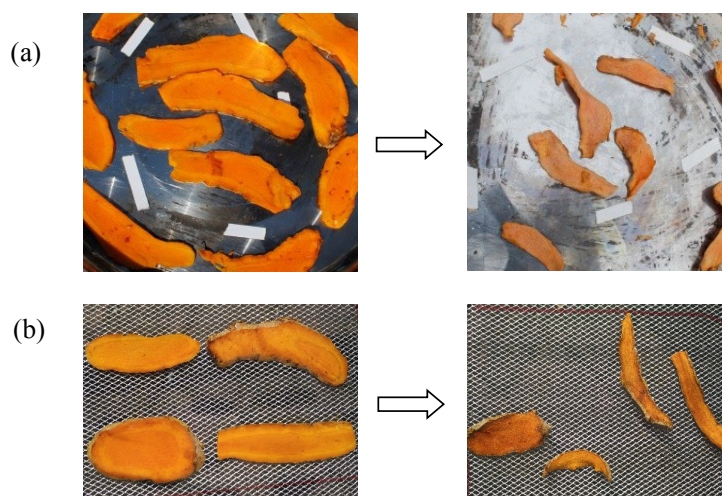
Figure 7. (a) Turmeric (*Curcuma longa*) before and after being dried under direct sun; (b) Turmeric (*Curcuma longa*) before and after being dried in the Self-designed solar dryer

CONCLUSION

Drying of turmeric (*Curcuma longa*) was conducted in a manually designed solar dryer which consisted of four trays. The drying process in all trays was quite similar and the average temperatures on the trays were about 40°C to 41°C. Overall, the drying process of turmeric using the solar dyer was better in comparison to that under the sun in terms of drying time and sample quality. It took about 450 minutes in the solar dryer compared to 480 minutes under the sun for the samples to reach the equilibrium moisture content. The colour of turmeric was better preserved in the solar dryer since the UV radiation did not directly impinge on the samples. Thus, solar dryers should be widely developed and used for producing improved quality of herbs.

REFERENCES

Borah, A., Hazarika, K., & Khayer, S. M. (2015). Drying kinetics of whole and sliced turmeric rhizomes (*Curcuma longa* L.) in a solar conduction dryer. *Information Processing in Agriculture, 2*(2), 85-92.

Geankoplis, C. J. (2003). *Transport process and separation process principles* (4th Ed.). New Jersey, USA: Prentice Hall.

Ministry of Health Republic of Indonesia (1994). *Keputusan Menteri Kesehatan Republik Indonesia Nomor: 661/IMENKES/SK/VII/ 1994 tentang persyaratan obat tradisional* [Decree of the minister of health of the republic of Indonesia number: 661 / IMENKES / SK / VII / 1994 about traditional medicines requirements]. Jakarta, Indonesia: Ministry of Health.

Prathapan, A., Lukhman, M., Arumughan, C., Sundaresan, A., & Raghu, K. G. (2009). Effect of heat treatment on curcuminoid, colour value and total polyphenols of fresh turmeric rhizome. *International Journal of Food Science Technology, 44*(7), 1438–1444.

Schweiggert, U., Carle, R., & Schieber, A. (2007). Conventional and alternative processes for spice production - A review. *Trends Food Science Technology, 18*(5), 260–268.

Suyanto, T. A., & Antoro, D. A. (2016). *Desain pengering tenaga surya (solar dryer) untuk pengeringan bahan herbal* [Design of solar dryers (diesel dryer) for drying herbs]. Surabaya, Indonesia: University of Surabaya.

Weiss, W., & Buchinger, J. (2011). *Solar drying*. Austria. Retrieved from http://www.aee- intec. at/0uploads/dateien553.pdf

Exploration of Tritrophic Interaction for Enhancing Conservation Biological Control of Insect Pest, the Role of Analytical Chemistry

Surjani Wonorahardjo[1]*, Nurindah[2], Dwi Adi Sunarto[2], Sujak[2] and Setya Ayu Aprilia[1]

[1]*Department of Chemistry, Faculty of Mathematics and Science, Universitas Negeri Malang, Jl. Semarang 5 Malang, Indonesia*
[2]*Indonesian Sweetener and Fibre Crops Research Institute, Jl. Raya Karangploso, 199 Malang, Indonesia*

ABSTRACT

Tritrophic interaction among plants, herbivores and natural enemies of the herbivores is mediated by volatiles. Tritrophic interaction model using rice plant, brown planthopper (BPH), and egg parasitoid of BPH were used to study the volatiles involved in the interaction. Extraction of chemicals in rice plants, determination of the extracted volatiles using GC-MS, and bioassays of the volatiles to analyse the orientation behavior of the egg parasitoid were methods used to study the involvement of the volatiles in the interaction. The knowledge of chemistry involved in such complicated natural interaction is beneficial for enhancing conservation biological control (CBC) of BPH using natural population of its egg parasitoid. Implementation of the positive volatile compounds attract the egg parasitoid as a component of CBC, materials involved that adsorb and release the volatiles efficiently. A material made of integrated bio-silica-cellulose was proven to be an effective adsorbent. Analytical chemistry was important in supporting these findings. Furthermore, in order to increase the efficiency of the volatiles that are positively involved in the tritrophic interaction, developing an attractant formula using synthetic-similar compounds and bio-silica-cellulose adsorbent is valuable in enhancing CBC.

Keywords: Biosilica-cellulose adsorbent, chemical cues, tritrophic interaction

INTRODUCTION

One important phenomenon in nature is tritrophic interaction, in which some volatile compounds are involved. It is the effect of plants on attack of natural enemies like plant parasites. Plant chemistry, morphology and resources will affect the behaviour and performance of natural enemies of the plant

parasites (Agrawal, 2000). More modern research in overcoming agricultural problems would take ecological approach rather than using poisonous chemicals to stop the presence of certain insect pests. Some studies of tritrophic interactions have revealed the mechanism of plant defense as well as the consequences in food chain and ecosystem.

When related to plant defense, a model using rice plant, brown planthopper (BPH), and egg parasitoid of BPH (*Anagrus nilaparvatae*) were used to study the volatiles involved in the interaction (Wonorahardjo, Nurindah, Sunarto, Sujak, & Zakia, 2015). Tritrophic interaction is a usual defense mechanism describing ecological impacts of each component which are the plant, herbivore, and parasitoid of the herbivore. While the BPH attack the plants, the wound would induce the plants to produce chemicals that are emitted into the air. This is the sign of host locator for the parasitoids (Agrawal, 2000). The defense mechanism of paddy (*Oryza sativa* L.) is the natural way to make tritrophic interaction work at a bigger scale. In this case the plant was chosen as an example. Biochemistry of the process is really complex but in the end, some new volatiles would be released together with old chemicals to give special scents for the sign of host location. The activation of some genes involved in encoding vegetative storage proteins and repressing genes encoding proteins involved in photosynthetic process is induced by compounds such as jasmonates, as derivations of long chained carboxylic acid (Bi, Zeng, Su, Min, & Luo, 2007) and the presence of these compounds can be found in any plant. The chemical compounds involved in such interactions are mainly called biological volatile organic compounds (BVOCs) and are defined as any organic compounds with vapour pressures high enough under normal conditions to be vaporised into the atmosphere and act as the chemical cues in opportunity or danger (Yuan, Himanen, Holopainen, Chen, & Stewart, 2009).

In tropical lands, there are more complex patterns of tritrophic interactions (Dyer, 2005). The soft climate changing compared to four seasonal places enable the ecosystem with more diverse components to live for multitrophical interactions. The ecosystems in altitudinal gradients show extreme differences in diversity, productivity as well as trophic abundance compared to those from the declined altitude. The ambient temperature and humidity (Dyer, 2005) plays a role for the optimum conditions of more complex possibilities.

It seems that analytical chemistry works in smaller scale of object, and is far too small compared to environmental scale, such as the tritrophic interaction. The task of analytical chemistry is to answer the questions of what and how much in chemistry scale, while environmental scale addresses questions of why and how to describe its phenomena as well as some steps to 'imitate' such mechanism.

This study works out the chemical cues in the tritrophic interaction of analytical chemistry. Extraction method provides many possibilities to take the real chemicals under investigation by choosing solvents, and providing the best conditions for the extraction. More apparatus and analyses as well as methods can be set up in modern phytochemical methods, which aim for analysis of secondary metabolites and their roles in wider scopes. After extraction, the analysis for compound separation and identification using GC/MS was done. Some critical analysis must be done to determine the real chemical structures of the volatiles obtained. However, bioassay works with the orientation of parasitoid towards the volatiles give clear indications

of preferable chemicals. In the modern investigations, bioassay has become more and more relevant and practical in deciding what chemicals are really involved.

When the description about chemical compounds presence was clear, an attempt for a similar system was also made. The extract of infested paddy containing volatiles were dropped into silica cellulose surface, the volatiles would escape into the air, leaving the surface depending on many physical aspects. This is a trial to make a similar system of wounded paddy stem surface from which the volatiles are released. The creation was based on information from analytical chemistry, which works in chemistry, and is brought into biological spheres.

However, the tritrophic interaction very much depends on spatial and temporal parameters (Raghava, Ravikumar, Hegde, & Kush, 2010). The physical properties of the volatiles as well as the releaser matrix play a significant role. When the place and time refer to humidity, temperature, air pressure, and presence of other chemicals associated with climate changes, then the mechanism of tritrophic interaction alters. Besides that, the intrinsic properties of plants as well as genetic variability amongst cultivars are also important to differentiate the processes. This indicates that complex biochemistry can possibly occur in each condition. With this complex chance and possibility, there are many factors also affecting in changing the release of volatiles. Some scientists have reported an analysis on the role of chemical diversity on defense mechanism, in this case the effect of pollution of ozone layer in changing the metabolism (Iriti & Faoro, 2009).

This proposed system of investigation was implemented under environmental considerations. When analytical chemistry can root the phenomenon from molecular level, then changes from molecular level can occur too. There are the steps of analytical chemistry, starting from sampling and pretreatment to data analysis. The result should be the new set of data for the next treatment. There was also some division of analysis across different levels of objects. The investigations started from chemistry investigation of infested paddy stems. The brown planthopper put the eggs inside the stem by making wounds in the cortex of the paddy stem. The wounds would induce certain biochemistry inside the paddy body and generate some chemicals, some are volatiles and readily spread all over the air. This is chemical cues for the parasitoid. In this case the parasitoid would scent the odor's source. The chemicals are extracted and investigated, and the chemical cues involved in the interaction can be known. To make a similar system, the researchers tried to create a similar situation using adsorbent, and tested them by olfactometer and the parasitoid in bioassay (Wonorahardjo et al., 2015). This is in fact a multidisciplinary approach in overcoming agricultural problems of pests and pesticides. This is also better from the environmentalist point of view by using biological pest controls, in the frame of green chemistry.

In the future, chemistry researches would deal more with biochemistry, to reveal the way some chemicals modulate the genetic expressions. Moreover, the genetic manipulation can also be done to make agriculture beneficial as well as for the environment. On the other hand, analytical chemistry works in different directions. The making of biocontrol pesticides which are environmentally friendly is to fulfill the need for short term necessity. In this case, the role of separation chemistry is also important as the underlying principles of many designs for the bio-pesticides and their carrier materials.

MATERIALS AND METHODS

Analytical chemistry methods can be divided into a few different steps, which were: 1) the sampling and pretreatment of infested paddy, 2) extraction process as well as pre-concentration, 3) separation and analysis by GC/MS, 4) biosilica-nanocellulose adsorbent making, 5) bioassay using parasitoid. All these stages were carried out in chemistry laboratories, at Universitas Negeri Malang, and entomology laboratory, Indonesian Sweetener and Fiber Crops Research Institute (Wonorahardjo, 2013).

The paddy was infested by brown plant hopper and grown under ambient circumstances. Two-month-old plants were chosen for the experiments. The stems of about 50 paddy plants were cut into 5 mm long pieces and immersed in solvent. The extraction was done using soxhlet apparatus and methanol as universal solvent was chosen to take the chemicals out of plant tissues. This polar solvent would be able to dissolve polar and semipolar compounds. The extractions were done in two rows: the healthy paddy as well as infested paddy. All the processes were completed after six circulations. The extracts then underwent pre-concentration using rotary evaporator. The final result was concentrated extract of healthy and infested paddy from the stem part of the plants.

Gas chromatography followed by GC-MS were done for both extracts to know the types of chemicals present in healthy or infested plants. The GC/MS runs were done using Shimadzu instruments (from Shimadzu Singapore). The carrier gas was Helium, the total speed was 20 mL/minute and column speed was 0.5 mL/minute. The initial temperature was 60°C for five minutes, then progressively increased to 300°C before settling it at around 40 minutes. Temperature of ion source was 250°C, interface temperature 305°C with cut time started at minute-3.75. Chromatograms obtained were called as per Library Wiley 8.

An important part of this research is the adsorbent/releaser making. In this study, the adsorbent was made of silica from rice husk ash via sol-gel processing and the method is already used in several occasions for separation purposes (Wonorahardjo et al., 2015). The silica was extracted from the rice husk ash by dissolving the ash into strong base before back gelation by strong acid according to Kalapathy, Proctor and Shultz (2000) with some modifications. In the gelation process, nanocellulose was incorporated into silica porous material. Nano cellulose itself was made by acid hydrolysis of *nata de coco* cellulose fiber for a certain period of time at warm temperature. These methods were registered with Directorate General of Intellectual Properties, Ministry of Law and Human Rights of Indonesia, as a simple patent at the end of 2015 (registration number P00201508272). The cellulose was incorporated into the silica network during gelling, together with sulfuric acid. In this part, nanocellulose was settled on surface of the silica granules.

The physical and chemical characterisation were done using scanning electron microscope (SEM) (PAN Analytical, Singapore) using magnification of up to 10000 times. Besides pictures, porosity and surface area was also done using BET method. Brunauer – Emmett – Teller theory is used to characterise the surface of materials which adsorbs physically gaseous inert molecules (Zhuravlev, 2000). Chemical characterisations were done using water and ash content as well as density and iodine number. Iodine number indicates chemical adsorption of some active molecules with the surface.

Bioassay is a useful method nowadays which includes biological parameters in big amounts. In this experiment, the assay was done using Y-tube olfactometer, in accordance with Lou, Ma and Cheng (2005) with modifications. The olfactometer contains Y-tube, cylinder glass for volatiles' source keeper, aerator, while glass tube contains active carbon as air filter, and cylinder glass with humidifier for keeping filtered air, flow meter light source (Figure 1). Bioassays were done at 8-11 am at $25 \pm 2°C$.
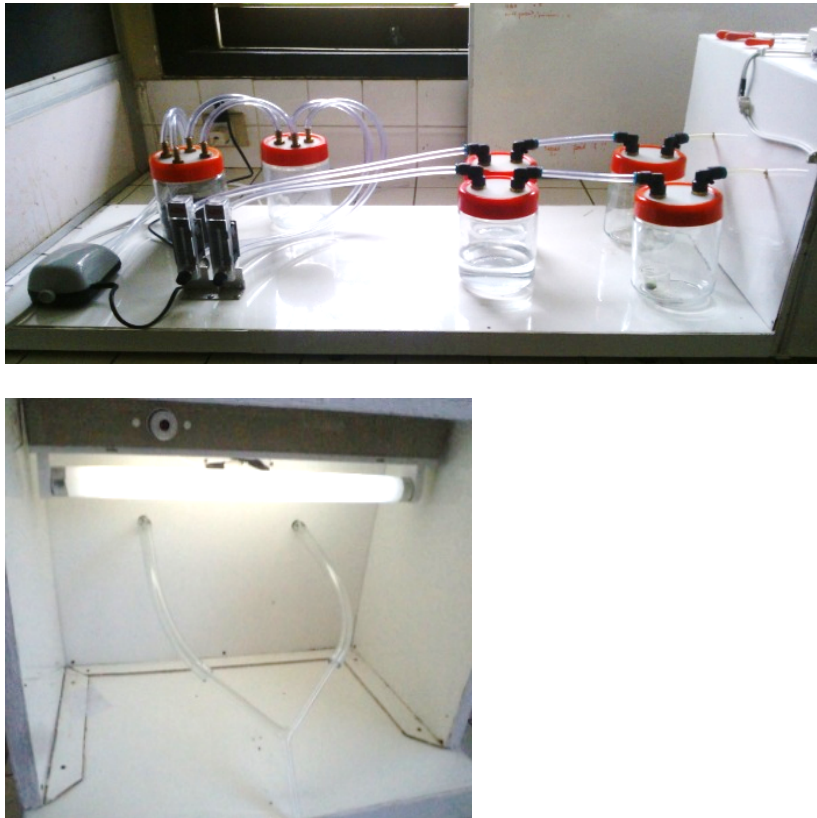
Figure 1. Olfactometer used in the bioassay step

Validation of the olfactometry was done by testing the parasitoids response to the natural volatiles of the infested plants. The uninfested and infested plants were placed on each odour source. The parasitoids were introduced singly into the Y-tube arm and the orientation of the parasitoid was observed. Parasitoids orientation toward the arm connected to the infected plant volatiles were categorised as having positive response (R+) and those to the uninfested plant were categorised as having negative response (R-). Those which did not show orientation to any arm were categorised as no response (NR). Bioassay of the adsorbent and extracted volatiles on the parasitoid orientation behaviour was done too to validate similar volatiles release from the plant tissue.

Field tests using the adsorbent / releaser and raw extract of infested paddy were done as well. The materials were put into plastic vials and set in wooden stakes and put in rice fields in certain distance. The wooden stake with glued wings to trap the parasitoid would give off the volatiles to attract parasitoid. The number of insects trapped in the glued wings were counted.

RESULTS AND DISCUSSION

The extraction process was the first analytical chemistry done for this project. The raw extract contained more or less the same chemicals as they presented in the plant tissues, before and after the infestation (Figure 2). Changes during process due to elevated temperature, excessive use of solvents in longer immersed time should be taken into account. However, the volatiles which consisted of longer chained hydrocarbon or other groups were relatively stable.

Figure 2. The extraction process and dxtracts obtained from infested and uninfested paddy plants

The different colours of the extracts indicated different contents. Some of them were pigment molecules which were damaged during the process of extraction. The soxhlet extraction is a perfect method in chemistry, however some unstable compounds would be different after extractions. The brown colour of the extract showed degradation compounds together, including the green chlorophyll pigments.

To investigate chemical presence in the extracts, the authors employed gas chromatography followed by mass spectrometry. The compounds were pushed through a semipolar column as the separation process took place. Each compound was carried by Helium carrier gas in certain retention time and would reach the end of the column within different retention time. Soon, after the compounds left the end of separation column, they were directed to the ionisation chamber of the mass spectrometer. Here, fragmentation occurred and the ionic fragments were separated by quadrupolar mass filter to be identified in the detector. From the pattern of fragmentation, the original compound can be matched to the database in Wiley 8.

The raw extracts were both rich of chemicals (Table 1), some were typical volatiles of paddys in other reports in this area (Lou et al., 2005). The similarity may be due to the paddy's types and species, but the difference may be due to different methods, ambient air, altitudes, climate as well as age of the plants (Nurindah, Wonorahardjo, Sunarto, & Sujak, 2017).

Table 1
Complete profiles of chemical compounds composition in infested and uninfested paddy stems (According to Wiley 8 Library)

No	Name of compounds	Uninfested % area	Infested % area
1	3-eicosyne	6,945958	3,991168
2	17-(1,5-dimethyl-hexyl)-10,13-dimethyl-1,7,8,9,10,11,12,13,14,15,16,17-dodecahydro-cyclopenta[a]phenanthren-4-one	6,465718	-
3	Stigmasta-5,22-dien-3-ol	5,447146	-
4	Stigmasta-5,24(28)-dien-3-ol, (3.beta.,24e)-	5,447146	-
5	Eicosanoic acid, methyl ester	5,239284	3,424039
6	9,12,15-octadecatrienoic acid, methyl ester, (z,z,z)-	4,152383	2,926125
7	2-hexadecen-1-ol, 3,7,11,15-tetramethyl-, [r-[r*,r*-(e)]]-	3,679248	3,290053
8	Hexadecanoic acid, 2,3-dihydroxypropyl ester	3,673684	-
9	9,12-octadecadienoic acid, methyl ester, (e,e)-	3,608907	2,352646
10	Eicosane, 7-hexyl-	3,544913	-
11	Tricosane	3,425545	2,708808
12	Pentatriacontane	3,425545	-
13	Tricosane	3,37482	-
14	Pentacosane	3,37482	4,224462
15	Methyl 5-oxo-2-pyrrolidinecarboxylate #	3,370902	-
16	Hexadecanoic acid	3,24187	-
17	Hexatriacontane	2,989082	4,935695
18	3-hydroxypregn-5-en-20-one	2,98135	-
19	9,12-octadecadienoic acid (z,z)-, 2-hydroxy-1-(hydroxymethyl) ethyl ester	2,948544	-
20	3-eicosyne	2,573149	1,525004
21	Octadecane	2,498707	-
22	Hentriacontane	2,354708	-
23	Octacosane	1,850176	4,877568
24	Pentatriacontane	1,665456	-
25	1-isopropoxy-3,3,3-trimethyl-1-[(trimethylsilyl)oxy]disiloxanyl tris(trimethylsilyl) orthosilicate #	1,503278	-
26	3-furylmethanol	1,469662	-
27	Docosane	1,275486	1,662665
28	Ethanone, 1-(2-hydroxy-5-methylphenyl)-	0,914431	-
29	1-decene, 8-methyl-	0,64665	-
30	Benzaldehyde, 4-methyl-	0,613582	-

Table 1 *(continue)*

No	Name of compounds	Uninfested	Infested
		% area	% area
31	(2e)-3,7,11,15-tetramethyl-2-hexadecene #	0,603291	-
32	Phenol, 2,6-dimethoxy-	0,229411	-
33	Phenol, 4-(3-hydroxy-1-propenyl)-2-methoxy-	0,226538	-
34	Heptadecane, 3-methyl-	-	4,877568
35	Dotriacontane	-	4,377178
36	Celidoniol, deoxy-	-	4,020032
37	Tetracosane, 11-decyl-	-	3,497129
38	Tetratetracontane	-	3,416504
39	Nonacosanol	-	3,027218
40	9,12,15-octadecatrienoic acid, methyl ester, (z,z,z)-	-	2,926125
41	Heptadecane	-	2,708808
42	9,12-octadecadienoic acid, methyl ester, (e,e)-	-	2,352646
43	Ergost-5-en-3-ol	-	2,271206
44	Ethanol, 2-(dimethylamino)-	-	1,570983
45	2-methoxy-4-vinylphenol	-	0,809865

It is also interesting to see that some chemicals (in red) are present in both healthy and infested plants. In this scheme of thinking, there are natural chemicals which play a role within the tritrophic interactions as equilibrium makers. The blend composition cannot be separated from each other in giving the power as bioattractant, and the biochemistry can go though it in deeper analysis (Jang et al., 2014; Xu et al., 2003). There must be several interesting topics as a sequence of these analytical steps.

In imitating the natural wound-induced volatiles release, hard cellulose materials having the chemicals of the surface were considered. From some investigations the cellulose surface combined with silica can be proven as the slow releaser. The adsorbent made for the bio-attractant gave white porous powder with silica and cellulose content bound together physically (Figure 3). This cellulose on surface shifted the adsorptivity of the surfaces (Wonorahardjo, Wijaya, & Suharti, 2016). This also changed the separation properties of the original adsorbents. By the presence of cellulose, the volatiles were released step by step depending on the temperature, air pressure, as well as surface interactions. The chemical reactions can be written as below:

$$SiO_2(s) + 2\ NaOH\ (aq) \rightarrow Na_2SiO_3(aq) + H_2O\ (l)$$

$$Na_2SiO_3(aq) + H_2SO_4(aq) \rightarrow SiO_2(s) + Na_2SO_4(aq) + H_2O\ (l)$$
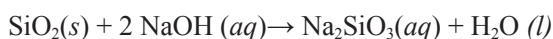
Figure 3. Adsorbent making for volatile releaser

Scanning electron microscopy pictures of the material can be seen in Figure 4. The texture of the surface can be seen clearly, that is, the porosity of the matrix plays a role in broadening the surface area. Surface interactions in adsorbing and releasing volatile compounds would regulate the amount of volatiles in the air before regulating the tritrophic interactions.
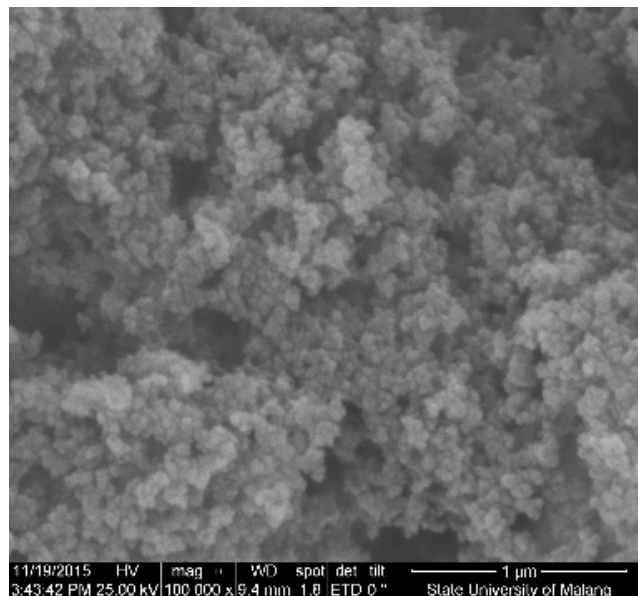
Figure 4. SEM picture of silica-cellulose surface

Adsorbent material made of silica and cellulose can bind the organic content of the extracts in a proper way. Having a good surface area (13.514 m²/g), the interaction between molecules with the surface can be considered (Table 2). Even though the BET result indicated physical interaction dominant on the surface, the high iodine adsorption numbers also gave an indication of more chemical interaction possible on two types of homogeneous surface (nonpolar and semipolar from cellulose and silica). This is how the act as adsorbent as well as releaser of both the polar and nonpolar compounds is possible (Table 1). Higher water content was the moisturiser of the surface and the system was almost similar to the original wounded-plant stems.

Table 2
BET characterisation, density, water content, ash content, and iodine adsorption number

Characterisation	
Surface area	13.514 m²/g
Density	3,657 g/mL
Water content	15,400%
Ash content	22,85%
Iodine adsorption number	4,822%

The analytical concepts of equilibrium on surface can be the underlying principles in surface interactions. In this case the solid-gas equilibrium on the interface occurred as far as the surface was covered homogenously by the blended compounds. Moreover, the porous structure they have suggested the in tortuosity played a role in surface interaction. The partly filled porous medium can undergo fast diffusion as studied thoroughly using Nuclear Magnetic Resonance decades before (Ardelean, Mattea, Farrher, Wonorahardjo, & Kimmich, 2003; Mattea, Kimmich, Ardelean, Wonorahardjo, & Farrher, 2004). Strange kinetics on such surface was already studied as the consequences of porous system (Kimmich, 2002). However, in application, discussion about chemistry scale physics was not considered widely. Attention about separation power of some surfaces was given in greater deal.

The biological assessment or bioassay provided useful information for the feasibility of analytical chemistry to get involved in biological research. The olfactometer served as the nature of paddy in field. The positive response towards volatiles from infested paddy was clear. For six experiments, the number of parasitoid which gave positive response was high compared to the negative ones. Only a few numbers showed no response. Since the mixture of compounds was used, blended components cannot be treated separately. The parasitoids involved gave indication for biological activities of the chemicals. This involvement was done since most of the biological systems work with more than one single compound as a blended mixture together. In this case, the knowledge of quantitative effect of each compound is lacking.

Field test using adsorbent/releaser materials in wooden stakes in rice field offered interesting results too (Figure 5). There were also other insects trapped in the glued wings besides the intended parasitoid. The number of parasitoid trapped within five days in a row can be seen in Table 3. In windy situations, more insects can be found. However, more descent experiments must be done to see the results in bioassay for real conditions.

(a)   (b)
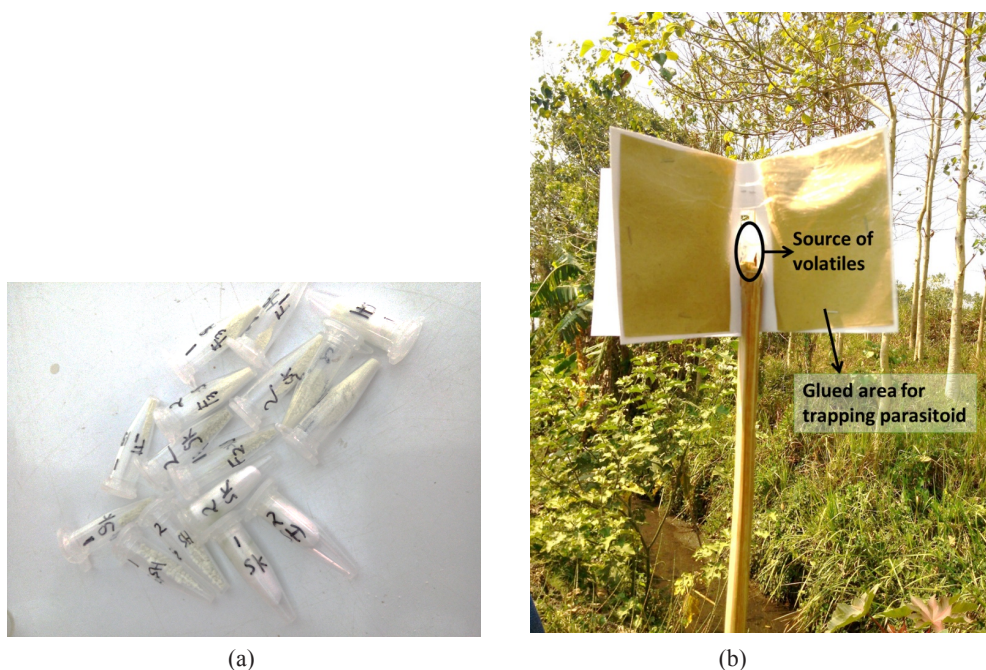
Figure 5. (a) Plastic vials containing adsorbent/released volatiles; (b) Field test using the source of volatiles

Table 3
Parasitoid entrapment in field test

Day	PSe	JPSe A	JPSa A
I	2	9	10
	0	4	8
II	4	5	7
	2	1	6
III	0	8	5
	0	0	1
IV	3	1	1
	5	0	1
V	1	0	2
	2	2	3

PSe : extract without adsorbent

JPSe A : extract adsorbed in silica cellulose

JPSa A : extract adsorbed in silica cellulose (duplo)

Within five days, the number of molecules given off from the surface of silica-cellulose would be decreased. In warmer temperature, the age of this bioattractant would be shorter. There must be more treatment in the surface to be applied so that the sequence of evaporating molecules can be calculated. In this case, the age of the attractant in the field, as far as the rain does not come, can be determined. Moreover, the expiring date of the formula must be given as well.

There is a lot to analyse in every step of analytical chemistry. The sensitivity, selectivity, and effectiveness of the bio-attractant must be followed too. When it goes deeper into plant biochemistry, more new insights can become key knowledge for the development of friendlier biological controls.

CONCLUSION

There are a lot of lessons learned from the natural behaviour of living things on earth. The mechanism of defense is one of the natural phenomenon and can be followed at least by analytical chemistry. The steps to analyse and imitate tritrophic interaction were done using paddy-brown planthopper-egg parasitoid system, from extraction and analysis followed by making adsorbent/releaser of volatiles, bioassay as well as field study. More analysis on bio-attractant making should be done in future, but the complex biochemical changes in between the steps are the objects of much follow-up research. Similarly, the study of surface interaction and material making can be crucial for agricultural purposes. The type of industrial bio-attractant can be considered and for that, analytical parameters should be set and optimised before they are applied.

ACKNOWLEDGMENTS

REFERENCES

Agrawal, A. A. (2000). Mechanisms, ecological consequences and agricultural implications of tri-trophic interactions. *Current Opinion in Plant Biology, 3*(4), 329-335.

Ardelean, I., Mattea, C., Farrher, G., Wonorahardjo, S., & Kimmich, R. (2003). Nuclear magnetic resonance study of the vapor phase contribution to diffusion in nanoporous glasses partially filled with water and cyclohexane. *The Journal of Chemical Physics, 119*(19), 10358-10362.

Bi, H. H., Zeng, R. S., Su, L. M., Min, A., & Luo, S. M., (2007). Rice allelopathy induced by methyl jasmonate and methyl salicylate. *Journal of Chemical Ecology, 3*3(5), 1089-1103.

Dyer, L. A. (2005) *Tropical tritrophic interactions: Nasty hosts and ubiquitous cascades in tropical forest community ecology.* Oxford: Blackwell Science.

Iriti, M., & Faoro, F. (2009), Chemical diversity and defence metabolism: How plants cope with pathogens and ozone pollution. *International Journal of Molecular Sciences, 10*(8), 3371-3399.

Jang, G., Shim, J. S., Jung, C., Song, J. T., Lee, H. Y., Chung, P. J., ... & Do Choi, Y. (2014). Volatile methyl jasmonate is a transmissible form of jasmonate and its biosynthesis is involved in systemic jasmonate response in wounding. *Plant Biotechnology Reports, 8*(6), 409-419.

Kalapathy, U., Proctor, A., & Shultz, J. (2000). A simple method for production of pure silica from rice hull as. *Bioresource Technology, 73*(3), 257-262.

Kimmich, R., (2002). Strange kinetics, porous media, and NMR. *Chemical Physics, 284*(1-2), 253–285.

Lou, Y. G., Ma, B. O., & Cheng, J. A. (2005). Attraction of the parasitoid anagrus nilaparvatae to rice volatiles induced by the rice brown planthopper nilaparvata lugens. *Journal of Chemical Ecology, 31*(10), 2357-2372.

Mattea, C., Kimmich, R., Ardelean, I., Wonorahardjo, S., & Farrher, G. (2004). Molecular exchange dynamics in partially filled microscale and nanoscale pores of silica glasses studied by field-cycling nuclear magnetic resonance relaxometry. *The Journal of Chemical Physics, 121*(21), 10648-10656.

Nurindah, N., Wonorahardjo, S., Sunarto, D. A., & Sujak, S. (2017). Chemical cues in tritropic interactions on biocontrol of insect pest. *The Journal of Pure and Applied Chemistry Research, 6*(1) 49-56.

Raghava, T., Ravikumar, P., Hegde, R., & Kush, A. (2010) Spatial and temporal volatile organic compound response of select tomato cultivars to herbivory and mechanical injury. *Plant Science, 179*(5), 520-526.

Wonorahardjo, S., (2013). *Metode-metode pemisahan kimia, sebuah pengantar* [Separation chemistry methods, an introduction]. Jakarta: Indeks Akademia.

Wonorahardjo, S., Nurindah, N., Sunarto, D. A., Sujak, S., & Zakia, N. (2015). Analisis senyawa volatil dari ekstrak tanaman yang berpotensi sebagai atraktan parasitoid telur wereng batang coklat, anagrus nilaparvatae (Pang et Wang) (Hymenoptera: Mymaridae) [Analysis of volatile compounds from plant extracts potentially as chickens parasitoids of brown planthopper aphis, anagrus nilaparvatae (Pang et Wang) (Hymenoptera: Mymaridae)]. *Jurnal Entomologi Indonesia, 12*(1), 48-57.

Wonorahardjo, S., Wijaya, A. R., & Suharti, S. (2016) Surface behavior of rhodamin and tartrazine on silica-cellulose sol-gel surfaces by thin layer elution. *The Journal of Pure and Applied Chemistry Research, 5*(2), 48-54.

Xu, T., Zhou, Q., Chen, W., Zhang, G., He, G., Gu, D., & Zhang, W. (2003). Involvement of jasmonate-signaling pathway in the herbivore-induced rice plant defense. *Chinese Science Bulletin, 48*(18), 1982-1987.

Yuan, J. S., Himanen, S. J., Holopainen, J. K., Chen, F., & Stewart, C. N. (2009). Smelling global climate change: mitigation of function for plant volatile organic compounds. *Trends in Ecology and Evolution, 24*(6), 323-331.

Zhuravlev, L. T. (2000). The surface chemistry of amorphous silica. Zhuravlev model. *Colloids and surfaces: A Physicochemical and Engineering Aspects, 173*(1-3), 31-38.

SCIENCE & TECHNOLOGY

OPH-LB: Optimal Physical Host for Load Balancing in Cloud Environment

Sakshi Chhabra* and Ashutosh Kumar Singh

Department of Computer Applications, National Institute of Technology, Kurukshetra, Haryana, India

ABSTRACT

Cloud computing has set a trend on a worldwide stage along with the rapid growth to enhance global technology standard and market scale in recent years. For the cloud users, load balancing in data center networks initiates the necessity of reducing the downtime for migrating overloaded virtual machines. To achieve better during-task deployment, optimal physical host must be selected efficiently. Nowadays, cloud customers are facing security risks in the context of load balancing of Virtual Machines (VM) which is infrequently addressed. This research addresses this pertinent issue and provides a different perspective of studying ways to develop VM deployment strategy by reducing the probability of VM co-tenancy with their targets. This will in turn make it difficult for attackers to evaluate the strategy. A model called Optimal Physical Host for Load Balancing (OPH-LB) is proposed to find the probability with probabilistic estimation in the form of its computing capability and performance in secure multi-tenant cloud. The proposed solution is evaluated via Cloudsim 3.0.3 and compared with two existing well-known algorithms. The reported results indicate that OPH-LB outperforms in improving the makespan, throughput, performance and reduces the failure number of task deployment. The results show that OPH-LB can effectively reduce the risks and security score and upgrades the utilisation of resources, with an improvement of 42.13% in all types of analyses for the experimental data.

Keywords: Cloud computing, deployment, optimality, performance, probability, security

INTRODUCTION

The paradigm of cloud computing reshapes all prospective users and is emerging as the fastest technology worldwide. It is an eventual and promising way of managing and boosting the utilisation of resources and delivering various computing, IT services (Zissis & Lekkas, 2012). In recent years, the number of network users has been growing linearly, but traffic has been increasing exponentially at

cloud data centers (Diaz, Martin, & Rubio, 2016). For this reason, load balancers are essential to balance the traffic and that is the reason why overloaded servers are able to process their backlog successfully. Virtual machine migrations and balancing the upcoming workload based on performance requirements are made possible by virtualisation. This is particularly useful when the workload is unpredictable or varies significantly. This XaaS (X-as-a service) model in cloud as infrastructure provides service which is parallel to hardware resources for its clients, such as Amazon EC2 and Amazon S3 (Ang, Por, & Liew, 2017). Because of the inadequate hardware possessions in a cloud system, it has become a big concern as to how to allocate the resources securely and choose the best machine for task deployment effectively. This will result in a strategy for load-balancing and obtaining reliable performance (Kavousi-Fard, Niknam, Taherpoor, & Abbasi, 2014). Many companies propose new techniques for fast and efficient application deliveries for the deployment of load, as KEMP technologies release vRealize Plugins for fast balancer deployments. VMs is also a commonly used resource in the cloud computing environment and acts like an isolated computing unit.

For cloud users, it facilitates the sharing of resources, on-demand resource scaling with high flexibility, controllability and predictable performance. However, apart from all these benefits, it also brings a new security threat as a number of concerns are emerging regarding the issue of multi-tenancy attacks where multiple tenants are residing on the same server. It is evident that if attackers and clients reside on the same server, there are more chances of co-resident attacks (Han, Chan, Alpcan, & Leckie, 2017). At the time of balancing the load, security issues are rarely addressed (Chhabra & Singh, 2018). So, a model for VM's security during load balancing and deployment of job requests is proposed. When virtual machines migrate, common tenants are concerned about VM placements from one host to strange hosts. Hence in this study, the researchers tried to analyse possible threats during load balancing. The general view of the cloud data centers is demonstrated in Figure 1. Here, cloud consumers send the workload requests for the deployment to cloud service providers (CSPs). The resource manager subsequently handles the heavy workload and monitors the whole scheduling of upcoming load in the cloud management portal. Then, CSPs divide the resources into fully managed resource pool chunks which deliver excellent flexibility and controls. To make changes in the resource pool itself is very simple and allocations can be scaled up or down as per requirement.

In facilitating effective usage of computing resources and reduced waiting time, two aspects, which are security and load balancing need to be considered (Lin, Chin, & Deng, 2014). In this paper, the safety of a VM allocation policy in its ability to defend against multi-tenant cloud attacks is measured. These metrics are modeled under basic resource provisioning VM allocation policy such as time-shared, and extensive experiments on the widely used simulation platform Cloudsim to validate the model. Basically, this new proposed secure load balancer policy OPH-LB significantly decreases the co-tenancy of virtual machines on the same physical host, but also satisfies the constraints in workload balance and security (Calheiros, Ranjan, Beloglazov, De Rose, & Buyya, 2011).
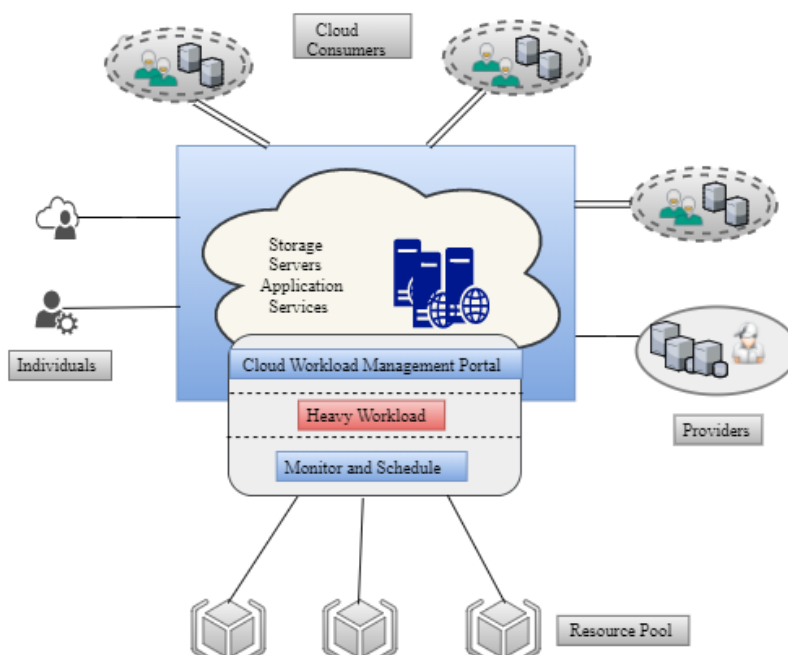
Figure 1. General view of cloud data center networks

Significance of the Research

This idea provides a generic formulation which distributes the load effectively to the respective servers in cloud data center networks. This can only be achieved by considering the current status of requested tasks for the cloud data centers intelligently, which helps to improve the efficiency of computing the resources and managing the incoming requests among the physical machines securely. This secure load balancer not only significantly decreases the co-tenancy of virtual machines on the same physical host but also satisfies the workload balance.

The contributions are threefold:

- To consider the dynamic simulated scenario which helps structured and flexible evaluation of the proposed model during secure load balancing. A reliable VMs for task deployment is able to diminish the risks from malicious hypervisor and ensures the original performance of the load balance.

- To guarantee the scalability of algorithm, no modifications are required in the guest OS, hypervisors or hardware platforms. It basically modifies the structure of the model to a very less extent by converting a simple load balancer to secure load balancer.

- OPH-LB achieves relatively accurate estimation with less communication overheads of the upcoming load and selects optimal physical host for processing the tasks.

The subsequent sections of this paper is organised as follows: In section 2, a brief review of the related work which achieved secure load balancing in cloud datacenters is offered. Section 3 introduces the Optimal Physical Host for Load Balancing (OPH-LB). The performance

evaluation which was obtained from the practical implementations is discussed in section 4. Finally, section 5 concludes the paper and gives future direction of the work.

RELATED WORK

Load balancing is the mechanism which decreases the possibility of VMs to be overloaded or underloaded. This leads to improvement in the concurrent user capacity and overall reliability of applications and also helps to achieve the best response time and good utilisation of resources. The methods of load balancing can be broadly classified into two categories: static and dynamic. In the static type of balancing, the prior knowledge of the system is needed and cannot change the requirement of resources at the run time because need is fixed before the programme execution. In the case of dynamic algorithm, it is based on the current usage and gives the facility of re-mapping according to their respective task requirement during run time. Dynamic algorithm deals with real-time network and finds the closest server in the whole system and prefers the deployment (Cho, Tsai, Tsai, & Yang, 2015). Here the present state of the system is used to make decisions to handle the load with the benefits of resource utilisation (Li, Qian, Lu, & Wu, 2013). With the aim of improving the performance, redistribution policy is applied in which they transfer the tasks from heavily loaded processors to lightly loaded processors. Literature search on previous work on optimal task deployment and its possible countermeasures has yielded various results. Zhao et al. (2016) proposed a model which focused on the selection problem to place the virtual machines optimally for deploying the requested tasks to achieve the immediate load balancing effect. The authors applied clustering approach with Bayes theorem to choose the optimal set of physical host. It was found that this approach improves the throughput, reduces the number of failures and optimises the load balancing effect. Domanal and Reddy (2014) implemented the method which ensures the usage of resources in an intelligent manner, so efficient utilisation of resources can be done. They also compared with previous Active-VM algorithms. Deng, Wu, Shen and He (2016) presented the model which combines both static and dynamic provisioning and makes the online power management system which helps to save power and reduce the operating cost of carbon emmissions of the data centers. Its algorithm is to adapt the usage of green data centers powered by renewable energy called EcoPower which performs better load balancing and eco-aware power management simultaneously (Deng, Lu, Lai, Luan, & Liang, 2016). Its main objective is to diminish the average time and cost without any compromise on quality. So, this system's performance proves that this model achieves a good balance between power savings, cost and quality. The cost can also be reduced by 20% and can use solar or wind complementary strategies for further purposes. An optimisation of virtual machine placement for energy efficiency algorithm was also proposed by Li et al. (2013) based on evolutionary population initialisation strategy which helps to minimise energy consumption and maximise load balance. This model is evaluated with many-objective optimisation problems (MaOPs) with energy consumption reduced to 23.01 kWh with a percentage of 0.00029 and minimises the 770 number of VM migrations.

The cross VM side channel attacks introduced by Martin (2010) proposed the fine-grained information extraction between VMs. They mapped the internal cloud infrastructure, where

a particular target VM and attacker VM was likely to reside in one physical machine. This approach is mainly focused on mitigating the side channel risks and employs the blinding techniques to minimise the information that can be leaked. An approach proposed by Zhang, Juels, Reiter and Ristenpart (2012) basically deals with extraction of fine-grained information from a victim VM running on the same server. These types of attacks happen on a symmetric micro processing virtualised system. Their algorithm is able to overcome a few problems: filter out numerous sources of noise, core migrations and extract the victim's key. There is a technique which helps to mitigate the arbitrary cloud side channel attacks which is recommended by Moon, Sekar and Reiter (2015). They presented "Nomad", a system that suggests vector-agnostic defense against known and future side-channels. It captures information leakage model by channels and required migration heuristics between VMs in shared cloud deployments. Sun, Shen, Li and Wu (2016) presented the model for analysis related to security threats and generated an information leakage model to secure the load balancing risks. This SeLance technique estimates and predicts the information leakage in the whole process during VM migrations and VM placement. Some researchers only focus on a particular threat suggested by using VM allocation policies to defend against co-resident attacks in cloud computing (Han et al., 2017). They basically follow three metrics: efficiency, coverage and VMmin. Their work satisfies these objectives such as security, workload balance and power consumption. Duan and Yang (2017) implemented the method which generates multi-tenancy oriented private clouds and allow multiple VMs to communicate with others under physical hosts. They have achieved global load balancing on the underlying physical networks. To attain efficient resource provisioning by optimal workload allocation is also studied in light of the max-min algorithm by Cao, Li and Stojmenovic (2014). Other solutions for secured load balancing with many techniques are by Papagianni et al. (2013), Ramezani, Lu, and Hussain (2014), and Zhao, Hu, Ding, Xu, and Hu (2014).

OPH-LB

Proposed Scenario and Assumptions

In IaaS cloud data centers to improve utilisation of resources, CSPs places the VMs which relates to different tenants in one physical host. The respective VMs share the host's resources and cloud service providers make sure the isolation between each VMs. In this multi-tenant cloud, there are chances of coinciding of VMs attacks. So, the objective of this study is to choose the optimal physical host for deploying the tasks in this secure cloud environment. When clients submit their job requests in resource pool of the cloud data centers, it generally chooses the physical host randomly for deploying the tasks as illustrated in Figure 2. However, it can become optimal if we can decide the most favorable machine for a particular task and will create a better load balancing effect. This idea will surely improve resource utilisation and provide high throughput. At the time of allocating tasks, firstly we need to check if the amount of resource requested by a task is greater than the available resources, then the physical host cannot deploy the task. When available memory is greater or close to the requested ones, then it can only deploy the tasks effectively. Whether it is possible for several physical machines to deploy the same task, we still need the physical host which is optimal among all

physical machines. As we know, parallel computing systems have many processors which run simultaneously and are scattered in multiple locations. For reducing the communication overheads between processors and to obtain the best scheduling performance, a hybrid system topology is pursued. At the time of finding a physical host which is the most favorable, we don't need to go and check for each and every processor because there is an index table named as DUIDX which preserves all information and estimations, and helps the load manager to maintain the records and reflect about any changes in the deployment time. When new jobs are coming in the system, then the load manager chooses the most optimal host for deployment so that communication overheads are reduced substantially.
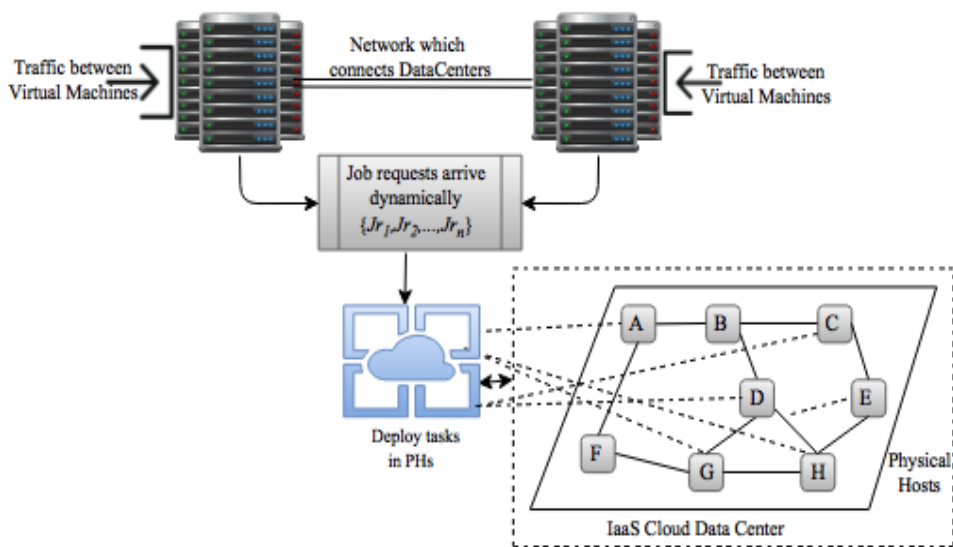
Figure 2. Requesting jobs into physical hosts for deployment in IaaS

This strategy is effective in high throughput with favourable scalability and less communication overheads. In multitenant cloud, attackers are always trying to achieve economic gain by utilising virtualisation and allow resource sharing. It occurs when two or more customers are using the same physical machine's services given by cloud service providers at one time. These risks come when both the attacker and clients are in the same cloud and are sharing the same server. In order for multi-tenancy, both virtualisation and resource sharing must be allowed by cloud service providers as shown in Eq. (1).

MultiTenancy = Virtualisation + Resource Sharing (1)

Here, resources are being shared between attackers and customers. Although the difficulty of deploying co-tenant VMs is actually reduced by load balancing, multi-tenancy threat occurs when the attacker and the target victims are in the same cloud and are sharing the same physical host. Given this better and secure load balancing scenario, the new strategy should satisfy these three main objectives:

Workload Balance - The importance of balancing the workload is twofold: For cloud providers, distributing VMs over multiple processors have to improve the whole parallel performance in the cloud. In the research policy, all the requesting tasks are not allocated together on the same server at one time. These providers adjust the workload periodically and dynamically so that all running tasks of different locations will complete their execution on time, so that idle time is minimised and resource utilisation can be maximised.

$$WB : \max \frac{1}{N} \sum_{C_n \in \pounds} [| \{s \mid s \in \mathrm{SR}, \text{€}_{\mathrm{C,Jr,PH}} = 1, \forall PH \in \wp\} |] \tag{2}$$

Computing Efficiency - How to effectively use the computing resources is a critical issue for cloud providers. There are many different techniques that have been widely discussed in various research papers (Zhang, Chong, Pezeshki, Moran, & Howard, 2017). However, there is yet to be a 100% foolproof solution. Therefore, here, the most straightforward approach is considered to pick out the most optimal hosts set and minimise the extra running servers.

$$CE : \min[| \{s \mid s \in \mathrm{SR}, \exists C_n \in \pounds, Jr \in \yen, \text{€}_{\mathrm{C,Jr,PH}} = 1\} |] \tag{3}$$

Secure Physical Host - The ability to optimise and secure the cloud service performance is of great importance and selects the most optimal physical host to reduce unnecessary computing resources. Based on minimum resource utilisation and less communication overhead between hosts in secure environments, a way to dynamically produce the best load balancing with low computation complexity is allocated as legal and attackers can't deploy the tasks to the same physical host. Based on such idea, he average number of clients per physical host is minimised.

$$SPH : \min \frac{1}{\wp} \sum_{s \in \mathrm{SR}} [| \{c \mid c \in \pounds, \text{€}_{\mathrm{C,Jr,PH}} = 1, \forall PH \in \wp\} |] \tag{4}$$

These targets have determined and accomplished in the analysis and simulation output of this paper. In addition, the following assumptions are made:

➢ The available CP_{avl}, MeM_{avl} and requested CP_{rqs}, MeM_{rqs} capacity of the physical hosts should be known by the load balancer manager and stored in DUIDX index table. However, when a new request is being processed, only the favorable physical hosts with adequate resources left are considered. In other words, designing an algorithm to find the optimal host with less communication overhead is the focus of this study.

➢ This multi-modal optimisation considers two objectives: CPU utilisation and memory. It is done for every incoming task requests when it arrives in the cloud data center, so that only current system capacity state and task requests are taken into consideration.

➢ A dynamic environment for implementing this approach is considered in order to reduce unnecessary computation complexity and assure optimal balancing effect.

➢ Cloud Service Providers (CSP_i) don't have foregoing knowledge of the attacker's capability and requests are considered identically. VM live migration is taken into consideration because when attack launches, intrusion detection system notices automatically. It is known when the current working VM is in secure or insecure state.

METHODS

Architecture

The proposed problem of task deployment can be formulated with the following scenario: In a cloud paradigm of N clients $£= \{C_1, C_2, ..., C_n\}$, R job requests $¥ = \{Jr_1, Jr_2, ..., Jr_r\}$ requesting to p physical hosts $\square = \{PH_1, PH_2, ..., PH_p\}$. A mapping $\epsilon: £ \times ¥ \rightarrow \square$ allocates each physical host from each user with specific job requests, $\epsilon_{£ \times ¥ \times \square} = \{\epsilon_{C,Jr,PH} | \epsilon_{C,Jr,PH} = 1$ if job request Jr of client C is allocated to optimal physical host$\}$. This obtained solution vector mapping ϵ is the deployment strategy and used to conclude which task will be deployed into which physical host (PH_p), and should be in secure state. The nomenclature of these notations is described in Table 1. These hosts are assumed to be heterogeneous and implemented in dynamic environment. When cloud data centers receive the request for deploying the tasks, then the OPH-LB problem is formulated in a stochastic framework to find out the final deployment strategy by utilising its algorithm mechanism. This problem for load balancing can be solved through optimising task deployment problem in every Δt time from a long perspective. The tuples are defined in $\square = \{Jr, PH, C_n, CP_{avl}, MeM_{avl}, CP_{rqs}, MeM_{rqs}, E\}$. These Jr_r defines the job requests which comes for deploying their tasks from $\{i = 1, 2, ..., r\}$. PH illustrates the set of available physical hosts $PH(p, tm) = \{PH_1, PH_2, ..., PH_p\}$, where tm represents the starting time for deploying the tasks. These CP_{avl}, MeM_{avl} parameters discuss the current available CPU and memory resource amount of the p physical host in the set $CP_{avl}^1, CP_{avl}^2, ..., CP_{avl}^n$ and $MeM_{avl}^1, MeM_{avl}^2, ..., MeM_{avl}^n$. CP_{req} and MeM_{rqs} is the requested resource amount by clients for deploying the r tasks in Jr.

Table 1
Description of notations to be used

Notation	Description
C_n	Number of Clients
£	Set Set of clients $\{1,2,...,n\}$
¥	Set of Job Requests $\{1,2,...,r\}$
€	Specifies mapping between hosts, job requests and clients
□	Octatuple
PH_p	Number of Physical hosts
CP_{avl}, MeM_{avl}	Available resource amount of CPU and Memory
CP_{rqs}, MeM_{rqs}	Requested resource amount of CPU and Memory
SR	Set of servers $\{1,2,...,s\}$
$VM_{i(t)}$	VM i which is located by tenant t
$VM_{i'(t')}$	VM i' which is located by tenant t' malicious one
WB	Workload Balance
SPH	Secure Physical Host
CE	Computing Efficiency
RD	Random Deployment

The goal is to protect the multi-tenant cloud in this way so that co-resident attacks never harm any physical host. A binary variable $\perp_p^i = 1$ indicates VM_i will be placed on ph_p and 0 otherwise. The load placement matrix is defined as T. Let $T_{i,j}$ be an element of matrix T, then $T_{i,j}$ defines the load placement between VM_i and VM_j. It should be noted that a feasible VM placement decision should satisfy the following resource constraints:

$$\sum_{i=1}^{\wp} \perp_p^i \, CP_{rq} \leq CP_{avl} \, i \in \{1,2,...,n\} \tag{5}$$

$$\sum_{i=1}^{\wp} \perp_p^i \, MeM_{rq} \leq MeM_{avl} \, i \in \{1,2,...,n\} \tag{6}$$

Eq. (5) and (6) ensures that the total required consumption of processors and memory resource amount should not exceed its total capacity. For avoiding the overflow condition of the node servers, we have the following load constraint:

$$\frac{1}{2} \sum_{i=1}^{V} \sum_{p=1, p \neq i}^{\wp} T_{i,j} \, \perp_p^i \perp_j^p \leq B_s \, \forall s_t, s_q \in \delta, t \neq q \tag{7}$$

where B_s denotes the bandwidth of servers st. The coefficient (1/2) defines the VM placement pair of respective servers, the load is calculated two times. It guarantees that each task is allocated in one VM as respective physical host with specific requests. This $VM_{i(t)}$ is defined

as VM i which is located by tenant t and $VM_{i'(t')}$ as $VM_{i'}$ located by tenant t' which is called malicious one. This $Co\text{-}Tenant_{t,i,t',i}(\xi t)$ represents a Boolean value which defines whether any $VM_{i(t)}$ and $VM_{i'(t')}$ are Co-Tenant at time ξt. As the main goal is to reduce the leakage introduced by Chhabra and Singh (2016) or insecure states which occur during this proposed process, a secure load balancing policy is designed where resource manager allocates the tasks effectively so that optimal VM is selected and placed successfully. For each strategy, when the upcoming load comes for deployment, the secure or insecure states are calculated and predicted for every decision. When clients (Cl_i) send their workload job requests to the resource manager, OPH-LB model identifies the reliable or unreliable states. It maintains a collection of all possible VM(VM_i) as in Eq. (8) to be selected and placement at first. Then, it calculates the secure or insecure states for all feasible migration paths. As the system has less insecure states, it should have more security advantages during the deployment of requested jobs. In case of safe states, the qualified hosts as required and available amount of resources are observed. The safe VMs are returned to the resource manager who will make the final decision, in appraisal of both security and load balancing, as illustrated in Figure 3.
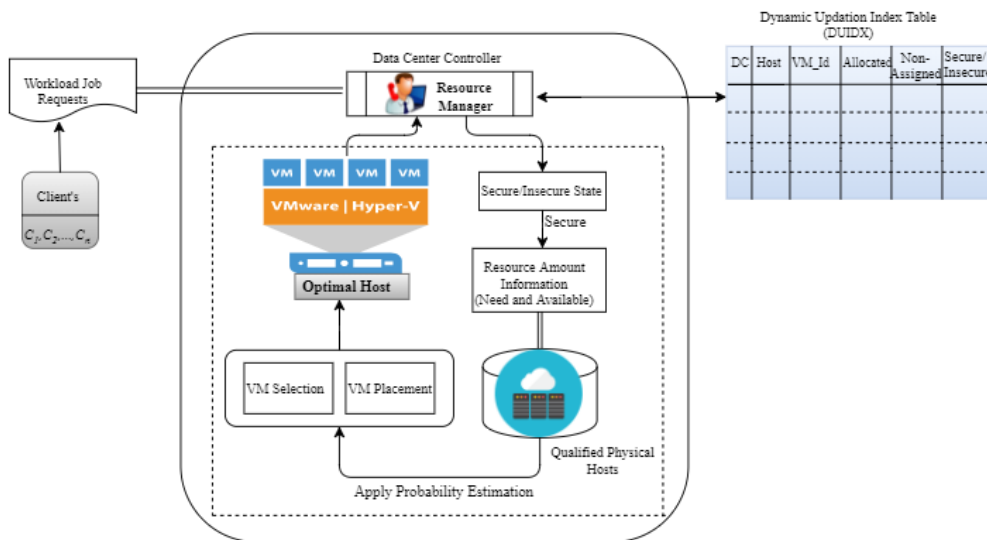
Figure 3. OPH-LB

Secure and insecure states were introduced, from which this model is able to show whether upcoming requests are deployed securely or not. For safe states, some conceptions have been initiated accordingly. Firstly, the reliability of particular tenants is estimated with the help of Co-Tenancy between those clients in historical time ξt. THR_S time to find the malicious VMs from all the co-tenant VMs is considered when VM P and VM Q have been co-tenant in the past ξt times and resides parallel for more than its threshold time THR_S. When the intrusion detection system doesn't ring an alarm, then we can generally think that both VMs are favourable to each other and there is no maliciousness between them as shown in Eq. (8) (9); these machines are ready for the deployment of tasks. For the trusted hosts, intrusion detection system needs

THR_{KH} time for finding the position of the hypervisor whether it is able to deploy the VM's tasks. If VM P is migrated to host R, in the past ξt times and P has been co-tenant in host R for less than THR_{KH} then R is unknown host to P and it can be the malicious one, otherwise it is called Known Host as in Eq. (10)

$$\sum_{i=1,t=1}^{n,m} VM_{i(t)} = (VM_{1(t1)}, VM_{2(t2)}, ..., VM_{i(ti)}, ..., VM_{n(tm)}) \tag{8}$$

$$[Co - Tenant_{i(t)} \times Favorable_{i(t)}](\xi t) > \prod THR_S \tag{9}$$

$$[Co - Tenant_{i(t)} \times Favorable_{i(t)} \times KnwHost_{i(t)}](\xi t) > \prod THR_{KH} \tag{10}$$

These equations help to observe the most favourable physical hosts which help to save the resources for the further deployments productively. By using probabilistic model (Chhabra & Singh, 2018), accurate estimations are calculated which depend on the particular set of data. The filtered ones are called favourable qualified sets ($FQPH$).

$$FQPH\{PH_1, PH_2, ..., PH_n \mid \Xi\} = FQPH(PH_1 \mid \Xi) \times FQPH(PH_2 \mid \Xi) \times ... \times FQPH(PH_n \mid \Xi) \tag{11}$$

where Ξ are the parameters, that is, CPU and Memory. For every host there should be fixed parameters of this function:

$$\aleph (\Xi; PH_1, PH_2, ..., PH_n) = FQPH(PH_1, PH_2, ..., PH_n \mid \Xi) = \prod_{i=1}^{n} FQPH(PH_i \mid \Xi) \tag{12}$$

Where ; denotes separation between two types of input. This equation demonstrates and helps to calculate the probability of each host for handling tasks. It assists in finding the qualified sets which meet the performance constraints. Tasks to particular hosts have been filtered out and assigned for the deployment which has maximum probabaility.

$$= \frac{Requested_i}{Capacity_i} = \frac{R_i^C}{C_i^C} \times \frac{R_i^M}{C_i^M}; \quad i \in \{1, 2, ..., fq\} \tag{13}$$

where fq is qualified set. R_i and C_i defines the requested and capacity computing power of CPU, Memory. Now, the maximum probability among all favourable qualified sets can be calculated.

$$\Phi_{OPT} = \{\max_{i=1}^{fq} \eta(\Xi; PH_1, PH_2, ..., PH_{fq})\} \tag{14}$$

Finally, the tasks are placed on that host for deployment in cloud data centers for executing tasks. For convenience, the operational summary of the proposed model is mentioned in Algorithm 1.

Algorithm 1 *OPH-LB (Cl$_n$,V$_v$,PH$_p$,CP$_{Avl}$,MeM$_{Avl}$,CP$_{req}$,MeM$_{req}$,FQPH,E)*

Initiation: Requesting the jobs by clients for secure load balancing in Multi-tenant cloud.

Output: Searching for secure and optimal physical host.

1:      PHList={}, QPHList={}, FQPH={}

2:      For each client's *CL$_n$* job request do

3:      $T_{i,j} \perp_p^i \perp_j^p \leq B_s \ \forall s_t, \ s_q \in \delta, \ t \neq q$

4:      end for

5:      if $\perp_p^i CP_{req} \leq CP_{avl} \ i \in \{1,2,...,n\}$then

6:          if $\perp_p^i MeM_{req} \leq MeM_{avl} \ i \in \{1,2,...,n\}$ then

7:      QPHList.add(*PHp*)

8:      else

9:      PHList.add(PHp)

10:          end if

11:      end if

12:      if two VMs are co-tenant in ξt times then

13:      $[Co-Tenant_{i(t)} \times Favorable_{i(t)}](\xi t) > \prod THR_S$

14:      FQPHList.add(PHp)

15:      $\aleph(\Xi;PH_1,PH_2,...,PH_n) = FQPH(PH_1,PH_2,...,PH_n \mid \Xi) = \prod_{i=1}^{n} FQPH(PH_i \mid \Xi)$

16:      end if

17:      secure states found

18:      for each physical host ϵ FQPHList do

19:          the probability estimation applies only on favorable qualified physical hosts PH

20:      $\varpi = \dfrac{Requested_i}{Capacity_i} = \dfrac{R_i^C}{C_i^C} \times \dfrac{R_i}{C_i}; \ \ i \in \{1,2,...,fq\}$

21:      end for

22:      for each find the maximum oneamong all out of favorable qualified sets do

23:      $\Phi_{OPT} = \{\max_{i=1}^{fq} \eta(\Xi;PH_1,PH_2,...,PH_{fq})\}$

24:      end for

25:      return secure optimal host

RESULTS AND DISCUSSION

Experiments are conducted with performance and efficiency of the proposed solution and evaluated by considering dynamic creation via Cloudsim simulation environment. It is extremely difficult to examine these long-term experiments on real infrastructures and compared with the proposed technique. OPH-LB is compared with other two well-known methods applied in the literature against the corresponding performances: Random Deployment (RD) and Dynamic Load Balancing (DLB).

Experimental Setup

The simulated cloud network is considered for achieving realistic results. The evaluation is conducted on 100 physical machines with different configurations which can efficiently fulfill the requirements of upcoming load simulation conditions. Every physical machine has its own *VMId* which increases/decreases according to this deployment of cloudlets in dynamic nature. In the space-shared, the machines are partitioned into a set of clusters and every cluster is allocated as a single job and shares the memory space. In time-shared, the computing power is divided by many users and each job runs for a quantum of time. The completion time is compared during analysis of the cloudlets in time-shared and space-shared allocation policies. These 100 physical machines have different available computing resource amount of CPU and Memory. There are 50 cloudlets running continuously on these physical machines. The information captured on these machines and tasks on the global blackboard of the OPH-LB is summarised in Table 2. Some parameters are defined in the range because of the different configurations used for every physical machine.

Table 2
Parameters used in simulation

Parameters	Value
Host Memory	204800
Host Storage	10000000
Host Bandwidth	100000
System architecture	x86
Operating system	Linux
MIPS	250-350
VM Image Size	1000-5000
VM Memory (RAM)	2048 MB
VM Bandwidth	1000-2000
VMM Name	Xen

Makespan

Makespan is calculated by the difference between start time and fininsh time for processing the tasks during scheduling. When we assign the workload requests to physical hosts, it is mainly used during context switching. The processing time increases with increase in the number of requested tasks. As can be seen from Figure 4, this method is analysed and compared with the RD and DLB approaches. In Random Deployment (RD) scheme, it processes the tasks randomly in the cloud data centers. If the number of requested tasks increases and is randomly selected, the capacity of handling that volume of tasks will also weaken. This DLB approach is basically based on the prediction model and is built on some repository knowledge or historical experience. The total processing time of these strategies is always greater than OPH-LB method. It is because OPH-LB method can quickly find the optimal host based on the required resource, so that it has a smaller makespan among all under the same conditions. It also saves the cost of utilisation of resources and indirectly saves power consumption for cloud data centers.
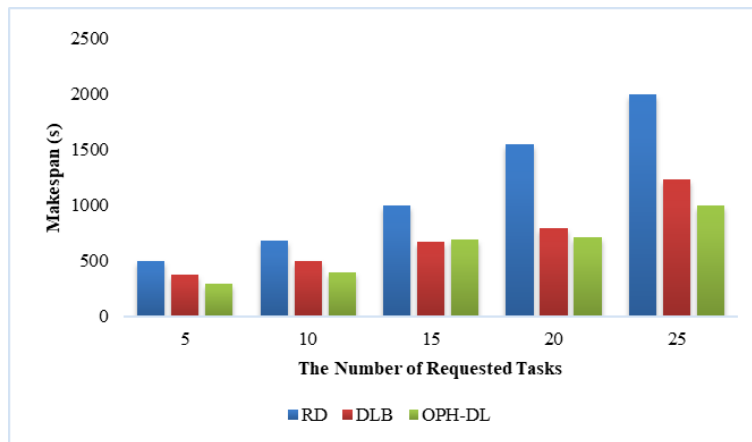
Figure 4. Comparison of makespan between random-migration, DLB and OPH-LB

Failure of Number of Processing Events

In this evaluation, the number of failure tasks is found by the cloudsim simulator during scheduling and deploying the tasks. The finding of these failure nodes in the dynamic environment can only be possible if the chosen physical machine is unable to fulfill some of the demands of requested tasks. When the number of requested tasks increases, chances of the number of failures increases gradually because the ability of handling tasks will weaken slowly. As shown in Figure 5, the framework approach is compared with other existing techniques based on the failed number of tasks during deployment in the simulated cloud network data center. In this figure, RD approach has so many failure sets of tasks because it deploys randomly and in DLB there is less failure than RD because of its knowledge repository experimental values. But these sets of failure cannot manage in the real time. In the case of OPH-LB the analysis shows that the quantity of failure is less and in the plotted experiment, up to 43 requested tasks have no failure node (when tasks = 100, just <15 or few failed tasks). To sum up, the OPH-LB has better solidity and effectiveness for large-scale cloud data centers.
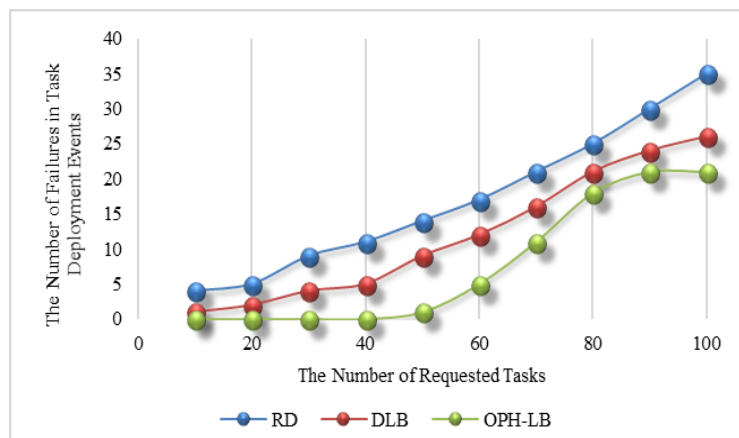
Figure 5. Comparison in failure between random-migration, DLB and OPH-LB

Throughput Performance

This evaluation measure mainly uses an effective measure of load balancing to analyse and evaluate whether it is good or bad in terms of its service performance. These effective measures mainly include the ability of dealing with tasks, the response time to calculate a task request and the number of completed services per unit time. Based on these parameters in cloud system, we can calculate the throughput rate and evaluate the external service performance with respect to increasing time. In this study, the throughput among various numbers of requested tasks was calculated by taking cloudlets as 1000, 2000, 3000 and 4000 as shown in Figure 6.
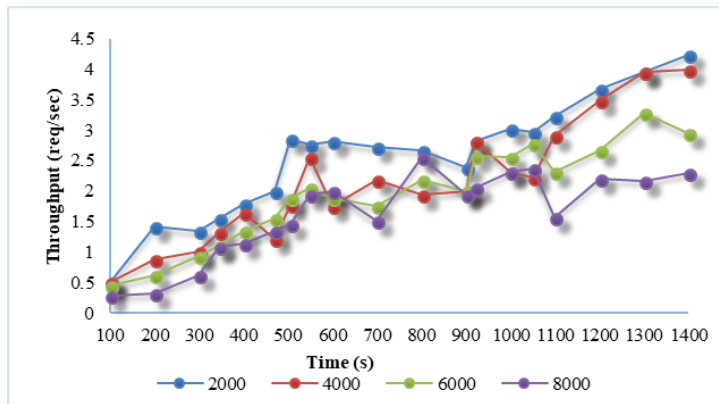
Figure 6. Throughput in different number of requested tasks

Throughput for Succeeded Cloudlets

In this experiment, the succeeded throughput for measuring the performance of requested tasks has been illustrated, except the failure number of nodes. The result in Figure 7 shows that throughput based on succeeded cloudlets gives much better execution than the above result. It also helps to improve the resource utilisation of the cloud data center effectively.
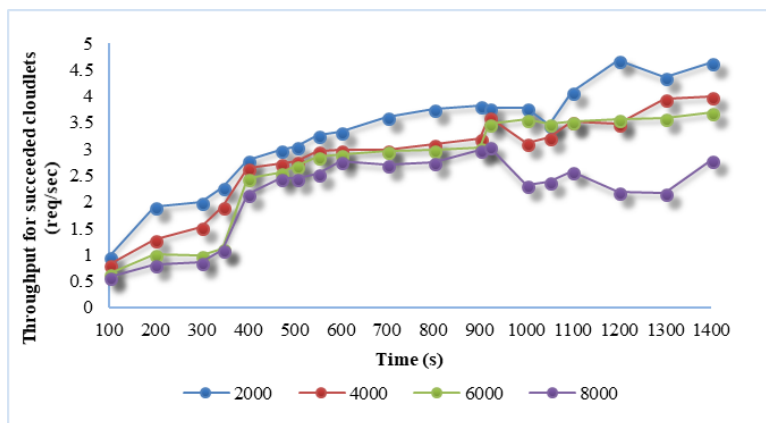
Figure 7. Throughput in different number of succeeded requested tasks

Number of Requested Tasks under Normal State and Secure State

Figure 8 illustrates cloudlet completion time in time-shared scheduler policy for their respective tasks. When tasks increase from 100 to 1000, the time taken by the cloudlets for task completion also increases. At n = 100, 75% of cloudlets complete their processing in time = 160.2 s while at n = 800, only 6% of cloudlets complete their processing in t = 160 s and remaining 94% cloudlets complete their processing in 320.6 or 480.8 s respectively in normal state. Figure 9 shows that some of the cloudlets complete their execution more than 640.8 s but it is quite secure than the previous graph. It shows that even thought it takes more time, it ensures the clients share the resources without any insecurity.
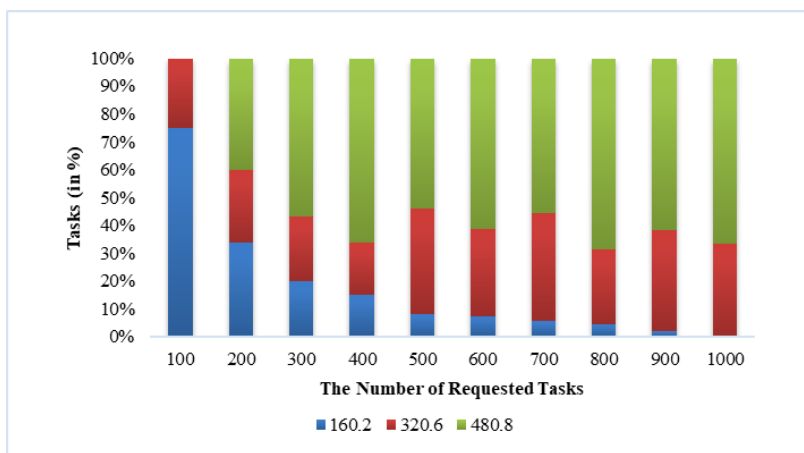
Figure 8. Number of requested tasks under normal state
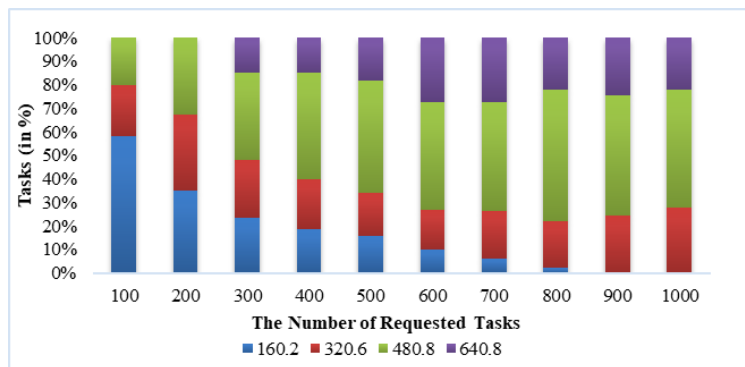
Figure 9. Number of requested tasks under secure state

CONCLUSION

This paper provides a new perspective of task deployment to see the positive effect for secure load balancing. This paper propose an improved solution by using probabilistic model for effective utilisation of resources to provide the service of clients in cloud data centers. Its main purpose is to research and design a novel secure LB policy OPH-LB, to calculate and predict the secure or insecure states for every possible decision of VM selection and VM placement. It certainly reduces information leakage during load balancing and improves security benefit in the cloud. The simulation results show that OPH-LB has many benefits such as it reduces failure rate, shows improvement in throughput, decreases the makespan time, upgrades the utilisation of computing resources and boosts the external service performance too.

REFERENCES

Ang, T. F., Por, L. Y., & Liew, C. S. (2017). Dynamic pricing scheme for resource allocation in multi-cloud environment. *Malaysian Journal of Computer Science, 30*(1), 1-17.

Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., & Buyya, R. (2011). CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience, 41*(1), 23-50.

Cao, J., Li, K., & Stojmenovic, I. (2014). Optimal power allocation and load distribution for multiple heterogeneous multicore server processors across clouds and data centers. *IEEE Transactions on Computers, 63*(1), 45-58.

Chhabra, S., & Singh, A. K. (2016). Dynamic data leakage detection model based approach for MapReduce computational security in cloud. In *2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS)* (pp. 13-19). Bhopal, India: IEEE.

Chhabra, S., & Singh, A. K. (2018). A probabilistic model for finding an optimal host framework and load distribution in cloud environment. *Procedia Computer Science, 125*, 683-690.

Chhabra, S., & Singh, A. K. (2018). Beyond lightning: A systematic review of information security in the age of cloud computing using key management. *International Journal of Computer Engineering and Applications, 11*(12), 299-315. Retrieved from www.ijcea.com ISSN 2321-3469.

Cho, K. M., Tsai, P. W., Tsai, C. W., & Yang, C. S. (2015). A hybrid meta-heuristic algorithm for VM scheduling with load balancing in cloud computing. *Neural Computing and Applications, 26*(6), 1297-1309.

Deng, R., Lu, R., Lai, C., Luan, T. H., & Liang, H. (2016). Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption. *IEEE Internet of Things Journal, 3*(6), 1171-1181.

Deng, X., Wu, D., Shen, J., & He, J. (2016). Eco-aware online power management and load scheduling for green cloud datacenters. *IEEE Systems Journal, 10*(1), 78-87.

Diaz, M., Martín, C., & Rubio, B. (2016). State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing. *Journal of Network and Computer Applications, 67*, 99-117.

Domanal, S. G., & Reddy, G. R. M. (2014). Optimal load balancing in cloud computing by efficient utilization of virtual machines. In *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)* (pp. 1-4). Bangalore, India: IEEE.

Duan, J., & Yang, Y. (2017). A load balancing and multi-tenancy oriented data center virtualization framework. *IEEE Transactions on Parallel and Distributed Systems, 28*(8), 2131-2144.

Han, Y., Chan, J., Alpcan, T., & Leckie, C. (2017). Using virtual machine allocation policies to defend against co-resident attacks in cloud computing. *IEEE Transactions on Dependable and Secure Computing, 14*(1), 95-108.

Kavousi-Fard, A., Niknam, T., Taherpoor, H., & Abbasi, A. (2014). Multi-objective probabilistic reconfiguration considering uncertainty and multi-level load model. *IET Science, Measurement and Technology, 9*(1), 44-55.

Li, X., Qian, Z., Lu, S., & Wu, J. (2013). Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data center. *Mathematical and Computer Modelling, 58*(5-6), 1222-1235.

Lin, C. C., Chin, H. H., & Deng, D. J. (2014). Dynamic multiservice load balancing in cloud-based multimedia system. *IEEE Systems Journal, 8*(1), 225-234.

Martin, T. D. (2010). Hey you - Get off of my cloud: Defining and protecting the metes and bounds of privacy, security, and property in cloud computing. *Journal of Patent and Trademark Office Society, 92*, 283-314.

Moon, S. J., Sekar, V., & Reiter, M. K. (2015). Nomad: Mitigating arbitrary cloud side channels via provider-assisted migration. In *Proceedings of the 22nd Acmsigsac Conference on Computer and Communications Security* (pp. 1595-1606). Denver, USA: ACM.

Papagianni, C., Leivadeas, A., Papavassiliou, S., Maglaris, V., Cervello-Pastor, C., & Monje, A. (2013). On the optimal allocation of virtual resources in cloud computing networks. *IEEE Transactions on Computers, 62*(6), 1060-1071.

Ramezani, F., Lu, J., & Hussain, F. K. (2014). Task-based system load balancing in cloud computing using particle swarm optimization. *International Journal of Parallel Programming, 42*(5), 739-754.

Sun, Q., Shen, Q., Li, C., & Wu, Z. (2016). Selance: Secure load balancing of virtual machines in cloud. *Trustcom/BigDataSE/I SPA, 2016 IEEE* (pp. 662-669). Tianjin, China: IEEE.

Zhang, Y., Juels, A., Reiter, M. K., & Ristenpart, T. (2012). Cross-VM side channels and their use to extract private keys. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security* (pp. 305-316). Raleigh, USA: ACM.

Zhang, Z., Chong, E. K., Pezeshki, A., Moran, B., & Howard, S. D. (2017). Near-optimal distributed detection in balanced binary relay trees. *IEEE Transactions on Control of Network Systems, 4*(4), 826-837.

Zhao, J., Hu, L., Ding, Y., Xu, G., & Hu, M. (2014). A heuristic placement selection of live virtual machine migration for energy-saving in cloud computing environment. *PloS One, 9*(9), e108275.

Zhao, J., Yang, K., Wei, X., Ding, Y., Hu, L., & Xu, G. (2016). A heuristic clustering-based task deployment approach for load balancing using bayes theorem in cloud environment. *IEEE Transactions on Parallel and Distributed Systems, 27*(2), 305-316.

Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems, 28*(3), 583-592.

On Evaluation of Network Intrusion Detection Systems: Statistical Analysis of CIDDS-001 Dataset Using Machine Learning Techniques

Abhishek Verma* and Virender Ranga

Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India

ABSTRACT

In this era of digital revolution, voluminous amount of data are generated from different networks on a daily basis. Security of this data is of utmost importance. Intrusion detection systems have been found to be one of the best solutions in detecting intrusions. Network intrusion detection systems are employed as a defence system to secure networks. Various techniques for the effective development of these defence systems are found in the literature. However, research on the development of datasets used for training and testing purposes of such defence systems is of equal concern. Better datasets improve the online and offline intrusion detection capabilities of detection models. Benchmark datasets like KDD 99 and NSL-KDD cup 99 are obsolete and do not contain network traces of modern attacks like Denial of Service, hence are unsuitable for the purpose of evaluation. In this study, a detailed analysis of CIDDS-001 dataset was conducted and the findings are presented. A wide range of well-known machine learning techniques were used to analyse the complexity of the dataset. Evaluation metrics including detection rate, accuracy, false positive rate, kappa statistics, and root mean squared error were utilised to assess the performance of employed machine learning techniques.

Keywords: Anomaly, decision tree, *k*-means clustering, *k*-nearest neighbour, labelled flow, metrics, random forests, signature

INTRODUCTION

Network security has rapidly become one of the most pressing issues and of concern for web users and service providers with a continual growth in web utilisation (Medaglia & Serbanati, 2010). A secure network can be characterised in terms of its hardware and software immunity against different intrusions. A network can be secured by

incorporating strong observing, examination and safeguard procedures. Network Intrusion Detection System (NIDS) incorporates these procedures to defend against network intrusions (Debar, Dacier & Wespi, 1999). These defence systems perform continuous monitoring of network traffic, analyse and report any intrusions. The major components of this system include traffic collector, analysis engine, signature database and alarm storage, as shown in Figure 1.
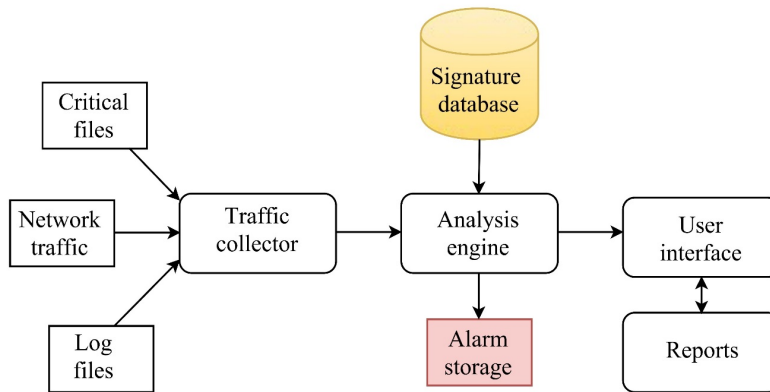
Figure 1. Components of intrusion detection system

Each component plays an important role in intrusion detection. Network traffic is captured by the traffic collector, that is, packet traces, analysis engine conducts a deep analysis of the captured traffic information and sends alarm signals to alarm storage when intrusion is detected. The signature database stores the signatures or patterns of known intruders, and these signatures are used for matching purpose. A typical NIDS is illustrated in Figure 2.
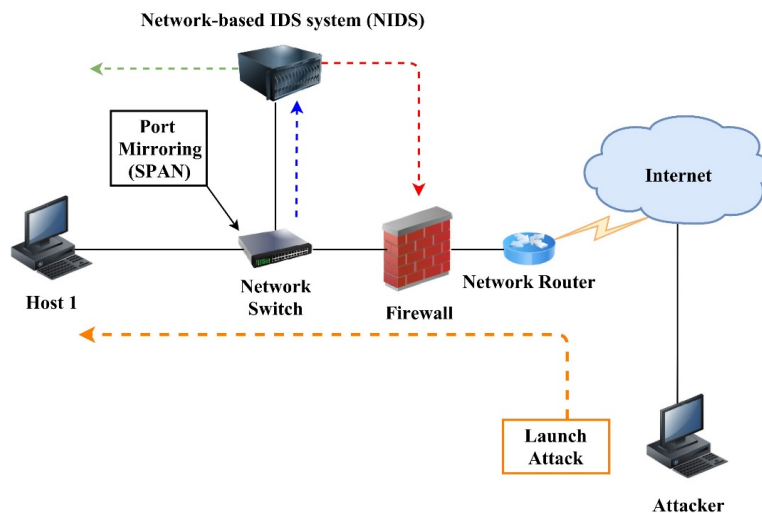
Figure 2. Illustration of network intrusion detection system

NIDS can be classified in two classes, that is, misuse detection (MD) (Zhengbing, Zhitang, & Junqi, 2008) and anomaly detection (AD) (Garcia-Teodoro, Diaz-Verdejo, Macia-Fernandez, & Vazquez, 2009). MD-based systems use traffic patterns of already known attacks for detecting intrusions in the network, while AD-based systems monitor any deviations from normal profiles of network behaviour. MD-based NIDS perform well in accuracy and have a significantly lower false alarm rate (FAR) but they perform poorly for unknown attacks. AD-based NIDS are capable of detecting novel intrusions or attacks, however, they score a higher FAR compared to MD-based NIDS.

Most of the benchmark datasets used for the evaluation of NIDS do not contain network traces of modern attacks (for example, denial of service, port scanning) which make them unsuitable for NIDS. This limitation is solved by CIDDS-001 dataset ("CIDDS-001 dataset", 2017) as it contains modern attack network traces. Machine Learning (ML) has proven to be very effective in the advancement of NIDS (Sommer & Paxson, 2010). It involves a detection system that learns from a dataset consisting of attack and normal packet traces and then classifies incoming network traffic into attack or normal class. The researchers used various well-known supervised and unsupervised learning-based ML models which are exhibited in Figure 3.
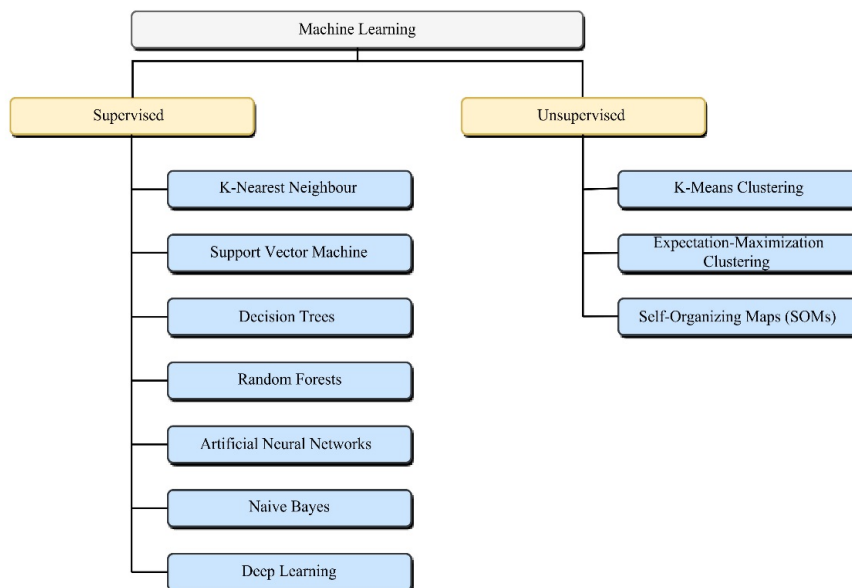
Figure 3. Machine learning techniques used in dataset analysis

CIDDS-001 Dataset

CIDDS-001 ("CIDDS-001", 2017) is a labelled flow-based dataset (Ring, Wunderlich, Grudl, Landes, & Hotho, 2017). It was developed primarily for the evaluation purpose of AD-based NIDS. The dataset consists of traffic from OpenStack and External Servers. CIDDS-001 has 13 features and one class attribute. A total of 11 features were used in this analytical study. The features of Attack ID and Attack Description were neglected in this study because they offer

more information about executed attacks. Hence, these attributes did not significantly contribute to the analysis. About 153,026 instances from external servers and 172,839 instances from OpenStack Server data were collected for analysis. Each instance of the dataset was labelled as normal, attacker, victim, suspicious and unknown class. Table 1 provides the description of CIDDS-001 dataset attributes.

Table 1
Details of CIDDS-001

Sl. No.	Attribute Name	Attribute Description
1	Src IP	IP Address of the source node
2	Src Port	Port of the source node
3	Dest IP	IP Address of the destination node
4	Dest Port	Port of the destination node
5	Proto	Protocol
6	Date first seen	Start time flow first seen
7	Duration	Flow duration
8	Bytes	Transmitted bytes
9	Packets	Transmitted packets
10	Flags	TCP Flags
11	AttackDescription	Additional information about attack
12	AttackType	Type of attack
13	AttackID	Unique attack ID (same type of attacks have same ID)
14	Class	Category or label of the instance

Note: Retrieved from CIDDS-001 dataset. Copyright 2017 by CIDDS-001 dataset (Reprinted with permission)

RELATED WORK

An analytical study on NSL-KDD cup 99 dataset was conducted by Aggarwal and Sharma (2015; "NSL-KDD cup 99 dataset", 1999). The attributes of the dataset were categorised in four classes, that is, basic, content, traffic and host. The contribution of every class was evaluated in terms of DR and FAR. Siddiqui and Naahid (2013) performed the analysis of NSL-KDD dataset for Intrusion Detection (ID) using clustering algorithm-based data mining techniques. They used k-means clustering to build 1000 clusters over 494,020 records and focused on establishing a relationship between attack types and protocols used in intrusion. Artificial neural network (ANN) was utilised for the analysis of NSL-KDD dataset (Ingre & Yadav, 2015).

DR for intrusion detection and attack type classification was found to be 81.2% and 72.9%, respectively. According to Moustafa and Slay (2015), the study on irrelevant features of KDD 99 ("KDD 99 dataset", 1999) and UNSW-NB15 ("UNSW-NB15 dataset", 2017) found that they lead to reduction of NIDS efficiency. An association rule mining algorithm was used for the strongest feature selection from the two datasets and subsequently, classifiers did the task of evaluation in accuracy and false alarm rate (FAR). The results indicate that features of UNSW-

NB15 are much more efficient than the KDD 99 dataset. Kayacik and Zincir-Heywood (2005) studied three Intrusion Detection System (IDS) benchmark datasets using ML algorithms. Clustering and neural network algorithms analysed the IDS datasets and distinguished the differences between synthetic and real-world traffic.

Parsazad, Saboori and Allahyar (2012) proposed a fast feature selection method that finds low-quality features in the dataset. The variance of a random variable is used as a measure for finding the quality of a feature. The authors presented a comparison between popular existing similarity-based algorithms like maximal information compression index, correlation coefficient and least square regression error. The output of these algorithms recommend some features which are then fed to naive bayes and k-nearest neighbour classifiers for the purpose of testing the proposed method. This proposed technique outperformed existing similarity-based algorithms in terms of computational cost.

Rampure and Tiwari (2015) suggested a rough set theory-based feature selection on KDD Cup99 dataset. This is based on the premise that the degree of precision in the data is lowered and data pattern visibility is increased. Based on this premise, facts from imperfect data were discovered. Feature selection using Random Forests was documented in Hasan, Nasser, Ahmad and Molla (2016). The researchers derived a new dataset, RRE-KDD, after removing redundant records from KDD99Train + and KDD99Test+ sets of the NSL-KDD dataset. RRE-KDD is then used for the evaluation of Random Forest (RF). RF technique selects the most important features needed for classification and increases accuracy with reduction in time complexity. Janarthanan and Zargari (2017) analysed the UNSW-NB15 dataset using Weka tool. Different attribute selection techniques like CfsSubsetEval (attribute evaluator) with Greedy Stepwise method, and InfoGainAttibuteEval (attribute evaluator) with Ranker method were used for selecting important features. The best selected subset of attributes was utilised for classification using a few machine learning algorithms including RF. It was found that kappa statistics improved due to classification using selected features. A weighted feature selection method for wifi impersonation detection using AWID ("AWID dataset", 2018) dataset was proposed in Aminanto, Choi, Tanuwidjaja, Yoo and Kim (2018). The researchers used deep-feature extraction and selection for feature reduction in the dataset. The proposed approach achieved an accuracy of 99.918% and a FAR of 0.012%. Verma and Ranga (2018) presented an analytical study on CIDDS-001 using distance-based machine learning methods, whereby kNN and k-means algorithms were used for complexity analysis.

EXPERIMENTAL SETUP

Research Methodology

Weka tool was utilised for performing the analysis (Hall et al., 2009).

- Dataset preprocessing involving handling missing values and feature normalization was performed.
- Supervised and unsupervised machine learning algorithms were executed.
- Results of the simulated algorithms were tabulated and analysed.

Supervised Learning Algorithms

***k*-nearest Neighbour (kNN).** kNN is an instance-based learning and classification technique (Cover & Hart, 1967). Basic founding of kNN algorithm is distance function that calculates the correspondence or dissimilarity between two instances or points. There are different distance measures used in *k*NN. The most common distance measure is Euclidean distance. It can be defined as *D (a, b)* as Equation 1 (Kaur, 2014)

$$D(a,b) = \sqrt{\sum_{i=1}^{r} (a_i - b_i)^2} \tag{1}$$

where a_i is the ith featured element of the instance *a*, b_i is the ith featured element of the instance *b* and *r* is the total number of the features in the dataset.

Support Vector Machine (SVM). SVM aims to find a hyper-plane which classifies all the training instances into different classes (binary classification or multiclass classification) (Suykens & Vandewelle, 1999). SVM algorithm takes observed instances and associated outputs, that is, binary or *N*-ary. Then, it designs a model that can classify new instances into different classes. Training instances are mapped as points in coordinate space, partitioning the instance input sets linearly. There can be the choice of many hyper-planes that can partition the training instance sets but the finest choice will be that with the maximum distance from the nearest instance of any class. In a case of two hyper-planes, *P* which classifies the instances correctly but has less distance from the nearest instance and *Q* which has maximum distance but has a small error in classification, hyper-plane *P* is selected in such case. SVM is effective for high dimensional spaces.

Decision Trees (DT). These are a type of supervised learning algorithms that are mostly used for solving classification problems of ML. Tree models in which the target variable can take discrete values as input are known as classification trees. DT consists of entities like leaves and branches. Leaves signify class labels and branches signify aggregations of attributes that lead to those class labels. It works with both discrete and continuous data. DT algorithm splits the samples into two or more homogeneous sets based on a most significant splitter in input variables. DT suffer from overfitting problem which can be handled by Bagging and Boosting (Quinlan, 1996). DT works effectively over discrete data. Figure 4 shows a typical example of DT (Bhargava, Sharma, Bhargava, & Mathuria, 2013).
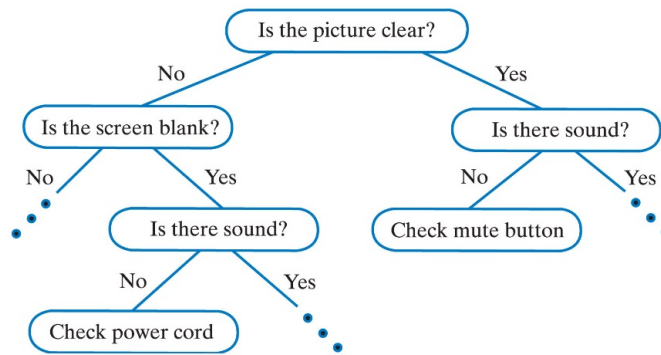
Figure 4. Decision tree [From "Decision Trees", para. 2, by Henrich 2018 (http://www.sfs.uni-tuebingen. de/~vhenrich/ss12/java/homework/hw7/decisionTrees.html). In the public domain.]

Random Forests (RF). As mentioned earlier DT suffers from overfitting problem. RF corrects this problem efficiently by averaging multiple deep decision trees (Breiman, 2001). RF is the ensemble learning algorithm used to solve classification and regression problems. Their operation involves the building of multiple DT during training time. The output is the mode of the classes of the distinct DT when classification task is being performed. RF gives better results than DT. A simple illustration of RF is shown in Figure 5.
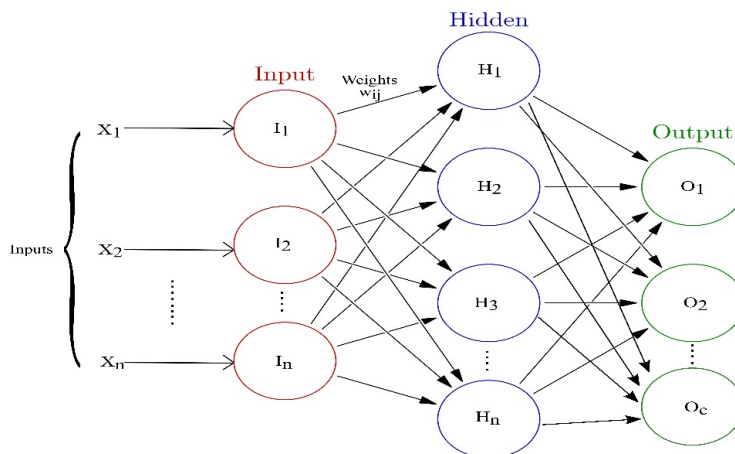
Figure 5. Artificial neural network

Artificial Neural Networks (ANN). ANN can be visualised as a weighted directed graph which consists of nodes and edges (Schalkoff, 1997). Nodes represent artificial neurons and directed edges with weights (strength between neurons) represent connections between artificial neurons. The output of one neuron acts as input to another neuron. ANN receives input from external world in the form of vector, that is, resembling some pattern or image. The weights are adjusted during learning of ANN which further help to solve the classification problems. ANN architecture consists of the input layer, output layer and hidden layer, each layer consists of

neurons. Input layer receives input from the external world, output layer responds to the input fed to input layer on the basis of its learning capability. Hidden layer is intermediary between the input layer and output layer; it transforms the input in some manner such that output layer can utilise. These layers can be partially or fully connected. In this study, the researchers used multilayer perceptron model with back propagation learning.

General ANN architecture (I-H-O) for c class is shown in Figure 6, where I represents the count of input nodes, H represents the count of hidden layer nodes and O represent the count of output nodes.
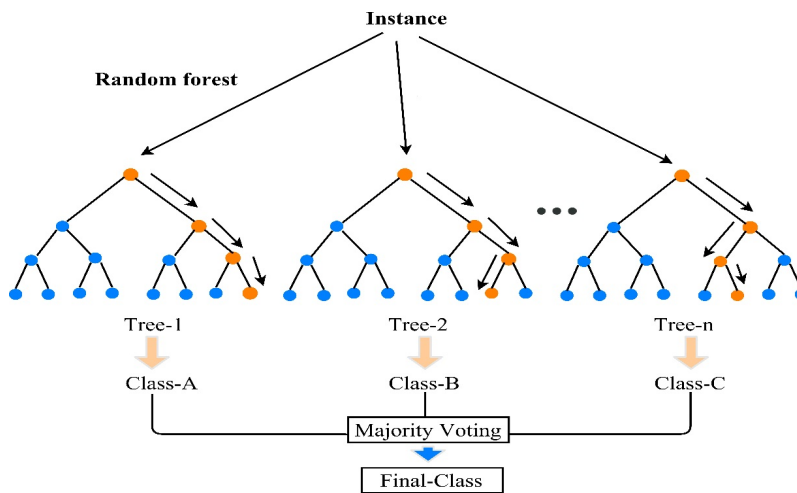
Figure 6. Random forests

Naive Bayes (NB). NB approaches are a family of simple probabilistic classifiers constructed by applying Bayes theorem. NB considers naive assumption of independence between every pair of features or attributes (Lewis, 1998). By applying a suitable pre-processing of training data NB can compete with most of the advanced approaches in its domain, that is, SVM and ANN. NB is easy to be trained using supervised learning configuration. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood. In other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Equation 2 represents Bayes theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

where *A* represents target attribute or dependent event, *B* represents predictor attribute or prior event. *P(A)* is said to be priori probability of *A* and *P(A|B)* is called as posteriori probability of *B* and *P(B|A)* is likelihood of *B* if the hypothesis *A* is true.

Deep Learning (DL). It is a method based on the learning of data representations in contrast to task definite methods without getting stuck to local minima. DL comprises ANN with more number of hidden layers making it more dense and complex (LeCun, Bengio, & Hinton, 2015).

It can be trained using supervised, semi-supervised or unsupervised learning. In this work, supervised learning is used. Cascaded multiple layers of neurons for feature extraction and transformation are used. It learns multiple representations of data that correspond to different levels of abstraction. Deep learning is applicable to many real-world problem solving situations. Figure 7 illustrates the deep learning model.
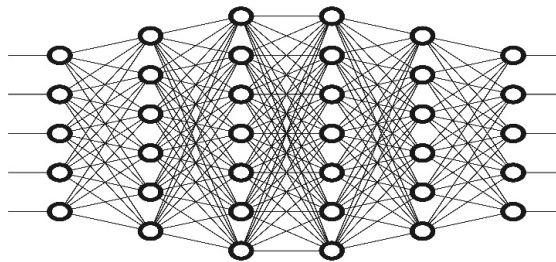
Figure 7. Deep learning network [From "Machine Learning, Deep Learning, and AI: What's the Difference?", para. 1, by Woodie, 2017 (https://www.datanami.com/2017/05/10/machine-learning-deep-learning-ai-whats-difference/). In the public domain.]

Unsupervised Learning Algorithms

k-means Clustering. k-means is known to be one of the simplest unsupervised learning algorithm from distance-based perspective. It partitions n instances into k clusters, where each instance is grouped with the cluster having the nearest mean. Given a set of instances ($p_1, p_2,…, p_n$), where each instance is a d-dimensional real vector. k-means clustering aims to partition p instances into k ($\leq p$) sets $Z = \{Z_1, Z_2,…, Z_k\}$ in order to minimise the variance. k-means can be illustrated as Equation 3 (Kriegel, Schubert, & Zimek, 2016).

$$\arg_Z \min \sum_{i=1}^{k} \sum_{p \in Z_i}^{\max} \|p - \mu_i\|^2 = \arg_z \min \sum_{i=1}^{k} |Z_i| Var Z_i \qquad (3)$$

where μ_i represents the mean of points in set Z_i.

Expectation-Maximisation Clustering (EM). EM clustering technique is very similar to k-means clustering (Moon, 1996). EM clustering extends the basic methodology of k-means clustering in two ways. EM algorithm calculates the probabilities of cluster memberships based on one or more probability distributions. EM aims to maximisze the overall probability of the data, given the final clusters.

Self-Organising Maps (SOM). It is based on unsupervised learning class of neural network models. SOM can perform clustering of data without prior knowledge of class categories of input data (Kohonen, 1998). SOM provides a topology preserving mapping from the high dimensional data space to map neurons (units). This mapping preserves the distance between points. Points which are near to each other are mapped to nearby maps units in the SOM. SOM network can recognise inputs which it has encountered before. Figure 8 represents SOM.
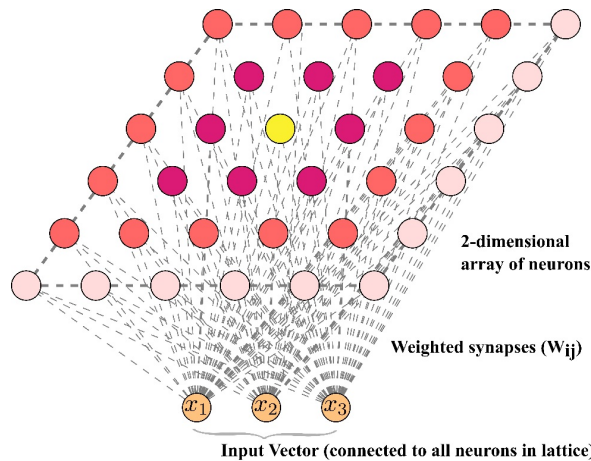
Figure 8. Self-Organising Maps (SOM)

Table 2 lists different Weka classes used for the analysis of CIDDS-001 dataset.

Table 2
Weka classes used for analysis of CIDDS-001

Machine Learning Techniques		Weka Class
Supervised Learning based Techniques	*k*-Nearest Neighbour	weka.classifiers.lazy.Ibk
	Support Vector Machine	weka.classifiers.functions.SMO
	Decision Trees	weka.classifiers.trees.J48
	Random Forests	weka.classifiers.trees.RandomForest
	Artificial Neural Networks	weka.classifiers.functions.MultilayerPerceptron
	Naive Bayes	weka.classifiers.bayes.NaiveBayes
	Deep Learning	weka.classifiers.functions.Dl4jMlpClassifier
	k-Means Clustering	weka.clusterers.SimpleKMeans
Unsupervised Learning based Techniques	Expectation-Maximization Clustering	weka.clusterers.EM
	Self-Organizing Maps	weka.clusterers.SelfOrganizingMap

Evaluation Metrics

Performance of machine learning classifiers in this analytical study were evaluated using eminent metrics, such as detection rate (DR), false positive rate (FPR), f-measure, accuracy, precision, root mean squared error and kappa statistics. All these metrics are evaluated from the elements of the confusion matrix. The elements of confusion matrix are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Typically, TP represents the number of instances that are correctly classified as the attack. TN represents the number of instances that are correctly classified as normal. FP is the count of incorrectly classified normal instances as attack instances. Similarly, FN is the count of incorrectly classified attack instances as normal

instances. Accuracy is defined as the ratio of all correctly classified instances (TP, TN) to all the instances (TP, TN, FP, and FN). Accuracy is denoted by Equation 4. DR (true positive rate) is the ratio of correctly classified instances (TP) as attacks to all the correctly classified attacks (TP) and normal instances (TN). DR is represented by Equation 5. Precision (positive predictive value) is the ratio of TP to a total of TP and FP. Equation 6 represents precision. The harmonic mean of precision and DR is known as f-measure. It is denoted by Equation 7.

Root mean squared error (RMSE) is a quadratic scoring rule which measures the average magnitude of the error (Levinson, 1946). It indicates the accuracy of the probability estimates that are generated by the classification model. Equation 8 represents the RMSE, where P is the original value or forecast value, O represents observed value and n is the number of samples.

In case of multi-class classification, evaluation measures like accuracy, precision and detection rate do not provide a full view of the classifier performance. Precision and detection rate are used in contrast to accuracy when there are imbalanced classes. Kappa statistics (K) is used in such case as it handles multi-class and imbalanced class like problems (Viera & Garrett, 2005). Kappa is defined in Equation 9, where Pr(a) is observed agreement and Pr(e) is expected agreement. K has value less than or equal to 1. Value of 0 or less represents that classifier is useless.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Detection \quad Rate = \frac{TP}{TP + TN} \tag{5}$$

$$Pr\,ecision = \frac{TP}{TP + FP} \tag{6}$$

$$F - measure = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{8}$$

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{9}$$

RESULTS AND DISCUSSION

Various supervised and unsupervised machine learning techniques were utilised for examining the complexity of CIDDS-001. In this study, 10 techniques were used which included classification techniques like kNN, SVM, DT, ANN, DL, RF, NB, and clustering techniques like k-means, EM, SOMs. All the experiments were done on Weka (version 3.9.1) using Intel(R) 7700 having a clock speed of 3.60 GHz processor with 8 GB primary memory running on Windows 10 Pro. Accuracy was given in scale between 0 and 1, that is, 0.36 would be 36% accuracy (multiplied by 100).

Analysis of Supervised Learning Algorithms

Analysis Using *k*NN. Firstly, *k*NN classifier is used for the analysis of external server traffic data. Features named "flows" and "tos" are not considered for the analytical study. Results of *k*NNclassifier execution are shown for 1, 2, 3, 4, and 5 neighbours in Table 3. Secondly, *k*NN classifier is analysed on OpenStack Server traffic data. The researchers selected 172839 instances from week 1 traffic data using reservoir sampling (Vitter, 1985). Results of kNN classifier execution are shown in Table 4. Approximately for every execution of *k*NN classifier on the external server traffic data, models average accuracy is 99%. Maximum accuracy of 99.6% was achieved with 2NN and minimum 99.3% with 5NN. Similarly, for *k*NN classifier execution on OpenStack traffic data models, average accuracy was 100% in each case, this may be due to a random sampling of instances from the dataset file which can lead to some biased instance selections. Dataset can be analysed using other evaluation metrics like ROC curve and FAR (Hand, 2009).

Table 3
Performance of kNN on external server data

Neighbours	Evaluation metrics					Class	Accuracy
	TP rate	FP rate	Precision	Detection rate	F-measure		
1NN	0.995	0.004	0.998	0.995	0.996	suspicious	0.995
	0.993	0.004	0.986	0.993	0.990	unknown	
	1.000	0.000	0.999	1.000	0.999	normal	
	1.000	0.000	1.000	1.000	1.000	attacker	
	1.000	0.000	1.000	1.000	1.000	victim	
2NN	0.997	0.006	0.997	0.997	0.997	suspicious	0.996
	0.990	0.003	0.991	0.990	0.990	unknown	
	1.000	0.000	0.999	1.000	1.000	normal	
	1.000	0.000	1.000	1.000	1.000	attacker	
	1.000	0.000	1.000	1.000	1.000	victim	
3NN	0.994	0.006	0.997	0.994	0.995	suspicious	0.994
	0.991	0.005	0.983	0.991	0.987	unknown	
	1.000	0.000	0.996	1.000	0.998	normal	
	1.000	0.000	1.000	1.000	1.000	attacker	
	1.000	0.000	0.996	1.000	0.998	victim	
4NN	0.996	0.007	0.996	0.994	0.996	suspicious	0.995
	0.989	0.003	0.988	0.989	0.988	unknown	
	1.000	0.000	0.996	1.000	0.998	normal	
	1.000	0.000	1.000	1.000	1.000	attacker	
	1.000	0.000	1.000	1.000	1.000	victim	
5NN	0.993	0.006	0.996	0.993	0.995	suspicious	0.993
	0.989	0.005	0.982	0.989	0.986	unknown	
	1.000	0.000	0.996	1.000	0.998	normal	
	1.000	0.000	1.000	1.000	1.000	attacker	
	1.000	0.000	1.000	1.000	1.000	victim	

Table 4
Performance of kNN on OpenStack server data

Neighbours	Evaluation metrics					Class	Accuracy
	TP rate	FP rate	Precision	Detection rate	F-measure		
1NN	1.000	0.000	1.000	1.000	1.000	victim	1.000
	1.000	0.001	1.000	1.000	1.000	normal	
	0.999	0.000	1.000	0.999	1.000	attacker	
2NN	1.000	0.000	1.000	1.000	1.000	victim	1.000
	1.000	0.001	1.000	1.000	1.000	normal	
	0.999	0.000	1.000	0.999	0.999	attacker	
3NN	1.000	0.000	1.000	1.000	1.000	victim	1.000
	1.000	0.001	1.000	1.000	1.000	normal	
	0.999	0.000	1.000	0.999	0.999	attacker	
4NN	1.000	0.000	1.000	1.000	1.000	victim	1.000
	1.000	0.001	1.000	1.000	1.000	normal	
	0.998	0.000	1.000	0.998	0.999	attacker	
5NN	1.000	0.000	1.000	1.000	1.000	victim	1.000
	1.000	0.001	1.000	1.000	1.000	normal	
	0.999	0.000	1.000	0.999	1.000	attacker	

Analysis Using Support Vector Machine. John Platt's sequential minimal optimisation algorithm was used to train SVM classifier (Platt, 1998). Firstly, SVM was trained over week 1 external server data. Accuracy of 95.3% was achieved in this case with RMSE of 0.320. Performance of SVM classifier on external server traffic is shown in Table 5. Secondly, SVM was trained over OpenStack server data.

In this case classifier achieved accuracy of 95.3% with RMSE of 0.272. Considerably good accuracy was achieved with SVM, hence a SVM based NIDS can be built. Modified algorithms for SVM training can be used to reduce model building time (Tsang, Kwok, & Cheung, 2005). Other variants of SVM can also be used to perform analysis of CIDDS-001. Performance of SVM classifier on OpenStack Traffic is tabulated in Table 6.

Table 5
Performance of SVM on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.976	0.088	0.951	0.976	0.964	suspicious	0.953
0.860	0.018	0.933	0.860	0.895	unknown	
0.981	0.001	0.968	0.981	0.974	normal	
1.000	0.000	0.999	1.000	1.000	attacker	
0.999	0.000	1.000	0.999	1.000	victim	

Table 6
Performance of SVM on OpenStack server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.997	0.000	0.999	0.997	0.998	victim	0.999
1.000	0.000	1.000	1.000	1.000	normal	
0.998	0.000	0.997	0.998	0.998	attacker	

Analysis Using Decision Trees. DT J48 (C4.5) is analysed over external server traffic data. It takes 4.61 seconds to build model for testing. Due to pruning characteristics of J48, model size significantly decreased the training and testing time. The accuracy of 99.7 % was achieved in the first case.

In second run J48 is trained over OpenStack server data. Model building time in this case is 1.27 seconds which is an acceptable time for NIDS training. Fortunately J48 provides 100% correct classifications. Efficient split decides the correctness of DT. Hence it can be concluded that J48 with pruning characteristics not only achieves a good accuracy but also manages space complexity. Table 7 shows the performance of DT on external server traffic data. Table 8 represents the performance of DT classifier on OpenStack server traffic data.

Table 7
Performance of DT on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	0.000	1.000	1.000	1.000	suspicious	0.997
1.000	0.000	0.999	0.999	0.000	unknown	
1.000	0.000	1.000	1.000	0.000	normal	
1.000	0.000	1.000	1.000	0.000	attacker	
1.000	0.000	1.000	1.000	0.000	victim	

Table 8
Performance of DT on OpenStack Server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	0.000	1.000	1.000	1.000	victim	1.000
1.000	0.000	1.000	1.000	1.000	normal	
1.000	0.000	1.000	1.000	1.000	attacker	

Analysis Using Random Forests. RF works by building many small classifiers and then collects votes from each one to decide the class of the test instance. This works on the voting method where small classifiers vote and the majority vote was selected as the output class. Firstly, RF was used for External Server data. It took 46.98 seconds to build the model. The accuracy of 99% was achieved in the first run. Performance of RF classifier on external server traffic is presented in Table 9.

Secondly, RF was used over OpenStack Server data. 100% accuracy was achieved second run model building time of 30.07 seconds. It can be observed that tree based algorithms perform well on CIDDS-001 dataset. RMSE is almost negligible in both cases. Performance of RF classifier on OpenStack server traffic is shown in Table 10.

Table 9
Performance of RF on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	0.000	1.000	1.000	1.000	suspicious	0.999
1.000	0.000	1.000	1.000	1.000	unknown	
1.000	0.000	1.000	1.000	1.000	normal	
1.000	0.000	1.000	1.000	1.000	attacker	
1.000	0.000	1.000	1.000	1.000	victim	

Table 10
Performance of RF on OpenStack server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	0.000	1.000	1.000	1.000	victim	1.000
1.000	0.000	1.000	1.000	1.000	normal	
1.000	0.000	1.000	1.000	1.000	attacker	

Analysis using Artificial Neural Networks. Analysis results show a very poor performance of ANN over both external and OpenStack server data as compared to other techniques employed. This may be possible due to improper dataset preprocessing. However, this can be improved by employing proper feature preprocessing methods (binary feature encoding). In the first run, ANN is tested over External Server data. About 63.8% accuracy is achieved while the model building time is 303.85 seconds which is very high. In second test ANN is tested on OpenStack server data which shows an accuracy of 8.26% with model building time of 413.63 seconds. Hence, ANN is unsuitable for NIDS development based on CIDDS-001. Table 11 and 12 show performance of ANN classifier on external and OpenStack server traffic data respectively.

Table 11
Performance of ANN on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	1.000	0.638	1.000	0.779	suspicious	0.638
0.000	0.000	0.000	0.000	0.000	unknown	
0.000	0.000	0.000	0.000	0.000	normal	
0.000	0.000	0.000	0.000	0.000	attacker	
0.000	0.000	0.000	0.000	0.000	victim	

Table 12
Performance of ANN on OpenStack server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
1.000	1.000	0.083	1.000	0.153	victim	0.083
0.000	0.000	0.000	0.000	0.000	normal	
0.000	0.000	0.000	0.000	0.000	attacker	

In this study, NB was tested first on external server traffic data. In the first case, NB yielded accuracy of 87.1% with 0.226 RMSE. NB took 0.27 seconds to build a model from training data. Secondly, NB is tested over OpenStack traffic. An accuracy of 99% was achieved with 0.074 RMSE. It took 0.34 seconds to build a model from training data in the second case. Results show the effectiveness of probabilistic classifiers, that is, NB takes less time in the model building while showing acceptable accuracy with minimum RMSE. Performance of NB classifier on external and OpenStack server traffic data is presented in Tables 13 and 14 respectively.

Table 13
Performance of NB on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.999	0.354	0.832	0.999	0.908	suspicious	0.871
0.426	0.001	0.994	0.426	0.597	unknown	
0.977	0.000	1.000	0.977	0.988	normal	
1.000	0.000	0.999	1.000	0.999	attacker	
0.999	0.000	0.999	0.999	0.999	victim	

Table 14
Performance of NB on OpenStack server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.998	0.000	1.000	0.998	0.999	victim	0.991
0.989	0.002	1.000	0.989	0.994	normal	
0.998	0.010	0.906	0.998	0.950	attacker	

Analysis using Deep Learning (deeplearning4j). Using the Java based deep learning class (deeplearning4j (http://Deeplearning4j.org)) CIDDS-001 was analysed. In the first run external server traffic data was analysed using DL. Model from training data was built in 916.07 seconds. An accuracy of 94.05% was achieved with RMSE of 0.139. Table 15 shows the performance of DL based classifier on external server traffic data.

In the second test OpenStack server traffic is analysed. The model is built in 457.47 seconds and instances are classified with 99.96% accuracy with 0.015 RMSE. It is quite clear that although it takes a long time to build the model, accuracy achieved is acceptable. DL can be used in high computation capable systems which aim to achieve higher accuracy in the long run. Performance of DL classifier on OpenStack server traffic data is presented in Table 16.

Table 15
Performance of DL on external server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.951	0.078	0.956	0.951	0.953	suspicious	0.941
0.874	0.039	0.864	0.874	0.869	unknown	
0.994	0.001	0.980	0.994	0.987	normal	
1.000	0.000	1.000	1.000	1.000	attacker	
1.000	0.000	1.000	1.000	1.000	victim	

Table 16
Performance of DL on OpenStack server data

Evaluation metrics					Class	Accuracy
TP rate	FP rate	Precision	Detection rate	F-measure		
0.997	0.000	0.999	0.997	0.998	victim	0.999
1.000	0.000	1.000	1.000	1.000	normal	
0.999	0.000	0.997	0.999	0.998	attacker	

Analysis of Unsupervised Learning Algorithms

Analysis using *k*-means Clustering. Firstly, *k*-means clustering was used for analysis of external server traffic data. Features named "flows" and "tos" were not considered for analytical study. Results of execution of k-means algorithm are shown in the form of Multi-class confusion matrix and tabulated in Table 17. A total 38.1086% instances are correctly clustered by the *k*-means algorithm. Secondly, *k*-means clustering is used over OpenStack server traffic data. About 150,000 instances from week 1 traffic data are selected using reservoir sampling. The results of the second execution are in Table 18. In this experiment 99.6627 instances are correctly clustered.

Table 17
Confusion matrix for k-means on external server data

k-means External server		Predicted class					Accuracy
		Suspicious	Unknown	Normal	Attacker	Victim	
Actual Class	suspicious	28952	3788	28061	17218	19833	0.381
	unknown	1977	14045	330	2545	14940	
	normal	32	20	3038	32	3058	
	attacker	0	4	719	8532	0	
	victim	2153	0	0	0	3749	

Table 18
Confusion matrix for k-means on OpenStack server data

k-means External server		Predicted class			Accuracy
		Victim	Attacker	Normal	
Actual class	victim	57955	0	0	0.997
	attacker	0	57963	0	
	normal	90	416	33576	

Analysis using Expectation-Maximisation Clustering. Firstly, EM algorithm was used to analyse external server traffic data. The same set of features as used in k-means clustering were used. In this experiment, accuracy of 45.9% was achieved with model building time of 32.07 seconds. The results of this experiment are shown in the form of multi-class confusion matrix and tabulated in Table 19. In the second experiment, EM was tested over OpenStack server traffic. In this analysis, accuracy of 49.3% was achieved and the model was built in 10.18 seconds. As compared to previously mentioned techniques, this method not only does time costly model building but also performs very badly. Confusion matrix for the second experiment is presented in Table 20.

Table 19
Confusion matrix for EM on external server data

EM External server		Predicted class					Accuracy
		Suspicious	Unknown	Normal	Attacker	Victim	
Actual Class	suspicious	2880	45238	5636	25	44073	0.459
	unknown	715	16202	15848	0	1072	
	normal	199	2	0	5898	81	
	attacker	0	0	0	8877	378	
	victim	2	0	0	5819	81	

Table 20
Confusion matrix for EM on OpenStack server data

EM OpenStack server		Predicted class			Accuracy
		Victim	Attacker	Normal	
Actual Class	victim	13094	0	1142	0.493
	normal	58905	71414	13083	
	attacker	0	14579	622	

Analysis using Self-Organising Maps. In the first experiment, SOM was used to analyse external server traffic data. SOM takes 601.37 seconds to build a model using training data. After applying testing data it was found that SOM correctly clustered 38.4% test instances. In the second experiment SOM was used to analyse OpenStack server traffic data. In this test, SOM built a model in 719.59 seconds. About 46.3% test instances were correctly clustered in this experiment. Table 21, and 22 show confusion matrix for first and second experiment respectively.

Table 21
Confusion matrix for SOM on external server data

SOM External server		Predicted class				Accuracy
		Attacker	Suspicious	No class	Unknown	
Actual Class	suspicious	15399	34252	339938	14263	0.384
	unknown	16030	913	934	15960	
	normal	3090	0	0	3090	
	attacker	8754	0	0	501	
	victim	215	0	0	5684	

Table 22
Confusion matrix for SOM on OpenStack server data

SOM OpenStack Server		Predicted class				Accuracy
		Attacker	Suspicious	No class	Unknown	
Actual Class	victim	0	62	14174	0	0.463
	attacker	50877	20672	21781	50072	
	normal	0	14918	283	0	

Overall Evaluation

From the analysis results, it can be interpreted that most of the supervised learning based classification ML techniques perform better, only ANN fails to give acceptable accuracy. Figure 9 shows the performance of all the used techniques in accuracy. Almost all unsupervised learning based clustering techniques perform poorly. However *k*-means clustering gives good accuracy on OpenStack Server traffic data. These can be improved by proper data cleaning, binary feature encoding, normalisation and data standardisation methods. As clustering techniques require input data to follow a normal distribution for achieving better accuracy, in CIDDS-001 case features do not follow normal distribution and hence, poor accuracy is achieved. Performance of clustering techniques can be improved by capping, flouring and normalisation of attributes. Removal of outliers from the dataset can also improve the clustering performance. Kappa statistics for all classification techniques other than ANN is above average and hence it can be said that anomaly based NIDS using kNN, SVM, DT, RF, NB and DL can be developed.

MACHINE LEARNING TECHNIQUE	kNN	SVM	DT	RF	ANN	NB	DL	k-means	EM	SOMs
External Server	0.995	0.953	0.999	0.999	0.638	0.871	0.941	0.381	0.459	0.384
OpenStack Server	1.000	0.999	1.000	1.000	0.083	0.991	0.999	0.996	0.493	0.463
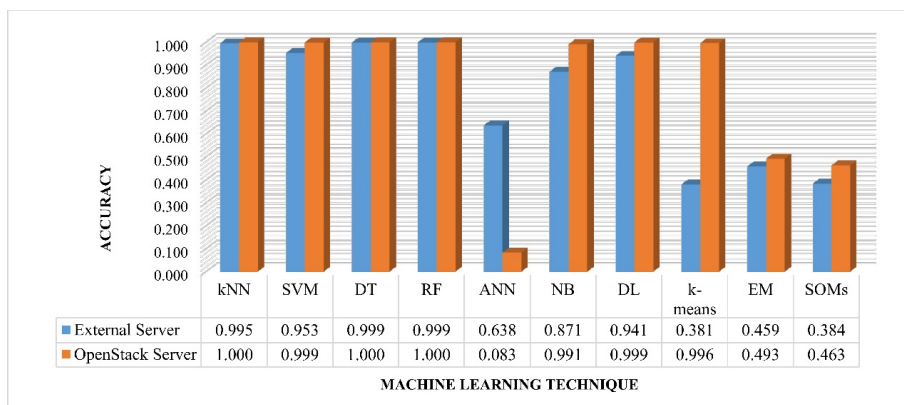
Figure 9. Performance of techniques in accuracy

Figure 10 and Figure 11 show the kappa statistics and model building time for different classification techniques used for the analysis. Model building time is the amount of time an algorithm (ML) takes to build a trained model from training data. This time should be less so that trained model can be employed for intrusion detection in minimum possible time. It can be observed that model building time for *k*NN, DT, RF, NB, *k*-means and EM methods is much

less while ANN, DL, SOM methods take a much higher time to build a model. In case of DL, although it takes a long time to build a trained model, DL gives better accuracy once a model gets completely built. Root mean squared error is one of the important factors to analyse the performance of classifiers on a particular dataset.

	kNN	SVM	DT	RF	ANN	NB	DL
■ External Server	0.992	0.911	0.999	0.999	0.000	0.734	0.889
■ OpenStack Server	0.999	0.999	1.000	1.000	0.000	0.969	0.999

MACHINE LEARNING TECHNIQUE
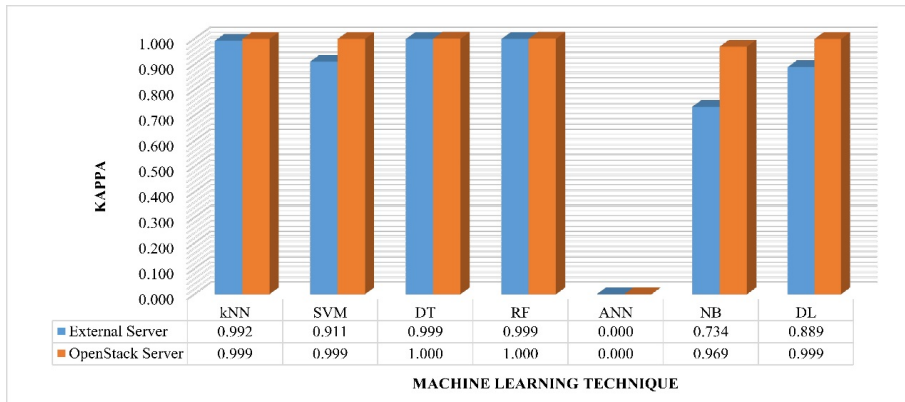
Figure 10. Performance of techniques in Kappa Statistics

	kNN	SVM	DT	RF	ANN	NB	DL	k-means	EM	SOMs
■ External Server	0.02	581.54	4.61	46.98	303.85	0.27	916.07	0.56	32.07	601.37
■ OpenStack Server	0.02	2.54	1.27	30.07	413.63	0.34	457.47	0.25	10.18	719.59

MACHINE LEARNING TECHNIQUE
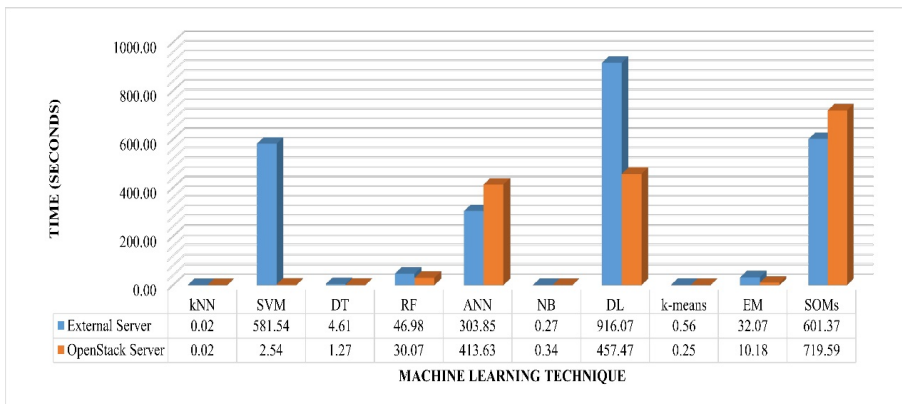
Figure 11. Performance of techniques in model building time

DL can be a suitable choice if target system (where NIDS is to be deployed, that is, a router or some dedicated analysing machine) is having high computational power. Tree model-based techniques like DT and RF can be a suitable choice if the target system is not capable of performing high computations and lack higher storage capabilities.

Figure 12 shows the RMSE in case of different ML techniques used in this study. In this case, ANN attains highest RMSE over both external server and OpenStack server data. SVM outputs RMSE are between 0.250 and 0.350. This is observed by running SVM multiple times. All the performance parameters are closely related and affect each other. Hence, it can be concluded that by making a trade-off between different performance parameters the best technique can be selected for developing anomaly based or signature-based NIDS.

	kNN	SVM	DT	RF	ANN	NB	DL
External Server	0.055	0.320	0.011	0.009	0.398	0.226	0.139
OpenStack Server	0.009	0.272	0.000	0.000	0.482	0.073	0.015
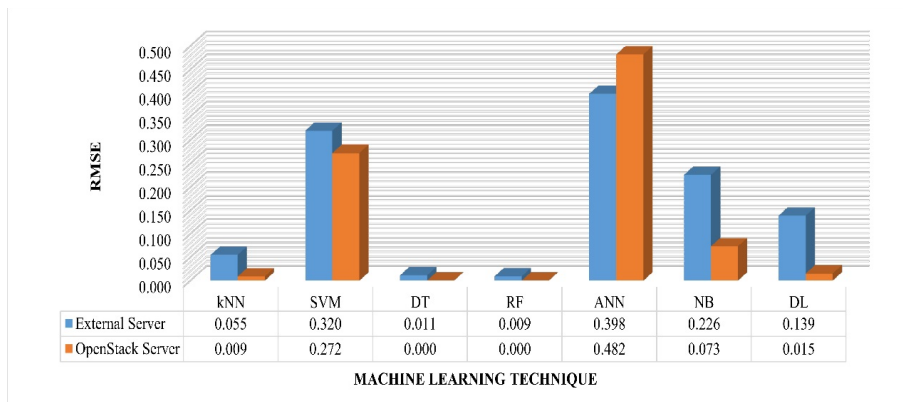
Figure 12. Performance of techniques in root mean squared error

CONCLUSION

In this paper, the statistical and complexity analysis of CIDDS-001 dataset is presented and discussed. Supervised and unsupervised machine learning techniques are utilised to analyse the complexity of the dataset in eminent evaluation metrics. Evaluation results show that k-nearest neighbour, decision trees, random forests, naive bayes and deep learning based classifiers can be used to develop an efficient network intrusion detection systems. Based on the evaluation results it is concluded that CIDDS-001 dataset is suitable for the evaluation of Anomaly-based Network Intrusion Detection Systems. As follow-up, the researchers aim to carry out an in-depth comparative study of the CIDDS-001 dataset with existing benchmarking datasets.

ACKNOWLEDGMENTS

REFERENCES

Aggarwal, P., & Sharma, S. K. (2015). Analysis of KDD dataset attributes-class wise for intrusion detection. *Procedia Computer Science, 57*, 842-851.

Aminanto, M. E., Choi, R., Tanuwidjaja, H. C., Yoo, P. D., & Kim, K. (2018). Deep abstraction and weighted feature selection for Wi-Fi impersonation detection. *IEEE Transactions on Information Forensics and Security, 13*(3), 621-636.

AWID. (2018). *AWID dataset.* Retrieved January 2, 2018, from http://icsdweb.aegean.gr/awid/download. html

Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(6), 1114-1119.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

CIDDS-001 dataset. (2017). CIDDS – Coburg intrusion detection data set. Retrieved January 22, 2018, from https://www.hs-coburg.de/forschung-kooperation/forschungsprojekte-oeffentlich/ ingenieurwissenschaften/cidds-coburg-intrusion-detection-data-sets.html

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21-27.

Debar, H., Dacier, M., & Wespi, A. (1999). Towards a taxonomy of intrusion-detection systems. *Computer Networks, 31*(8), 805-822.

Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers and Security, 28*(1), 18-28.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10-18.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning, 77*(1), 103-123.

Hasan, M. A. M., Nasser, M., Ahmad, S., & Molla, K. I. (2016). Feature selection for intrusion detection using random forest. *Journal of Information Security, 7*(03), 129-140.

Henrich, V. (2018). *Decision trees.* Retrieved January 8, 2018, from http://www.sfs.uni-tuebingen. de/~vhenrich/ss12/java/homework/hw7/decisionTrees.html

Ingre, B., & Yadav, A. (2015). Performance analysis of NSL-KDD dataset using ANN. In *Proceedings of International Conference on Signal Processing and Communication Engineering Systems (SPACES)* (pp. 92-96). Guntur, India: IEEE.

Janarthanan, T., & Zargari, S. (2017). Feature selection in UNSW-NB15 and KDDCUP'99 datasets. In *Proceedings of 26th International Symposium on Industrial Electronics (ISIE)* (pp. 1881-1886). Edinburgh, UK: IEEE.

Kaur, D. (2014). A comparative study of various distance measures for software fault prediction. *International Journal of Computer Trends and Technology, 17*(3), 117-120.

Kayacik, H. G., & Zincir-Heywood, N. (2005). Analysis of three intrusion detection system benchmark datasets using machine learning algorithms. In *Proceedings of International Conference on Intelligence and Security Informatics* (pp. 362-367). Atlanta, GA, USA: Springer.

KDD 99 dataset. (1999). *Index of /ml/machine-learning-databases/kddcup99-mld.* Retrieved January 5, 2018, from http://archive.ics.uci.edu/ml/machine-learning-databases/kddcup99-mld/.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing, 21*(1), 1-6.

Kriegel, H. P., Schubert, E., & Zimek, A. (2016). The (black) art of runtime evaluation: Are we comparing algorithms or implementations? *Knowledge and Information Systems, 52*(2), 1-38.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Levinson, N. (1946). The Wiener (root mean square) error criterion in filter design and prediction. *Studies in Applied Mathematics, 25*(1-4), 261-278.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *Proceedings of European Conference on Machine Learning* (pp. 4-15). Chemnitz, Germany: Springer.

Medaglia, C. M., & Serbanati, A. (2010). An overview of privacy and security issues in the internet of things. *The Internet of Things* (pp. 389-395). New York, USA: Springer.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine, 13*(6), 47-60.

Moustafa, N., & Slay, J. (2015). The significant features of the UNSW-NB15 and the KDD99 data sets for Network Intrusion Detection Systems. In *Proceedings of 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)* (pp. 25-31). Kyoto, Japan: IEEE.

NSL-KDD dataset. (1999). *KDD Cup 1999 Data Data Set*. Retrieved January 18, 2012, from http://archive.ics.uci.edu/ml/datasets/kdd+cup+1999+data.

Parsazad, S., Saboori, E., & Allahyar, A. (2012). Fast feature reduction in intrusion detection datasets. In *Proceedings of the 35th International Convention MIPRO* (pp. 1023-1029). Opatija, Croatia: IEEE.

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report MSR-TR, 98*(14), 1-21. Retrieved from https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/.

Quinlan, J. R. (1996, August 04-08). Bagging, boosting, and C4. 5. In *Proceedings of Fourteenth National Conference on Artificial Intelligence* (pp. 725-730). Providence, Rhode Island: ACM.

Rampure, V., & Tiwari, A. (2015). A rough set based feature selection on KDD CUP 99 data set. *International Journal of Database Theory and Application, 8*(1), 149-156.

Ring, M., Wunderlich, S., Grüdl, D., Landes, D., & Hotho, A. (2017). Flow-based benchmark data sets for intrusion detection. In *Proceedings of the 16th European Conference on Cyber Warfare and Security* (pp. 361-369). Dublin, Ireland: ACPI.

Schalkoff, R. J. (1997). *Artificial neural networks* (Vol. 1). New York: McGraw-Hill.

Siddiqui, M. K., & Naahid, S. (2013). Analysis of KDD CUP 99 dataset using clustering based data mining. *International Journal of Database Theory and Application, 6*(5), 23-34.

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *Proceedings of IEEE Symposium on Security and Privacy (SP), 2010* (pp. 305-316). Berkeley/Oakland, CA, USA: IEEE.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters, 9*(3), 293-300.

Tsang, I. W., Kwok, J. T., & Cheung, P. M. (2005). Core vector machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research, 6*(Apr), 363-392.

UNSW-NB15 dataset. (2017). *The UNSW-NB15 data set description.* Retrieved January 12, 2018, from http://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets/.

Verma, A., & Ranga, V. (2018). Statistical analysis of CIDDS-001 dataset for Network Intrusion Detection Systems using distance-based machine learning. *Procedia Computer Science, 125*, 709-716.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360-363.

Vitter, J. S. (1985). Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS), 11*(1), 37-57.

Woodie, A. (2017). *Machine learning, deep learning, and AI: What's the difference?* Retrieved January 6, 2018, from https://www.datanami.com/2017/05/10/machine-learning-deep-learning-ai-whats-difference/

Zhengbing, H., Zhitang, L., & Junqi, W. (2008). A novel Network Intrusion Detection System (NIDS) based on signatures search of data mining. In *Proceedings of First International Workshop on the Knowledge Discovery and Data Mining, WKDD 2008.* (pp. 10-16). Adelaide, SA, Australia: IEEE.

Protocols Performance Investigation using Ad Hoc WLAN for Healthcare Applications

Piyush Yadav[1]*, Rajeev Agrawal[1] and Komal Kashish[2]

[1]*Department of Electronics and Communication Engineering,*
G L Bajaj Institute of Technology and Management, Greater Noida, India
[2]*RWTH Aachen University, 52062 Aachen, Germany*

ABSTRACT

Medical emergencies are life-threatening situations which need immediate interventions. Response to such emergency situation will depend upon the resources available at that instant. A response time of any healthcare system will make a difference between life and death of a victim. In this work, the authors investigated the performance of various routing protocols for Ad Hoc network application in healthcare services for emergency situations where the data from the patients were collected via wearable devices. The comparison study is done to test the suitability of the Medium Access Control (MAC) protocols for such m-health applications established through an ad hoc network. The established network, however, diagnoses the patient's parameters on real-time basis and sends the data to the doctors and to a database within the hospital, which is constantly monitored, creating a virtual Intensive Care Unit (ICU) for a victim so as to address any abnormal behaviour during remote process. The proposed architecture was designed with the aim to manage emergency situations by reducing delays and increase the effectiveness of the patient monitoring mechanism in situations where the delay in providing medical aid to the patient maked a difference to his / her survival. To demonstrate the effectiveness of the study, an Electrocardiogram (ECG) pattern of a patient was monitored on real-time basis. A custom packet frame was designed for data transmission having a special field to store ECG data along with other general information as per the packet standards.

Keywords: Ad hoc network, ECG, healthcare system, MAC protocols, performance evaluation, telemedicine, WLAN

INTRODUCTION

In the past few decades the advancement in communication technology has helped health care professionals to adopt the technology available, making their routine job easier and safer. Although such technical advancements

may be efficient, fast and easy to use, the reliability and fault-free services, along with adoption and getting used to such new technologies, especially for critical applications from a practitioner's point of view are some of the challenging issues. The impact of Internet growth in telemedicine advances is very significant. With networks having capabilities of packet switching, it's become possible for different types of data to be transmitted on a single transmission medium. While the US, European nations and other developed countries have such applications happening in actual practice, it still remains a mere concept for developing and underdeveloped nations. There are many reasons for this, such as financial support, lack of infrastructure and unavailability of technical resources, being some of them. In such limitations one can deploy or choose a network that is simpler and widely available in every part of one's country for one's healthcare applications. The capability of communication network can support healthcare applications in many areas, from information sharing to audio (tele-consultation), still images, video images database, vital signs (ECG: electrocardiogram, EEG: electroencephalogram) and reliability (Cicalo, Mazzotti, Moretti, Tralli, & Chiani, 2016). It is important to note that the reliability aspect of such critical applications is very important as it deal with the quality of service QoS assurance. Further, such work should demonstrate a quantitative reliability aspect and promise a better-quality assurance when supporting such lifesaving applications, as quality is perhaps the most important attribute in creating a trust of users for such healthcare services.

Taking into consideration the aspect of easy deployment and scalability in view of developing countries such as India, Wireless Local Network (WLAN) can be a better choice, which is based on IEEE 802.11 standard. WLAN provides mobility to its users within a few meters, which is suitable for scenarios where patient nodes are assumed to have mobility within the hospital premises (Batistatos, Tsoulos, & Athanasiadou, 2012; Koutsakis, 2017). For this application, the authors simulated WLAN Ad Hoc network where the patients served as the nodes of the network. Ad Hoc allows ease of data transfer by being an autonomous system, where every node is responsible for transmitting and receiving the desired data. Since we are considering a case of emergency where multiple new nodes need to be added to the network, Ad Hoc offers an advantage of quick adaptation to topology changes (Yadav, Agrawal, & Kashish, 2018). The authors used the structure of ECG packets to serve as the data packets in their simulated environment. As WLAN is chosen as the network in the paper, the authors aimed to analyse the performance of three routing protocols, Ad Hoc On-Demand Distance Vector (AODV), Destination-Sequenced Distance-Routing (DSDV) and Dynamic Source Routing (DSR) on different MAC protocols, IEEE 802.11-DCF (Distributed Coordination Function), Carrier Sense Multiple Access (CSMA), Carrier Sense Multiple Access / Collision Avoidance (CSMA/CA) and their influence, application and suitability in remote healthcare services.

In this paper, first, general information about the wireless routing protocols, along with the various MAC protocols which are widely used has been introduced. This will give basic understanding to the reader to clearly know the performance of these protocols for the proposed application. The next section presents the proposed network architecture and its simulation framework. Further, the parameters on which the three protocols are evaluated are highlighted, followed by the simulation results and discussions. Finally, conclusions are drawn in the last section.

WIRELESS ROUTING PROTOCOLS IN AD HOC NETWORKS

The major categories of WLAN networks can be chalked up to two types - infrastructure-oriented and Ad Hoc Networks or infrastructure-less (Vetrivelan & Reddy, 2008). In infrastructure-oriented networks, nodes are always connected to an access point which acts as a central point. Meanwhile, in Ad Hoc networks, the mobile nodes create a decentralised network, without the need of a central node (Perkins & Watson, 1994). The pattern of the network changes with the change in routing table, which is done according to the requirement of the application.

Even though infrastructure-based networks outperform Ad Hoc network in efficiency, however, when additional nodes (or emergency case patients in the case of this study) are to be added immediately, infrastructure based network will fail to provide such support. Hence, in this paper, the authors decided to keep the focus on Ad Hoc networks for implementing the desired patient monitoring system. Routing protocols in Ad Hoc networks are further classified into three categories: Table Driven, On Demand and Hybrid (Subramanya, Shwetha, & Devaraju, 2012; Kim, Moon, Cho, 2009; Mbarushimana & Shahrabi, 2007; Toh, 2002; Corson & Macker, 1999).

Table-Driven routing protocols have the characteristic feature of the routing tables which are updated periodically on the information of the nodes within the network (Abolhasan, Wysocki, & Dutkiewicz, 2004). On-demand Routing Protocol decides the route on the basis of the demand response that it obtains after the entire network is flooded by Route Request Packets (RRPs). The combination of the best features of these two protocols, known as the hybrid protocol is presented in Mbarushimana and Shahrabi (2007) and Subramanya et al (2012). The routing protocols discussed in this paper are AODV, DSDV and DSR, while AODV and DSR are examples of On-Demand, and DSDV is an example of Table-Driven routing protocol (Royer & Toh, 1999).

Ad Hoc On-Demand Distance Vector

The primary feature of Ad-Hoc On-Demand Distance Vector (AODV) protocol is the routing tables, which give continuous updates to design the desired routes (Corson & Ephremides, 1995; Perkins & Watson, 1994). These tables have updated information about the active nodes in the network. This is done with the use of messages such as route request (RREQ), route reply (RREP) and route error (RRER) (Liu, Yang, & Wang, 2013). The changes in network configuration (topology) are periodically detected via the table and the route designing is done accordingly. In AODV, the route of packet (information) between two nodes is initiated only by exploring the best possible route, taking into consideration the information about other nodes from the table.

Destination-Sequenced Distance Vector

Destination-Sequenced Distance vector (DSDV), often defined as an advancement of the classical Bellman-Ford Routing Algorithm is a source-initiated protocol, that is, a route is designed after a source requests the path discovery to a fixed destination in the network (Ford

& Fulkerson, 1962). It is a hop to hop protocol (Perkins & Watson, 1994). Every individual node maintains a table continuously with information regarding the next hop and the number of hops to all its neighboring nodes and the nodes that it can access. DSDV solves the problem of routing algorithm problem in networks.

Dynamic Source Routing

Dynamic Source Routing is an "on-demand" routing protocol, that is, the path discovery is only done when a source responds to any of the RRPs that flood the network periodically with a request for path to any destination in the network. DSR is self-organised and self-configuring, since it eliminates the need for both infrastructure and monitoring (Johnson & Maltz, 1996). The senders under DSR protocols are aware of the entire hop-to-hop route to the destination (Basavaraju, Sarkar, & Puttamadappa, 2006).

MEDIUM ACCESS CONTROL PROTOCOLS METHODOLOGY OVERVIEW

IEEE 802.11-Distributed Coordination Function (DCF)

802.11 MAC is one of the distributed Wireless MAC protocols and specifies two modes: Distributed Coordination Function (DCF) and Point Coordination Function (PCF). DCF, used for ad-hoc purposes, is a distributed algorithm, where all the nodes run the algorithm individually, whereas PCF needs an infrastructure to operate as an access point runs the centralised algorithm (Youssef, Vasan, & Miller, 2002). The DCF mode of 802.11 is a Carrier Sense Multiple Access (CSMA) with Collision Avoidance type protocol and is often referred to as a combination of the CSMA and Multiple Access with Collision Avoidance (MACA) schemes (Basavaraju et al., 2006).

In DCF protocol, the sender node senses the medium to check if another node is transmitting currently. If the medium is found to be free, the node waits for varying DCF Interframe (DIFS) Space time slots. If the medium is available for the decided time slot, the node transmits the data; lest a back-off process is enabled where the node waits till the current transmission is complete.

Carrier Sense Multiple Access (CSMA)

CSMA is the simplest and earliest of the MAC protocols. Operating on CSMA, the sender node senses the medium to check if it is available. If it is not, it waits for a random time period and then attempts again. If the medium is free the transmission of data is performed by the sender node.

Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA)

An enhanced version of CSMA, is CSMA Collision Avoidance (CSMA/CA), the node listens to the medium for a specified period of time and transmits the data packets only when the medium is checked to be idle. The receiver node is programmed to send an acknowledgement

'ACK' to the sender. If no such acknowledgment is received, the packet is assumed to be lost and another transmission takes place.

PROPOSED NETWORK ARCHITECTURE FOR HEALTHCARE MONITORING USING WLAN

Attributable to WLAN's limited range coverage and mobility, an indoor scenario of a hospital is assumed. As in Figure 1, each patient is provided a unique ID tag. This helps in maintaining and retrieving the patient's data efficiently. Various forms of patient data, such as ECG records, scanned images, and patient vitals (blood pressure, oxygen saturation level, heart rate) are stored into the database whenever required. This database ensures smooth contact between patients' data collection and providing this data to the nurse stations or to the doctor concerned at any moment in time. The transmission of data uses routing protocols depending on the requirements of the particular data type which will be discussed in this paper. In the implementation of the scenario, the authors assumed the patients as nodes in a network design, which transmit ted data to a central database which could be accessed by doctors as well as nurse stations for regular monitoring and emergency situations. For the purpose of simplicity, the authors took only one parameter, ECG of the patient that was supposed to be sampled and transmitted from the patient's device. For accurate data samples, actual ECG data samples available on online database were used.
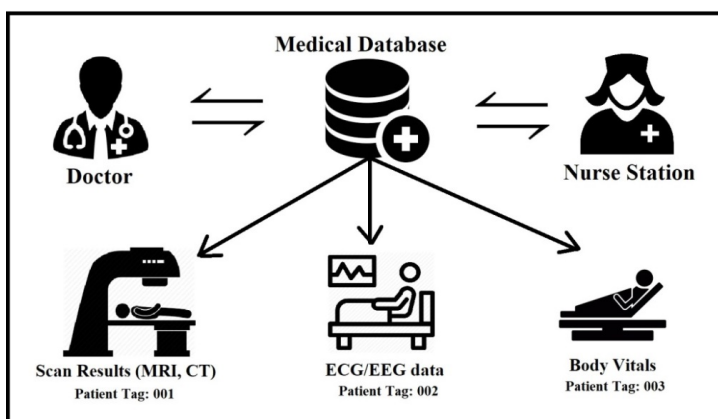
Figure 1. System architecture for proposed scenario

It is important for any support system to have the flexibility of scaling based on the need. The proposed system/framework can be scaled by adding additional nodes representing a situation where there is a large number of patients. Such situation can very well envisage in case of disaster management where the size of the network is not known. In such situation, the patients can be added to this Ad Hoc network and the vital parameters can be transferred to the main hospital or control stations for possible diagnosis of the disease and course of action. The

suggested framework can be an effective solution in such cases where it is difficult to carry all the equipment or staff. A virtual ICU or hospital can be formed to provide medical help to critical patients and could save lives of many patients. Through this, one can also minimise the delays caused by taking up the vital parameters of the patients physically.

Table 1
WLAN Properties

Technology	Frequency	Range	Upstream	Downstream
OFDM/MIMO	5, 2.4 GHz	30m	54 Mbps	54 Mbps

Note: Adapted from Standards ISO/IEEE (2008)

PERFORMANCE METRICS OF ROUTING PROTOCOLS

The evaluation of the performance of the three routing protocols was done on the basis of the following parameters:

Packet Delivery Ratio

Packet Delivery Ratio (PDR) is the number of packets successfully received by the receiver (R_i) over a number of packets sent by the source node (S_i) (Tamilselvan & Sankaranarayanan, 2007). It measures the reliability of the routing protocols. The packet delivery ratio of channel is N and i represents sequence number.

$$\text{Packet delivery ratio} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{R_i}{S_i}\right)$$

Throughput

Throughput means the total number of bits transferred (B_i) over the receiver in per unit time (T_i). The capacity of the channel depends upon the throughput. The throughput capacity of the channel is N and i represents sequence number.

$$\text{Throughput} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{B_i}{T_i}\right)$$

Average End-to-End Delay

It defines the total delay (D_i) over number of packets received by a receiver (S_i). Average end-to-end delay defines the average time taken by the packets to reach the receiver. Average end-to-end delay includes time like transmission queuing, processing, propagation delay, the lesser the E2E delay, the better is the performance of routing protocol. The average end-to-end delay with channel capacity is N and i represents sequence number.

$$\text{Average End to End delay} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i}{S_i}\right)$$

SIMULATION ENVIRONMENT

Simulation Model

In order to conduct the Performance Comparison of AODV, DSDV and DSR for different MAC protocols, the authors used the open source Network Simulator (NS2) for the simulation purpose. NS2 is a discrete network simulator that simulates the behaviour of both types of routing protocols wired and wireless networks for both single hop and multi-hop (Kaur, Kaur, & Mahajan, 2017; Chen, Yang, & Huang, 2016; Breslau et al., 2000). In this work, simulations were done by two types:

Scenario 1: In the first scenario, the simulations are based on keeping the speed of the node constant and varying the number of nodes (network size), The network size is increased from 30 to 50 to study the scalability of the routing protocols. This will give a better insight into the conduct of routing protocols for large, medium and small networks. This represents the variation in the number of patients at a particular instant within the hospital being monitored.

Scenario 2: In the second case, the simulations are carried out by keeping the number of nodes constant and varying the speed of the nodes (varying traffic), that is, if the patient is moving. The speed of the nodes is varied with pause times of 5 ms, 10 ms and 20 ms. Mobility enables us to study how the protocol performs in dynamically evolving networks.
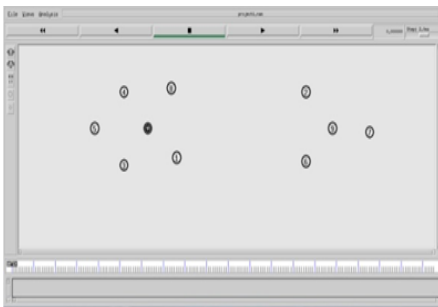
Figure 2(a). Network setup
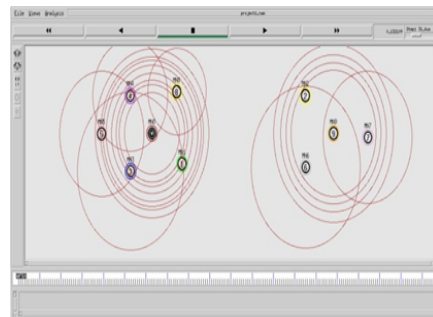
Figure 2(b). Network simulation (Scenario-1)
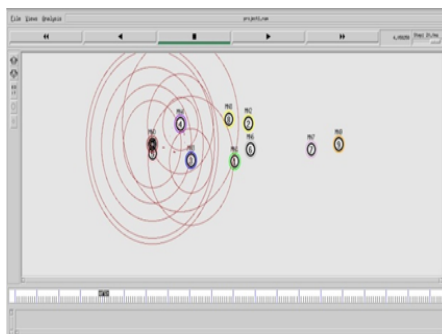
Figure 2(c). Network simulation (Scenario-2)

Figure 2(a), 2(b) and 2(c) show the setup and the simulation of the network in NS2. The figures also show the movement of nodes before and after simulation.

The Test scenarios of varying nodes and speed were applied to the three MAC protocols to evaluate their performance on varying network environment. In this work, the authors considered the simulation area 900 m x 900 m.

Simulation Parameters

For the simulation, the authors took the data to be ECG packets. An ECG packet, as shown in Figure 3, is made of the IP header (20 bytes), UDP header (12 bytes), RTP header (8 bytes) and the ECG data.
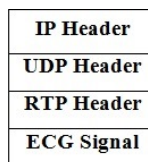
Figure 3. ECG Packet

National Institute of General Medical Sciences (NIGMS) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) have provided various ECG records ("Apnea-ECG Database", 2017); the authors considered record 'apnea ecg/a01' with some of the data values available in Table 2. The range of the ECG samples in this record varies from -0.885mV to 1.75mV. In this work, the data is taken as the ECG payload or symbols. To represent these values in a packet, a maximum of 12 bytes of data is enough.

Table 2
ECG data records

Time Elapsed (min)	ECG Values (mV)
0:00.000	-0.06
0:03.770	1.7
0:06.540	0.35
0:13.930	-0.035
0:18.990	0.045
0:26.090	-0.885
0:26.870	1.75
0:29.740	-0.025
0:35.700	-0.075
0:49.100	-0.02
0:52.960	-0.005
0:57.180	-0.135
0:59.750	-0.235
0:59.780	1.18
0:59.990	0.09

Hence, for the simulation scenario, packet size as 52 bytes was selected, its headers occupying 40 bytes and 12 bytes of ECG message. In the simulation, the authors streamed ECG signals (10,000 per node) for over an hour. The total ECG packets varied with the number of nodes active in the scenario.

Table 3
Simulation parameters

Parameter Name	Propagation	Simulator and Version	Packet Size	Antenna	Link Layer	Routing Protocols	Performance Parameters
Parameter Value	Two Ray Ground	NS2 (NS-allinone-2.35)	52 bytes (40 bytes headers; 12 bytes ECG symbol)	Omni Antenna	LL	AODV, DSDV, DSR	Packet Delivery Ratio, Throughput, End-to-End Delay

SIMULATION RESULTS AND DISCUSSION

The value of simulated parameters was obtained for two cases – for varying nodes and speed. The simulated results show that the performance of the three protocols varies with respect to changes in the number of nodes or speed of the nodes for the different MAC protocols.

This variation in nodes and speed reflects are associated with the variation in the number of patients connected and the mobility of the users in the network. These two variations mimic the real scenario in terms of usage where the patient/practitioner/staff is added or taken off from the network or they are on move. The objective is to simulate the study on these practical aspects of the usage so as to suggest a routing protocol which is best suited to transmit the data for such applications. Addressing these two issues in the system leads to saving of time in pre-operative and vital sign collection, which is the necessary and basic step for extending any medical aid to the patient.

The performance parameters for AODV, DSDV and DSR protocols with different MAC protocols are presented in the following sub-sections.

Packet Delivery Ratio

Figure 4(a) shows the Packet delivery ratio of AODV, DSDV and DSR for MAC 802.11-DCF when number of nodes are varied.
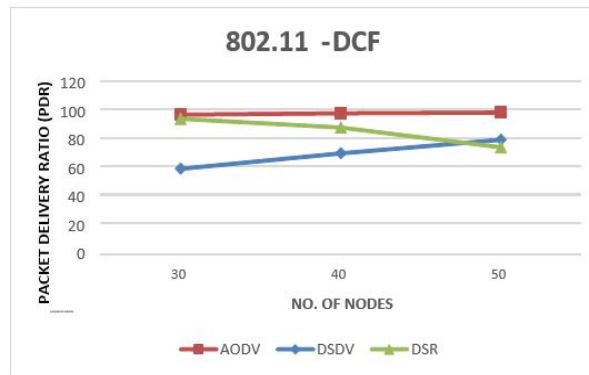
Figure 4(a). Packet delivery ratio over MAC 802.11-DCF with varying number of nodes
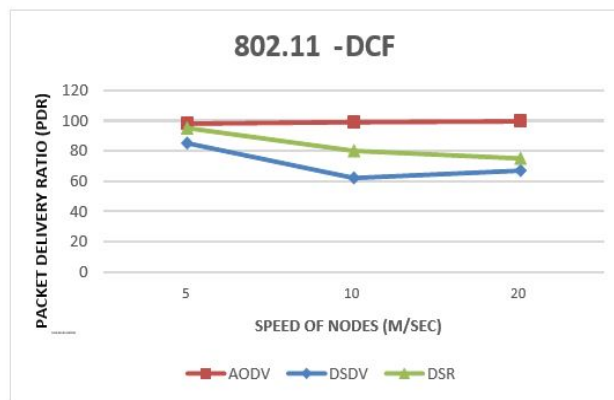
Figure 4(b). Packet delivery ratio over MAC 802.11-DCF with varying speed of nodes

Figure 4(b) shows the PDR, when the speed of the nodes is varied. AODV performs uniformly across both, increasing the speed and number of nodes and outperforms over other two protocols while DSR shows a better performance than DSDV. However, it shows a non-uniform behaviour with variations of speed and number of present nodes.
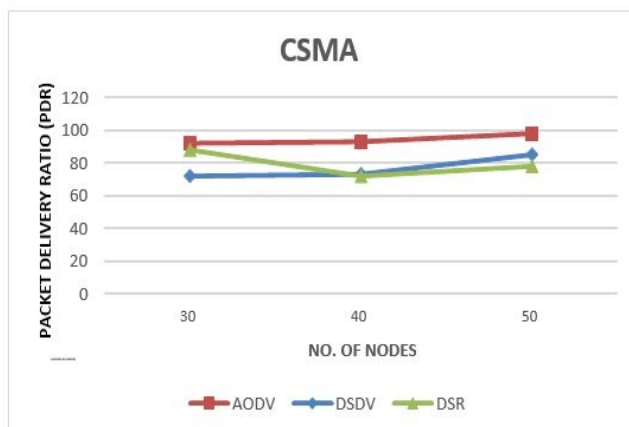
Figure 5(a). Packet delivery ratio over MAC CSMA with varying number of nodes
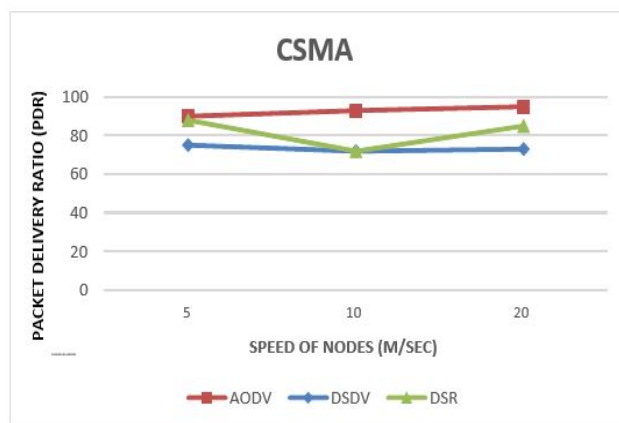
Figure 5(b). Packet delivery ratio over MAC CSMA with varying speed of nodes

Figure 5(a) shows the packet delivery ratio of DSR, AODV and DSDV for MAC CSMA when the numbers of nodes are varied and Figure 5(b) shows the same, when the speed of the nodes is varied. Similar to the case of 802.11-DCF, when run over CSMA, AODV shows a higher value of Packet Delivery Ratio (PDR) constantly than DSDV and DSR.

Figure 6(a) shows the result of PDR of DSR, AODV and DSDV when ran over MAC CSMA-CA when the number of nodes are varied and Figure 6(b) is the PDR performance of the routing protocols when the speed of the nodes is varied. AODV, similar to the other two MAC shows a uniformly good performance on CSMA/CA as well for varying nodes and speed. However, here DSR also shows a much better performance than other MAC protocols and clearly outperforms DSDV and even AODV for certain scenarios.
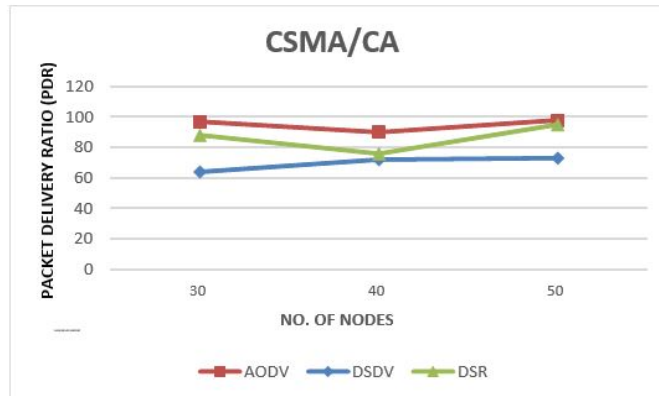
Figure 6(a). Packet delivery ratio over MAC CSMA/CA with varying number of nodes
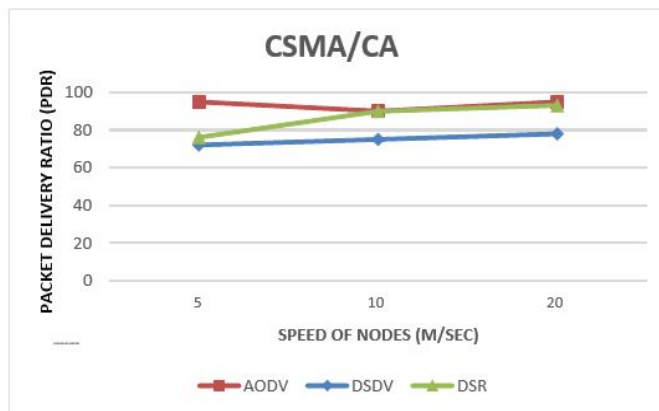
Figure 6(b). Packet delivery ratio over MAC CSMA/CA with varying speed of nodes

Throughput

From Figure 7(a) and 7(b), it is noticed that the throughput of AODV is significantly higher and reliable compared to DSR and DSDV. However, throughput of AODV almost remains constant except that it is decreased, when either the number of nodes or speed of nodes is increased to high values.
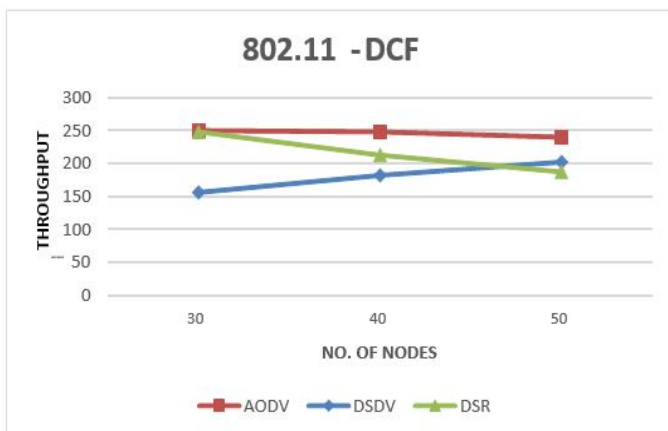
Figure 7(a). Throughput over MAC 802.11-DCF with varying number of nodes
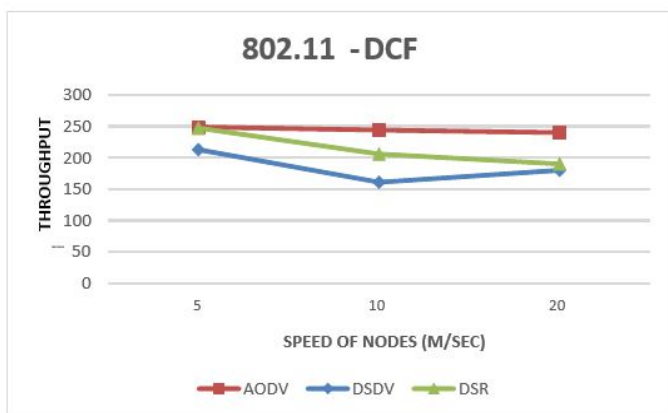
Figure 7(b). Throughput over MAC 802.11-DCF with varying speed of nodes

The throughput of DSR demonstrates a performance, although not as good as AODV but better than DSDV over 802.11-DCF. The throughput of DSDV increases with increase in the number of nodes. Further, it lowers with increase in the speed of nodes. It is observed that AODV has better throughput performance up to even 77.09% than DSDV in the case of varying network size.

In case of CSMA as observed from Figure 8(a) and Figure 8(b) and CSMA/CA as seen in Figure 9(a) and Figure 9(b), the performance of the protocols somewhat remains the same as the PDR performance with AODV outperforming others and with DSR performing better than DSDV.
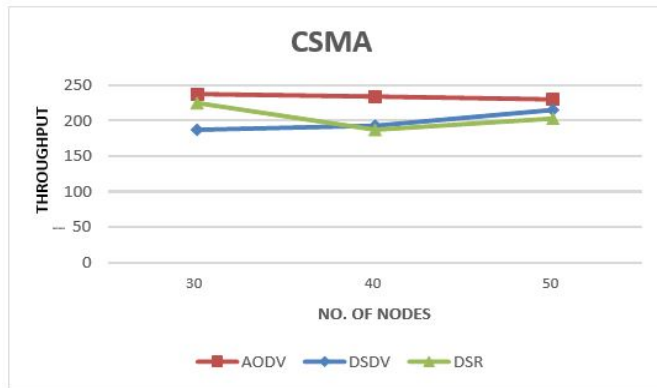
Figure 8(a). Throughput over MAC CSMA with varying number of nodes
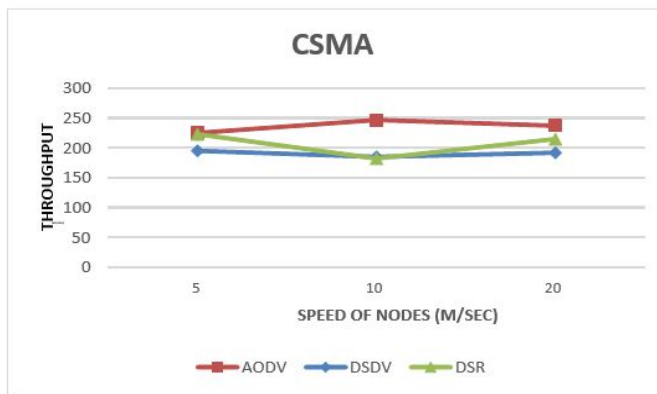
Figure 8(b). Throughput over MAC CSMA with varying speed of nodes
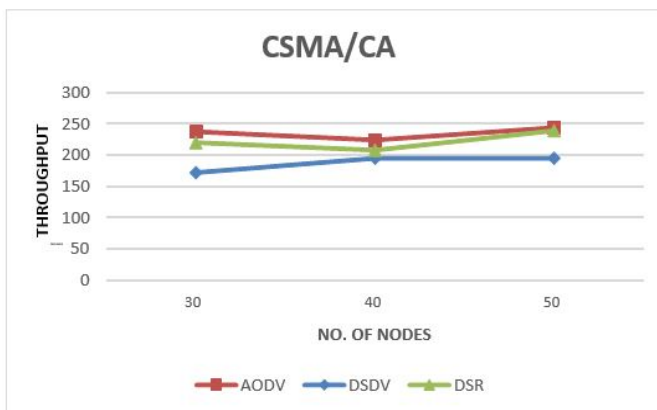
Figure 9(a). Throughput over MAC CSMA/CA with varying number of nodes
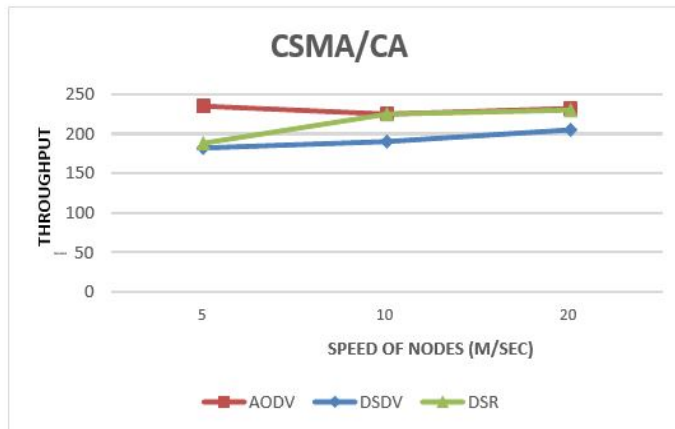
Figure 9(b). Throughput over MAC CSMA/CA with varying speed of nodes

Based on the above, it reflects that AODV is suitable in emergency cases, showing better throughput performance than DSR and DSDV.

End-to-End Delay

End-to-End delay of AODV and DSDV is calculated as shown in Figure 10(a), when the number of nodes is varied. Figure 10(b) represents the calculation of end-to-end delay of AODV and DSDV, when the speed of the nodes is varied over MAC 802.11-DCF. Due to the provision of maintaining a route only when its activity level is high, AODV suffers from the least end-to-end delay than DSR and DSDV. In case of MAC 802.11-DCF, DSR performs much better than DSDV by establishing the source-destination route much more efficiently.
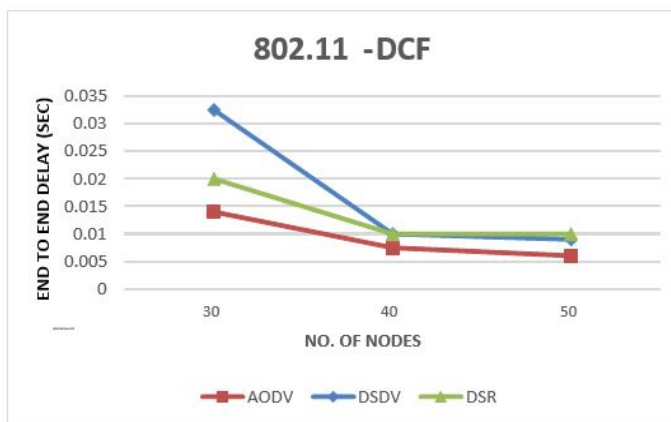
Figure 10(a). End-to-End Delay over MAC 802.11-DCF with varying number of nodes
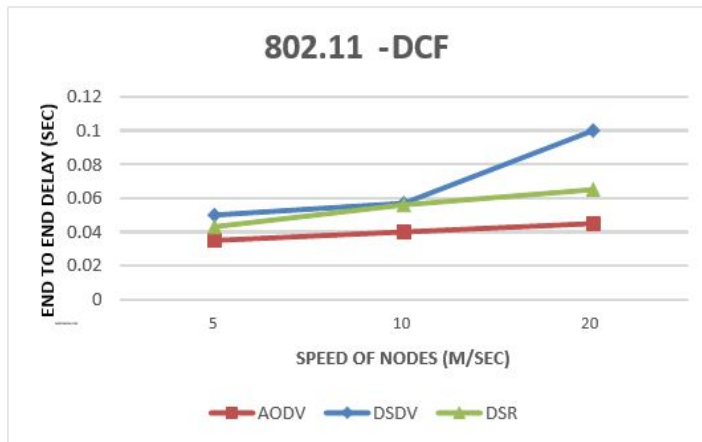
Figure 10(b). End-to-End Delay over MAC 802.11-DCF with varying speed of nodes

Figure 11(a) and 11(b) are the End-to-End Delay performance of the routing protocols over MAC CSMA with varying number of nodes and speed of nodes respectively. In both the cases, the performance of AODV is recorded as much better than DSDV and DSR with DSR depicting an inconsistent behaviour with changing cases.
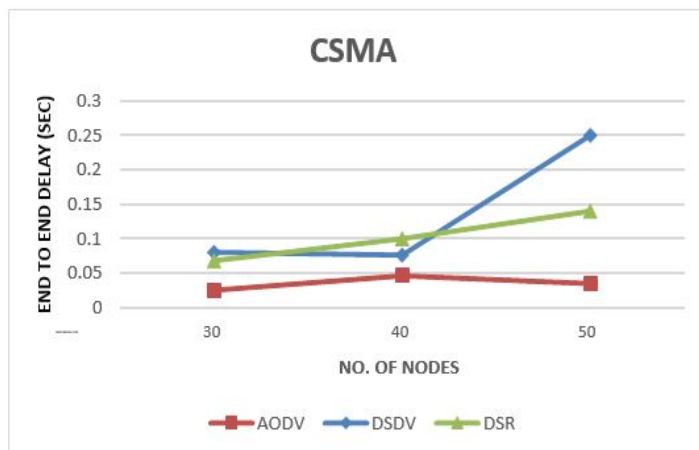
Figure 11(a). End-to-End Delay over MAC CSMA with varying number of nodes
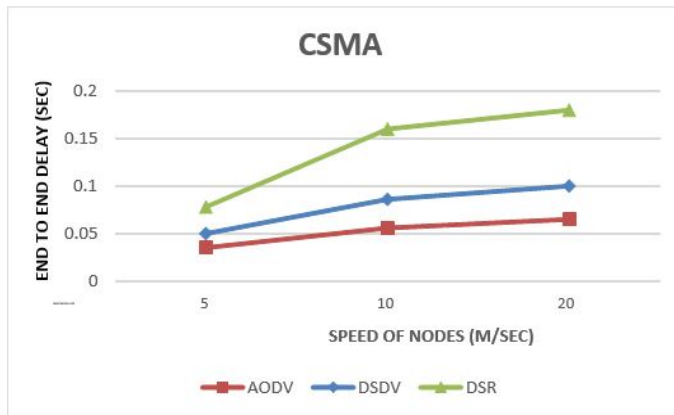
Figure 11(b). End-to-End Delay over MAC CSMA with varying speed of nodes

Meanwhile, when ran over MAC CSMA/CA as in Figure 12(a) and Figure 12(b), DSR outperforms the other two protocols with AODV as the second-best choice.
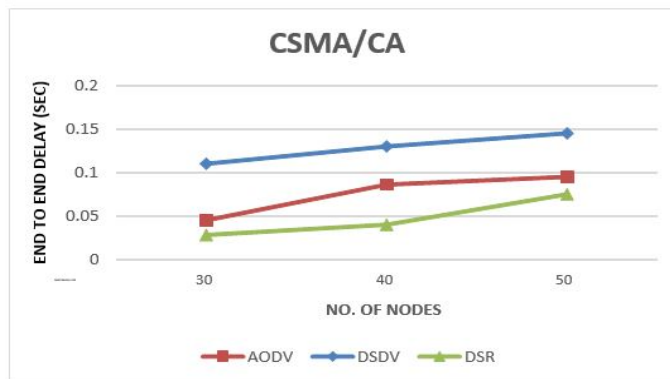
Figure 12(a). End-to-End Delay over MAC CSMA/CA with varying number of nodes
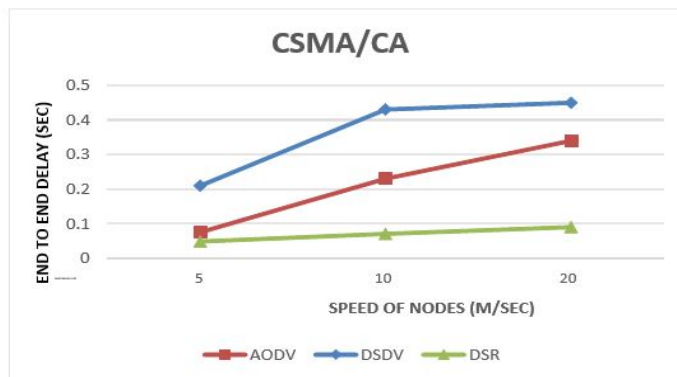
Figure 12(b). End-to-End Delay over MAC CSMA/CA with varying speed of nodes

Table 4
Parameter values using 802.11 DCF

Parameter Values by Varying Number of Nodes				Parameter Values by Varying Speed of Nodes			
Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)	Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)
AODV				AODV			
30 Nodes	100	250	0.014	5mps	99	250	0.035
40 Nodes	102	250	0.0075	10mps	101	246	0.04
50 Nodes	100	242	0.006	20mps	100	242	0.0425
DSDV				DSDV			
30 Nodes	60	155	0.0325	5mps	82	214	0.05
40 Nodes	72	181	0.01	10mps	60	157	0.06
50 Nodes	80	203	0.008	20mps	67	180	0.1
DSR				DSR			
30 Nodes	97	250	0.02	5mps	96	247	0.0435
40 Nodes	91	211	0.01	10mps	80	208	0.06
50 Nodes	77	182	0.009	20mps	76	192	0.062

Table 5
Parameter values using CSMA

Parameter Values by Varying Number of Nodes				Parameter Values by Varying Speed of Nodes			
Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)	Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)
AODV				AODV			
30 Nodes	94	234	0.03	5mps	90	225	0.04
40 Nodes	95	234	0.05	10mps	92	251	0.05
50 Nodes	99	227	0.042	20mps	97	242	0.063
DSDV				DSDV			
30 Nodes	70	181	0.084	5mps	78	248	0.05
40 Nodes	72	193	0.075	10mps	76	239	0.08
50 Nodes	85	214	0.25	20mps	77	248	0.1
DSR				DSR			
30 Nodes	85	223	0.064	5mps	88	224	0.075
40 Nodes	72	188	0.1	10mps	72	237	0.17
50 Nodes	79	200	0.14	20mps	83	216	0.18

Table 6
Parameter values using CSMA/CA

Parameter Values by Varying Number of Nodes				Parameter Values by Varying Speed of Nodes			
Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)	Parameter Measured	Packet Delivery Ratio	Throughput	End-to-End Delay (sec)
AODV				AODV			
30 Nodes	98	242	0.048	5mps	97	241	0.08
40 Nodes	92	224	0.08	10mps	90	225	0.22
50 Nodes	101	247	0.094	20mps	93	228	0.32
DSDV				DSDV			
30 Nodes	64	172	0.116	5mps	73	181	0.205
40 Nodes	76	192	0.182	10mps	75	189	0.424
50 Nodes	76	196	0.148	20mps	79	206	0.45
DSR				DSR			
30 Nodes	86	223	0.03	5mps	76	183	0.052
40 Nodes	78	202	0.038	10mps	90	225	0.08
50 Nodes	99	243	0.077	20mps	93	227	0.097

Table 4, 5, and 6 sum up the parameter values obtained in the previous section using 802.11-DCF, CSMA and CSMA/CA.

Thus, AODV and DSR are two best choices when End-to-End performances of the routing protocols are considered over three different MAC protocols with DSR giving an exceptional behaviour on CSMA/CA and AODV when ran on 802.11-DCF and CSMA.

CONCLUSION

The performance investigation study of three routing protocols, AODV, DSDV and DSR, over the various types of MAC protocols, namely 802.11-DCF, CSMA and CSMA/CA was carried out. The potential of Wireless Local Area Network in telemedicine applications was discussed based on QoS parameters of the three Ad Hoc WLAN protocols on the basis of their end to end delay, throughput and packet delivery ratio behaviour. To demonstrate the efficacy of these protocols, an ECG data packet is taken as the sample data to be transferred. The performance evaluation of AODV, DSDV and DSR was carried out by varying the network size along with mobility of user (node) to mimic an actual hospital scenario, where the practitioners and patients are supposed to be mobile and their numbers are variable. The results show that over IEEE 802.11- DCF and MAC CSMA, AODV exhibit almost constant behaviour and a better performance than the other two protocols in terms of throughput when the number of nodes and mobility rates are varied. However, DSR shows better performance in End-to-End delay and packet delivery ratio. It is important to note that in telemedicine application minimisation of patient delay and packet loss is an important aspect for effective and successful transmission of medical data.

Hence, it can be concluded and suggested that AODV and DSR should be the preferred protocols as they show better performances than DSDV routing protocol for the proposed applications. However, further studies with large number QoS parameters and high level of scalability ratio will be useful with other MAC protocols.

REFERENCES

Abolhasan, M., Wysocki, T., & Dutkiewicz, E. (2004). A review of routing protocols for mobile ad hoc networks. *Adhoc Networks, 2*(1), 1-22.

Apnea-ECG Database.(2017). *PhysioBank's automated teller machine.* Retrieved from http://www. physionet.org/cgi-bin/atm/ATM.

Basavaraju, T. G., Sarkar, S. K., & Puttamadappa, C. (2006). Impact of Mac layer on the performance of routing protocols in mobile ad hoc networks. *International Journal of Information Technology, 3*(4), 237-244.

Batistatos, M. C., Tsoulos, G. V., & Athanasiadou, G. E. (2012). Mobile telemedicine for moving vehicle scenarios: Wireless technology options and challenges. *Journal of Network and Computer Applications, 35*(3), 1140-1150.

Breslau, L., Estrin, D., Fall, K., Floyd, S., Heidemann, J., Helmy, A., ... & Yu, H. (2000). Advances in network simulation. *IEEE Computer Journals and Magzines, 33*(5), 59-67.

Chen, L., Yang, R., & Huang, M. (2016). Ad hoc high-dynamic routing protocol simulation and research. In Q. A. Zeng (Ed.), *Wireless communications, networking and applications* (pp. 399-408). New Delhi, India: Springer.

Cicalo, S., Mazzotti, M., Moretti, S., Tralli, V., & Chiani, M. (2016). Multiple video delivery in m-health emergency applications. *IEEE Transactions on Multimedia, 18*(10), 1988-2001.

Corson, M. S., & Ephremides, A. (1995). A distributed routing algorithm for mobile wireless networks. *ACM/Baltzer Wireless Networks Journal, 1*(1), 61-81.

Corson, S., & Macker, J. (1999). *Mobile Ad Hoc Networking (MANET): Routing protocol performance issues and evaluation considerations.* Retrieved from http://www.ietf.org/rfc/rfc2501.txt

Ford, L. R., & Fulkerson, D. R. (1962). *Flows in Networks* (1st Ed.). New Jersey, USA: Princeton University Press.

Johnson, D. B., & Maltz, D. A. (1996). Dynamic source routing in ad hoc wireless networks. In *T. Imielinski, & H. F. Korth (Eds.), Mobile computing* (pp. 153-181). Boston, USA: Springer.

Kaur, P., Kaur, D., & Mahajan, R. (2017). Simulation based comparative study of routing protocols under wormhole attack in Manet. *Wireless Personal Communications, 96*(1), 47-63.

Kim, Y. D., Moon, I. Y., & Cho, S. J. (2009). A comparison of improved AODV routing protocol based on IEEE 802.11 and IEEE 802.15.4. *Journal of Engineering Science and Technology. 4*(2),132-141.

Koutsakis, P. (2017). Scheduling for telemedicine traffic transmission over WLANs. *Computer Communications, 108*, 17-26.

Liu, S., Yang, Y., & Wang, W. (2013). Research of AODV routing protocol for Ad Hoc. *AASRI Procedia, 5*, 21-31.

Mbarushimana, C., & Shahrabi, A. (2007). Comparative study of reactive and proactive routing protocols performance in mobile ad hoc networks. In *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops* (Vol. 2, pp. 679-684). Washington, DC, USA: ACM.

Perkins, C. E., & Watson, T. J. (1994). Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers. In *Proceedings of the Conference on Communications Architectures, Protocols and Applications, SIGCOMM, 1994* (pp. 234-244). London, United Kingdom: ACM.

Royer, E., & Toh, C. (1999). A review of current routing protocols for Ad Hoc mobile wireless networks. *IEEE Personal Communications, 6*(2), 46–55.

Standards ISO/IEEE. (2008). *Health informatics PoC medical device communication, Part 00101: guide-guidelines for the use of RF wireless technology.* Piscataway: NJ, USA.

Subramanya, B. M., Shwetha, D., & Devaraju, J. T. (2012). A performance study of proactive, reactive and hybrid routing protocols using Qualnet Simulator. *International Journal of Computer Applications, 28*(5), 10-17.

Tamilselvan, L., & Sankaranarayanan, V. (2007). Prevention of black hole attack in MANET. In *Proceedings of 2nd International Conference on Wireless Broadband and Ultra-Wideband Communications, AusWireless 2007* (pp. 21-27). Sydney, NSW, Australia: IEEE.

Toh, C. K. (2002). *Ad hoc mobile wireless networks: Protocols and systems* (2nd Ed.). New Jersey, USA: Prentice Hall Inc.

Vetrivelan, N., & Reddy A. V. (2008). Performance analysis of three routing protocols for varying MANET size. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists IMECS 2008* (Vol.2, pp. 1178-1182). Hong Kong, China: Newswood Limited.

Yadav, P., Agrawal, R., & Kashish, K. (2018). Performance evaluation of ad hoc wireless local area network in telemedicine applications. *Procedia Computer Science, 125*, 267-274.

Youssef, M. A., Vasan, A., & Miller, R. E. (2002). Specification and analysis of the DCF and PSF protocols in the 802.11 standard using systems of communicating machines. In *Proceedings of the 10th IEEE International Conference on Network Protocols ICNP'02* (pp. 132-141). Paris, France: IEEE.

Detection of Spam Using Particle Swarm Optimisation in Feature Selection

Surender Singh[1,2]* and Ashutosh Kumar Singh[2]

[1]*Department of Information Technology, Maharaja Surajmal Institute of Technology, New Delhi, India*
[2]*Department of Computer Applications, National Institute of Technology, Kurukshetra, Haryana, India*

ABSTRACT

Spamming is a major issue in the area of web search. There are many features (Link & Content based) which are used for spam and non-spam classification. This paper recommends CFS+PSO, which takes the advantages of swarm behaviour (uses randomness and global communication between particles) and Correlation Based Feature Selection Technique (CFS). The objective of feature selection is to build logical model with improved performance in time and accuracy. The performance of CFS+PSO is evaluated on WEBSPAM-UK2006 with Multilayer Perceptron (MLP), Naïve Bayes, Support Vector Machine (SVM), J48 & AdaBoost. Experimental results show great decline in existing features and computational time while increases in the accuracy measures (F_1 Score and AUC).

Keywords: Content and link based features, correlation based feature selection, data mining, filter and wrapper model, particle swarm optimization, spam

INTRODUCTION

Web spamming (which is known as spamdexing) is recognised as one of the main problems of search engines (Gyongyi & Garcia-Molina, 2005). Nowadays, information

retrieval (IR) is a main concern of search engine industries. Spam not only corrupts the search quality but along with it, weakens the trust of users in a particular search engine and leads to phishing (Webber, Maria de Fátima, & Hepp, 2012). The manipulation can be done in different forms like adding content spam and link spam. Content spam is a common area for spammers because search engines (like Google & Yahoo!) use models which are based on the rank and content of websites. Primarily based on the web document structure, content spamming is subdivided into five categories, namely, title, body, meta-tags, anchor text, and URL

spamming (Gyongyi & Garcia-Molina, 2005). In link spamming, the spammer creates a large number of links to a page just to increase the link based rank.

Many conventional algorithms are developed for spam detection but infeasible due to the dynamic growth of the Web (Becchetti, Castillo, Donato, Baeza-Yates, & Leonardi, 2008). However, machine learning (ML) algorithms give better results due to their ability to study the necessary patterns (Goh & Singh, 2015). Along with ML algorithms, feature selection (FS) shows a crucial part in the success of spam detection. Most machine learning strategies or algorithms degrade in execution when executed with most features that are not essential or repetitive for anticipating the desired results (Li, Li, & Liu, 2017).

In a broad way, FS is divided into the following: filter technique, wrapper technique and hybrid technique (Yu & Liu, 2004). Statistical analysis is required for features without any ML structure while the ML model is assumed in wrapper method to confirm (validate) the learning performance of the particular model (Guyon & Elisseeff, 2003; Dash & Liu, 1997). The hybrid model or technique takes the strengths of both models (Huang, Cai, & Xu, 2007). The representation of all three models or techniques are given in Figure 1.
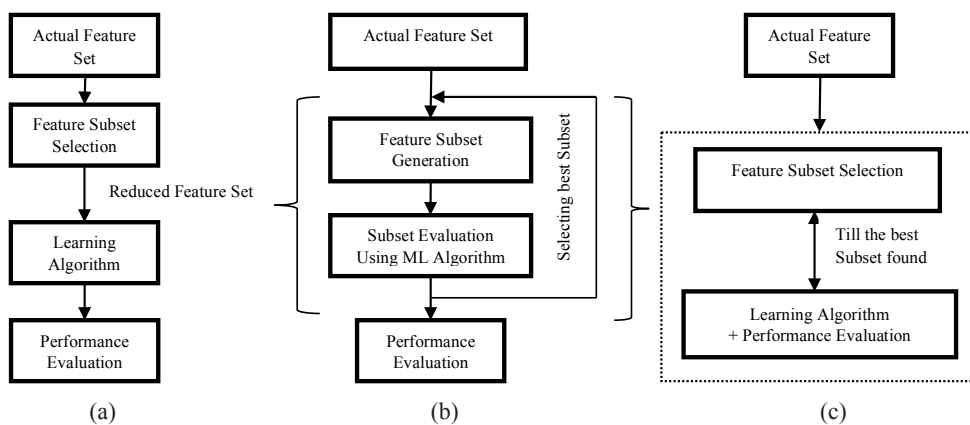
Figure 1. (a) Filter Technique   (b) Wrapper Technique   (c) Hybrid Technique

The best (optimum) features can be created and the search process is carried out in some ways, like SFS (sequential forward search), SBS (sequential backward search) and Bidirectional search. In SFS (Guan, Liu, & Qi, 2004; Reunanen, 2003), the search process starts with a blank set and adds features successfully but SBS search starts with a complete set and then eliminates features (Gasca, Sanchez, & Alonso, 2006; Hsu, Huang, & Dietrich, 2002). Bidirectional selection search starts with both sides and increases and eliminates features at the same time (Caruana & Freitag, 1994). There is an alternative search called complete search but is not possible with huge number of features (like in web spam area). According to Pudil, Novovicova and Kittles (1994), an additional search algorithm, known as floating search is proposed because of nesting effect in SFS or SBS.

Many of the above search techniques (or strategies) use local instead of global search (Kabir, Shahjahan, & Murase, 2012). These algorithms suffer from computational complexity due to partial search over the features set (space). Thus, most of the research is carried out on metaheuristics (nature inspired) algorithms (Ke, Feng, & Ren, 2008). The two primary segments of metaheuristic algorithms are determination of best solutions and randomisation. The former affirms that solutions will join to optimality while the latter keeps away the solution being stuck at local optima and increase the differing potentials of solutions. To accomplish these objectives, researchers have tried many metaheuristic methods including Firefly algorithms (Yang & He, 2013), Harmony algorithms (Ramos, Souza, Chiachia, Falcao, & Papa, 2011), Bee algorithms (Karaboga & Basturk, 2007), GA (Yang & Honavar, 1998), Ant Colony Optimisation (Kabir, Shahjahan, & Murase, 2009), Simulated Annealing (Filippone, Masulli, & Rovetta, 2006) and Wolf-Search algorithms (Song, Fong, & Tang, 2016; Tang, Fong, Yang, & Deb, 2012) in solving feature subset selection in various problems.

In this work, a CFS+PSO technique is suggested, which is combination of correlation based feature selection approach and particle swarm optimisation methodology that uses a hybrid search approach in feature space of web-spam area. The key emphasis of this technique is to create subsets of significant features of lesser size. This method exploits swarm intelligence in search strategy which combines with filter technique. The objective is to search the global best solution among the current best. Also, hybrid techniques are proficient in discovering a better answer, when a single method is frequently restricted with insufficient solution.

The rest of the paper is sequenced in five segments. Related work regarding feature selection and search methods are explored in segment 2. Segment 3 explains the CFS+PSO technique. Experimental and parameters setting are discussed in segment 4. The outcome results are displayed in segment 5. Finally, the conclusion with future scope is described in segment 6.

RELATED WORK

Basically, FS is a method of rejecting the irrelevant features to enhance the performance (time & accuracy) of ML algorithms. FS can be categorised mainly in two types: feature subset determination and feature ranking based on how the features are joined for assessment. Space and computational complexity is high in feature subset selection approach because it evaluates the individual feature subset with a feature selection metric (correlation or consistency) using any one of the searching techniques (Bolón-Canedo, Sánchez-Maroño, & Alonso-Betanzos, 2013). In feature ranking, each feature is ranked using selection metric such as chi-square feature evaluation, information gain, gain ratio attribute assessment and symmetric uncertainty. Then, top ranking attributes or features are selected as significant features by some threshold value. Space and time complexity is less compared to subset selection for this approach. Furthermore, FS algorithms can be divided in three classes: filter, wrapper and hybrid. Filter is done by ranking of features and determination of feature subset while wrapper creates subsets by use of any search technique, then assesses these sets using machine learning classifiers (Dash & Liu, 1997). Run time and search overhead is increased in comparison with filter technique.

Neural Network (NN), Support Vector Machines (SVM) and Boosting Algorithm can be used as a classification function in wrapper or filter method (Breiman, 1996, 2001; Cortes & Vapnik, 1995; Haykin & Lippmann, 1994; Quinlin, 1993).

In actual fact, filter methods are fast and easy to implement because of the assessment of features without any model presumed between outputs and inputs of the data. Chow and Huang (2005) present an idea of mutual information. PCA method is implemented to remove redundancy between the segments of high-dimensional vector data which empowers a lower-dimensional data without main loss of information (Kambhatla & Leen, 1997). Abdulla and Kasabov (2003) designed a model where features are reduced by refining the dominance effects.

The new FS technique is based on chi-square statistical measure (CHIR) given by Li, Luo and Chung (2008). Song, Ni and Wang (2013) propose a clustering technique in which features are separated into clusters, then highly illustrative feature that is intensely connected to objective classes is selected from clusters to build a new subset of features. Sotoca and Pla (2010) have applied hierarchical clustering technique.

Searching also shows a very critical role in finding of significant features from a given dataset for any feature selection method. Different types of sequential or metaheuristic algorithms are proposed for the searching problem. Sequential search is applied by many researchers (Gasca et al., 2006; Guan et al., 2004; & Hsu et al., 2002). Uğuz (2011) proposed Information Gain (IG) and GA & PCA (feature selection and extraction methods). Shunmugapriya and Kanmani (2017) applied a hybrid technique, which takes the benefits of ACO as well as Artificial Bee Colony (ABC) techniques to optimise FS. Some bio-inspired metaheuristic algorithms are also invented which includes Firefly (Yang, 2009), Cuckoos (Yang & Deb, 2009) and Bats (Yang, 2010).

PROPOSED METHODOLOGY

Feature selection method includes four steps, which are defined as: (a) subset generation - selection of an initial point (feature subset) because it can influence the search direction, (b) evaluation function - this step assesses the subset produced in the previous step by using filter or wrapper approach, since previous approach is autonomous of the induction algorithm while wrapper techniques use induction algorithm for assessing the weight of highlight subsets, (c) stopping criteria - a stopping point must be chosen because dependent upon the valuation policy, a feature selector may leave including (or expelling) features (elements) when the quality value of a present feature subset is not increasing, and (d) validation methodology - validation technique is to check whether the feature subset choice is substantial or not. Usually the result of the original feature set is compared with the feature set chosen in the previous step as input to some induction algorithm utilising datasets (Dash & Liu, 1997).

Feature Selection (CFS)

CFS algorithm selects features depending on correlation based heuristic assessment function (Hall, 1999). The preference of the assessment function is towards subsets whose features are extremely related with the class but independent of other features. Unrelated features are disregarded on the fact that they show less association with the class while other features are

separated out as they will be very much linked with at least some of the features. CFS evaluation (or assessment) function (Senliol, Gulgezen, Yu, & Cataltepe, 2008) is as:

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{1}$$

where M_s is the heuristic "function" of a feature subset S containing k features, $\overline{r_{cf}}$ is the mean feature-class correlation ($f \in S$), and $\overline{r_{ff}}$ is the average of feature-feature intercorrelation. $k\overline{r_{cf}}$ denotes predictiveness of the class with a set of features where $\sqrt{k + k(k-1)\overline{r_{ff}}}$ shows redundancy between the features. The above condition is the fundamental of CFS and the set of features with the highest value found during the process is utilised to reduce the size of both the training and testing set. Figure 2 demonstrates the steps of the CFS algorithms.

```
Input:
  S – Data set with features F, |F| = total no of features
  M₁ – Heuristic merit to be maximized
  SG – Successor Generation operator
Output:
  P = Starting Point (F);
  Solution = Best of P according to M₁;
Repeat
  P = Search_Strategy (P, SG (M₁), F);
  F'= Best of P according to M₁;
  If M₁ (F') > M₁ (Solution) or (M₁ (F') = M₁ (Solution) and |F'| < |Solution|) then Solution = F';
Until Stop (M₁, P).
```
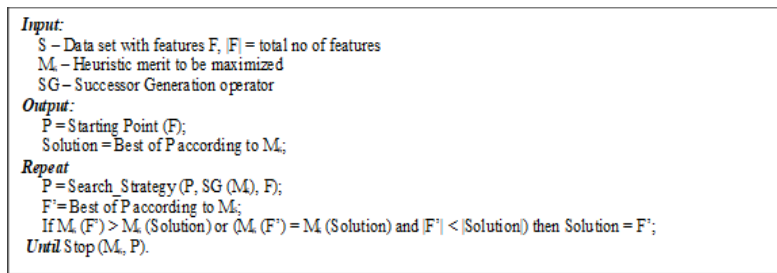
Figure 2. Steps involved in CFS Algorithm

Optimisation Strategy

Optimisation is dominant in many applications. The main idea behind optimisation is to minimise the cost (computation, time and resources) and to maximise the performance and proficiency. Due to the various constraints in real world applications, we have to find optimal solutions. PSO was developed in 1995, based on the swarm nature to monitor the particles for searching global best solutions. However, PSO has many resemblances with GA and Virtual Ant Algorithms but instead of using crossover (or mutation), it takes advantage of global communication between the particles.

PSO searches for space of an objective function by adjusting the paths of individual particles. The movement of particles can be described by two main segments: stochastic and deterministic. The particles are involved towards the position of gb^* (global best) and their best location (x_i^*), while in the interim it tends to move arbitrarily. When the particles find a position which is superior to any earlier found position, it updates it as the new current best for that particle. The main objective is to search the globally best solution between all the existing best solutions until there is no more improvement. The main steps of PSO is summarised in algorithm shown in Figure 3. Each particle has a position in the search area, which is denoted by $x_i = (x_1, x_2, x_3, x_4 \dots x_n)$. Particles scan or move in search space for the best solutions. Every particle has a velocity, which is denoted as $v_i = (v_1, v_2, v_3, v_4 \dots v_n)$. During this movement, each particle refreshes (updates) its velocity and position according to its own and its neighbour's experience (Eq. 2 & Eq. 3).

$$v_i^{t+1} = w * v_i^t + c_1 * r_1 * (gb^* - x_i^t) + c_2 * r_2 * (x_i^* - x_i^t) \qquad (2)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \qquad (3)$$

Here t signifies the count of iteration and w is inertia weight in the optimisation process, which is for controlling the influence of the previous velocities on the current velocity. The c_1 and c_2 are learning parameters (or acceleration constants) and r_1 and r_2 are random values distributed in between 0 and 1. The algorithm ends when a predetermined condition is achieved, which can be a good fitness value or a maximum number of iterations (Xue, Zhang, & Browne, 2013).


```
Start
    Objective function F(x) where x= x1, x2, x3 ...........xD
    Initialze locations xi and velocity vi of n particles.
    Initial minimum fmin^t=0= min {f(x1), f(x2), .... f(xn)} where t=0
    While (stopping criterion)
    t= t+1 (counter)
        For (loop over n particles & all D dimensions)
        Generate new velocity vi^t+1 using vi^t+1 = w * vi^d + c1 * r1 * (gb^* - xi^t) + c2 * r2 * (xi^* - xi^t)
        Calculate new locations xi^t+1 = xi^t + vi^t+1
        Evaluate objective functions at new locations x
        Find the current minimum fmin^t+1
        End for
        Find the current best xi^* & current global best gb^*
    End while
    Output the results xi^* & gd^*
End
```

Figure 3. Algorithm for Particle Swarm Optimisation

CFS+PSO Algorithm

In this segment, CFS+PSO algorithm is recommended, which is used to assess the importance and the dismissal of the selected feature subset. CFS+PSO utilises correlation based feature technique to form the fitness functions and assessment of integrity of the reduced feature subset. For a feature subset X with m features, $X = (x_1, x_2, x_3, x_4 \ldots \ldots x_m)$, CFS assesses mean of association between feature-class and average of intercorrelation between feature-feature to decrease the classification error or increase accuracy (by using Eq.1). The subset of independently good features may not be the best combination because of redundancy between features.
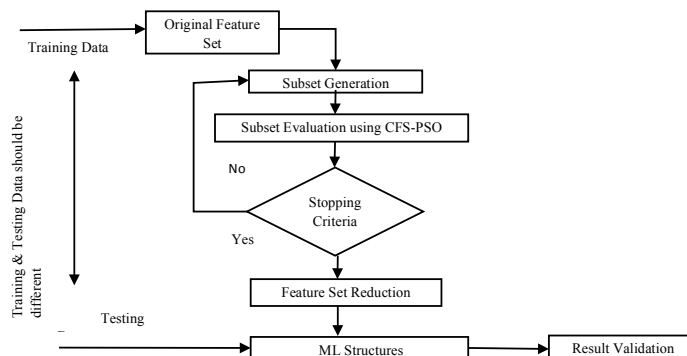
Figure 4. Correlation based feature selection method with PSO

The removal of redundant features can increase performance due to the reduction of the dimensionality. In PSO, every solution of the problem is denoted by a particle, which can be represented by array (or vector). Particles move in search space to search for the best solutions and during this movement, every particle can remember its best experience with its neighbours. So, all particles search for the optimal (best) answer by updating the position of each particle, based on its best experience and its nearby particles. Steps involved in CFS+PSO algorithm is shown in Figure 5.

```
Input: Training Data Set and Test Data Set;
Output: gbest (selected feature subset); // global best

Start
    Initialize the xi & vi; // position & velocity of each particle
    While (till stopping is not reached)
    Evaluate the fitness of each particle // using eq. (1)
    For (loop over n particles)
    Update the ith particle best position; // pbest
    Update the global best position of ith particle; // gbest
    For (loop over n particles)
        For (loop over dimension) // total no. of possible direction
        Update the position of ith particle; // using eq. (3)
        Update the velocity of ith particle; // using eq. (2)
    Calculate the classification accuracy of the selected feature subset on test set;
    Return the selected feature subset; // gbest
End
```

Figure 5. Algorithm for correlation based feature selection with PSO

BENCHMARK AND PARAMETER SETTING

In this paper we use publicly available benchmark WEBSPAM UK-2006, which consists of 777, 410, 46 pages in 11,402 hosts in the UK domain ("WEBSPAM UK-2006", 2006). Features are distributed as content and link-based. In Table 1, A denotes the content-based features (Ntoulas, Najork, Manasse, & Fetterly, 2006). There are a total of 96 full content-based features which are denoted by B (Castillo, Donato, Gionis, Murdock, & Silvestri, 2007). Label C represents the link-based features. The transformed link-based features are designated by set D. This conversion works better for classification than the raw link-based features. More detailing on these features is described by (Becchetti, Castillo, Donato, Leonardi, & Baeza-Yates, 2006). For experimental purpose, the authors combined some feature sets, like A+C and B+D (Singh & Singh, 2018; Goh & Singh, 2015; Goh, Singh, & Lim, 2013; Singh, Kumar, & Leng, 2011;).

Table 1
Distribution of features in dataset (WEBSPAM UK-2006)

Total Features	Name of feature Set
24	Content Based Features (A)
96	Full Content Based Features (B)
41	Link Based Features (C)
138	Transformed Link Based Features (D)
65	Content + Link Based Features (A+C)
234	Full Content + Transformed Link Based Features (B+D)

The classification (training and testing set) in spam database is presented in Table 2.

Table 2
Classification of Spam and Ham in Benchmark

	Benchmark	
	Training Set (Set 1)	Testing Set (Set 2)
Spam	553	1250
Ham (Non-Spam)	3810	601
Total	4363	1851

For the classification, the following algorithms were used: Naïve-Bayes, SVM, J48, MLP and AdaBoost. The Naive Bayes is a basic version of Bayes formula with strong independent assumptions between features and concludes which class a unique instance belongs to. J48 is the implementation of algorithm C4.5, and its predecessor, that summarises training data in the form of a decision tree. Random forest builds multiple decision trees and outputs the class which is mean prediction of the individual trees. SVM classifier gives high-dimension features using hyperplanes which provide the largest minimum distance to divide data points between classes. MLP is a feedforward neural network model that maps the weighted inputs to the output of each neuron using multi-weights connections. AdaBoost also called Adaptive Boosting is used for building strong classifiers with linear combination of weak classifiers.

Evaluation Criteria

After construction of the classifier, it must be assessed for accurateness. Effective estimation is also significant because without knowing the expected accuracy, it cannot be used in real-world problems. Confusion matrix for binary classification (spam or non-spam) was used. The main measure is classification accuracy which is totally correct prediction divided by total cases in data set. But here, the authors took two other evaluation criteria - area under the ROC (AUC) and F_1 score (F-measure) because it cannot be said that any one of the classifiers is strictly better than the other. ROC is 2-dimensional graphs where FP rate (FPR) and TP rate is plotted on X and Y axis respectively. It represents trade-off between costs (FP) and benefits (TP). F_1 score is the harmonic mean of precision and recall (Eq. 4). Precision is correct positive cases divided by total positive predicted cases whereas recall is termed as the count of correct positive predictions divided by the total count of positives cases.

$$F - measure = 2.\frac{precesion * recall}{precesion + recall} \qquad (4)$$

Parameter Setting

Parameter selection is one of the most essential parts of any algorithm. The parameters in PSO are taken according to the settings suggested by Clerc and Kennedy (2002). The detailed

settings are shown as follows: $c1 = c2 = 1.496$, $w = 0.7298$. Stopping criteria is taken as 30 (maximum iteration) in this experiment. Swarm (particles) size also affects performance of PSO because with few particles, it tends to confine with local maxima while the use of too many particles degrades (slows down) the algorithm. So, the authors took the swarm size as half of the total features used. While using NB, J48 and AdaBoost classifiers default parameters were considered. In SVM, radial basis function was taken because it gives value to each point based on its distance from the origin or a fixed centre. In the MLP structure, learning rate and momentum is taken as 0.3 and 0.2 respectively. The total number of epoch used for training is 500, validation threshold for testing is 20 and the number of features used in feature set is used as hidden neurons.

RESULTS AND DISCUSSION

For each feature set in WEBSPAM UK-2006, the authors ran CFS-PSO algorithm for the selection of optimal features. The reduction of features in link based feature set (C) was 88% (maximum) and 68% (minimum) for full content based feature set (B). Reduction for other feature sets is shown in Table 3. To test the performance of CFS-PSO (in terms of accuracy), five classification algorithms were applied: Naïve Bayes, J48, AdaBoost, SVM and MLP. After that results were compared using original number of features.

Table 3
Optimal selection of features after applying CFS+PSO

Label	Original features	Optimal features after CFS+PSO	Reduction in features
A	24	6	75%
C	41	5	88%
A+C	65	12	82%
B	96	31	68%
D	138	40	71%
B+D	234	59	75%

Evaluation of CFS+PSO for Accuracy Parameters

F_1 Score and AUC with existing (original) and optimum features (after applying CFS+PSO) for Naïve Bayes classifier is shown in Table 4. The maximum improvement of F1 Score (with optimal features) is found for feature set B+D (21.7%) followed by 15.53% in case of feature set D, 6.41% for feature set A and 2.29% for feature set C. Naïve Bayes showed improvement in AUC and also 8.23% in case of B+D feature set while 5.41%, 3.42%, 2.92% and 1.16% was seen for A+C, D, C and A feature sets respectively. But for feature set B, F measure and AUC (with optimal features) are decreased by 7.61% and 2.17% respectively (shown in Figure 6).

Table 4
F₁ score and AUC with existing (Original) and optimum features for Naïve Bayes classifier

Label	Naïve Bayes Classifier			
	F$_1$ score with existing features	AUC with existing features	F$_1$ score after CFS+PSO	AUC after CFS+PSO
A	0.39	0.687	0.415	0.695
C	0.7	0.72	0.716	0.741
A+C	0.699	0.739	0.64	0.779
B	0.67	0.738	0.619	0.722
D	0.657	0.731	0.759	0.756
B+D	0.647	0.741	0.784	0.802

(a)   (b)
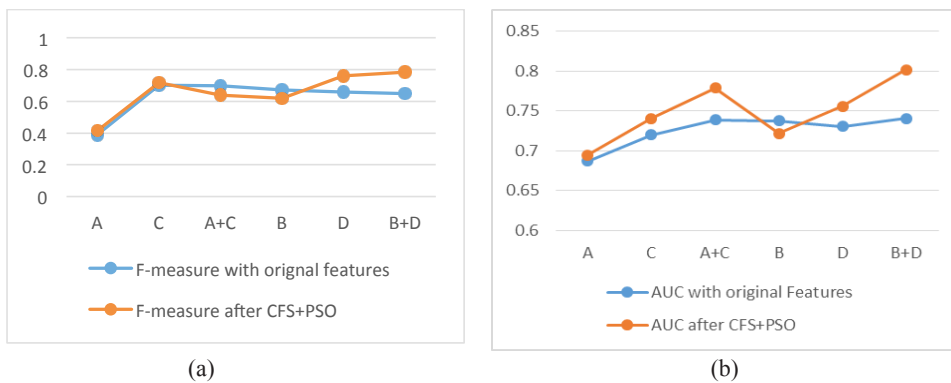
Figure 6. (a) F$_1$ score (F-measure) for NB (b) AUC for NB

For J-48, AUC (with optimal features) is increased with maximum improvement of 12.60% in case of link based features, as shown in Table 5). F1 score is improved by 9.54% for link based, 8.43% for transformed link based and 2.30% for B+D feature sets while decreased for A+C and content based feature sets (Figure 7).

Table 5
F₁ score and AUC with existing and optimum features for J-48 classifier

Label	J48 Classifier			
	F$_1$ score with existing features	AUC with existing features	F$_1$ score after CFS+PSO	AUC after CFS+PSO
A	0.563	0.706	0.514	0.719
C	0.629	0.627	0.689	0.706
A+C	0.677	0.695	0.665	0.768
B	0.646	0.694	0.646	0.717
D	0.676	0.723	0.733	0.761
B+D	0.697	0.701	0.713	0.74

(a)   (b)
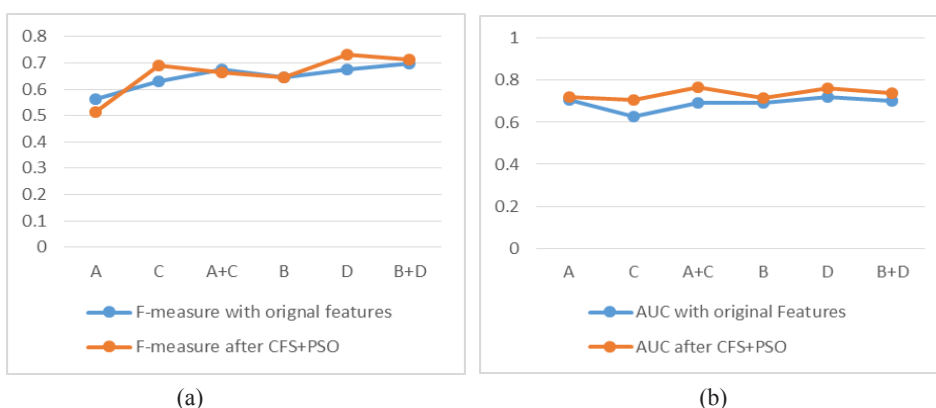
Figure 7. (a) F_1 score (F-measure) for J48 (b) AUC for J48

With AdaBoost, F1 score increased by 33.02% for transformed link based feature (denoted by D) followed by minor enhancement of 4.67% for link based feature set while decreased by 3.44% and 6.23% for A+C and full content feature set respectively (Figure 8 and Table 6). However, there is no improvement in AUC by CFS+PSO (with exception of C feature set).

Table 6
F_1 Score and AUC with existing (Original) and optimum features for AdaBoost classifier

Label	AdaBoost Classifier			
	F_1 score with existing features	AUC with existing features	F_1 score after CFS+PSO	AUC after CFS+PSO
A	0.406	0.759	0.438	0.738
C	0.643	0.68	0.673	0.685
A+C	0.668	0.773	0.645	0.748
B	0.61	0.811	0.572	0.799
D	0.315	0.763	0.419	0.744
B+D	0.655	0.84	0.726	0.822

(a)   (b)
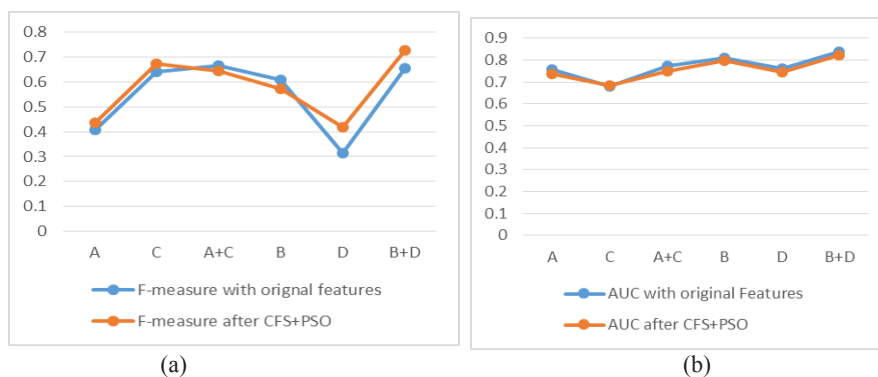
Figure 8. (a) F_1 score (F-measure) for AdaBoost (b) AUC for AdaBoost

AUC and F1 Score for every feature set is increased for SVM with optimum features (Figure 9 and Table 7). The maximum improvement in AUC and F1 Score is 3.78% and 45.83% for C feature set respectively.

Table 7
F_1 score and AUC with existing (Original) and optimum features for SVM classifier

Label	SVM Classifier			
	F_1 score with existing features	AUC with existing features	F_1 score after CFS+PSO	AUC after CFS+PSO
A	0.182	0.506	0.199	0.514
C	0.168	0.503	0.245	0.522
A+C	0.159	0.5	0.21	0.515
B	0.163	0.501	0.189	0.509
D	0.573	0.678	0.613	0.703
B+D	0.169	0.504	0.18	0.507

(a)   (b)
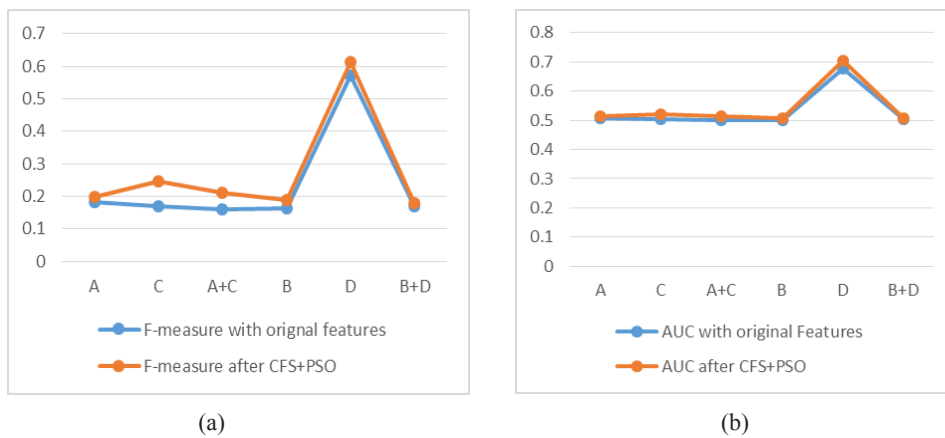
Figure 9. (a) F_1 score (F-measure) for SVM (b) AUC for SVM

In case of MLP, the maximum enhancement of F1 Score is for A+C (2.38%), full content based (8.57%) and transformed link based feature set (10.38%) while there is a decrease for all other feature sets. AUC is increased by 16.13% and 4.46% only for B+D and D feature set.

Table 8
F₁ score and AUC with existing (Original) and optimum features for MLP classifier

Label	MLP Classifier			
	F₁ score with existing features	AUC with existing features	F₁ score after CFS+PSO	AUC after CFS+PSO
A	0.576	0.801	0.557	0.783
C	0.655	0.81	0.586	0.755
A+C	0.673	0.864	0.689	0.835
B	0.595	0.827	0.646	0.814
D	0.607	0.807	0.67	0.843
B+D	0.764	0.75	0.739	0.871

(a)　　　　　　　　　　　　　　　(b)
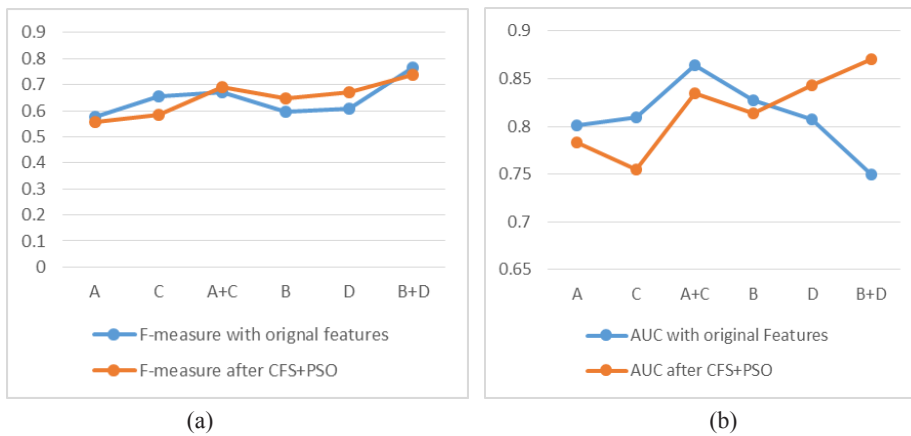
Figure 10. (a) F₁ score (F-measure) for MLP (b) AUC for MLP

Analysis of Time Complexity

The complexity of the algorithm is reduced due to the removal of irrelevant and redundant features. Hence the computational time for every classifier is also reduced. But here the authors have only shown the analysis of SVM and MLP because they are computationally very expensive in comparison to other classifiers. From Table 9, it can be seen that time taken by classifier with optimal features (by applying CFS-PSO) is less than the time taken by the same classifier with original features for every feature set.

Table 9
Computational time (in seconds) to build a model with / without CFS-PSO for SVM and MLP classifier

Label	SVM classifier		MLP classifier	
	Computational time with existing features	Computational time with optimal features	Computational time with existing features	Computational time with optimal features
A	7.48	1.16	26.06	4.9
C	58.75	15.64	62.41	2.83
A+C	67.3	24.37	143.11	9.73
B	71.73	32.35	320.13	44.17
D	23.83	8.89	662.28	70.5
B+D	108.14	46.7	1864.53	138.78

CONCLUSION

In this paper, the authors integrated CFS and PSO and comparison was done on two accuracy measures (AUC and F_1 Score) using five different classifiers The maximum increase of AUC for B+D feature set is 16.13% and 8.23% in MLP and NB respectively but in case of J48 classifier it is improved by 12.60% for link based feature set. F_1 score is improved by 33.02% in transformed link based features using AdaBoost but AUC is decreased for all feature sets except link based features. In the current analysis, CFS+PSO technique is best suited for SVM as both accuracy parameters are increased for every feature set and computational cost is also very low. In future, other meta-heuristic algorithms would be considered and comparison of this technique with wrapper and hybrid approach can also be made.

ACKNOWLEDGEMENTS

REFERENCES

Abdulla, W., & Kasabov, N. (2003). Reduced feature-set based parallel CHMM speech recognition systems. *Information Sciences, 156*(1-2), 21-38.

Becchetti, L., Castillo, C., Donato, D., Baeza-Yates, R., & Leonardi, S. (2008). Link analysis for Web spam detection. *ACM Transactions on the Web, 2*(1), 1-42.

Becchetti, L., Castillo, C., Donato, D., Leonardi, S., & Baeza-Yates, R. A. (2006). Link-based characterization and detection of web spam. In *Second International Workshop on Adversarial Information Retrieval on the Web* (pp. 1-8). Seattle, Washington, USA: ACM.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*(3), 483-519.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Caruana, R., & Freitag, D. (1994). Greedy attribute selection. In *Machine Learning Proceedings* (pp. 28-36). New Brunswick, NJ, USA: ACM.

Castillo, C., Donato, D., Gionis, A., Murdock, V., & Silvestri, F. (2007). Know your neighbors: Web spam detection using the web topology. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 423-430). Amsterdam, Netherlands: ACM.

Chow, T., & Huang, D. (2005). Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information. *IEEE Transactions on Neural Networks, 16*(1), 213-224.

Clerc, M., & Kennedy, J. (2002). The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation, 6*(1), 58-73.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*(1-4), 131-156.

Filippone, M., Masulli, F., & Rovetta, S. (2006). Supervised classification and gene selection using simulated annealing. In *International Joint Conference on Neural Networks* (pp. 3566-3571). Vancouver, BC, Canada: IEEE

Gasca, E., Sánchez, J., & Alonso, R. (2006). Eliminating redundancy and irrelevance using a new MLP-based feature selection method. *Pattern Recognition, 39*(2), 313-315.

Goh, K. L., & Singh, A. K. (2015). Comprehensive literature review on machine learning structures for web spam classification. *Procedia Computer Science, 70*, 434-441.

Goh, K. L., Singh, A. K., & Lim, K. H. (2013). Multilayer perceptrons neural network based web spam detection application. In *IEEE China Summit and International Conference on Signal and Information Processing* (pp. 636-640). Beijing, China: IEEE.

Guan, S., Liu, J., & Qi, Y. (2004). An incremental approach to contribution-based feature selection. *Journal of Intelligent Systems, 13*(1), 15-42.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1157-1182.

Gyongyi, Z., & Garcia-Molina, H. (2005). Web spam taxonomy. In *First International Workshop on Adversarial Information Retrieval on the Web* (pp. 1-9). Chiba, Japan: ACM.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning.* (Unpublished doctoral dissertation). University of Waikato, Hamilton, New Zealand. Retrieved from https://www.cs.waikato.ac.nz/~mhall/thesis.pdf

Haykin, S., & Lippmann, R. (1994). Neural networks, a comprehensive foundation. *International Journal of Neural Systems, 5*(4), 363-364.

Hsu, C. N., Huang, H. J., & Dietrich, S. (2002). The ANNIGMA-wrapper approach to fast feature selection for neural nets. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 32*(2), 207-212.

Huang, J., Cai, Y., & Xu, X. (2007). A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters, 28*(13), 1825-1844.

Kabir, M., Shahjahan, M. & Murase, K. (2012). A new hybrid ant colony optimization algorithm for feature selection. *Expert Systems with Applications, 39*(3), 3747-3763.

Kabir, M. M., Shahjahan, M., & Murase, K. (2009). An efficient feature selection using ant colony optimization algorithm. In *International Conference on Neural Information Processing* (pp. 242-252). Berlin, Heidelberg: Springer.

Kambhatla, N., & Leen, T. (1997). Dimension reduction by local principal component analysis. *Neural Computation, 9*(7), 1493-1516.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization, 39*(3), 459-471.

Ke, L., Feng, Z., & Ren, Z. (2008). An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recognition Letters, 29*(9), 1351-1357.

Li, Y., Li, T., & Liu, H. (2017). Recent advances in feature selection and its applications. *Knowledge and Information Systems, 53*(3), 551-577.

Li, Y., Luo, C., & Chung, S. M. (2008). Text clustering with feature selection by using statistical data. *IEEE Transactions on Knowledge and Data Engineering, 20*(5), 641-652.

Ntoulas, A., Najork, M., Manasse, M., & Fetterly, D. (2006). Detecting spam web pages through content analysis. In *Proceedings of the 15ᵗʰ International Conference on World Wide Web* (pp. 83-92). New York, USA: ACM.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters, 15*(11), 1119-1125.

Quinlan, J. R. (1993). *C4. 5: Programming for machine learning.* San Mateo, CA: Morgan Kaufmann Publishers.

Ramos, C. C., Souza, A. N., Chiachia, G., Falcão, A. X., & Papa, J. P. (2011). A novel algorithm for feature selection using harmony search and its application for non-technical losses detection. *Computers and Electrical Engineering, 37*(6), 886-894.

Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research, 3*(Mar), 1371-1382.

Senliol, B., Gulgezen, G., Yu, L., & Cataltepe, Z. (2008). Fast Correlation Based Filter (FCBF) with a different search strategy. In *23ʳᵈ International Symposium on Computer and Information Sciences* (pp. 1-4). Istanbul, Turkey: IEEE

Shunmugapriya, P., & Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm and Evolutionary Computation, 36*, 27-36.

Singh, A. K., Kumar, R., & Leng, A. G. K. (2011). An experimental study on spam detection algorithms. In *Proceeding of IEEE Conference TENCON* (pp. 1382-1385). Bali, Indonesia: IEEE.

Singh, S., & Singh, A. K. (2018). Web-spam features selection using CFS-PSO. *Procedia Computer Science, 125*, 568-575.

Song, Q., Fong, S., & Tang, R. (2016). Self-adaptive wolf search algorithm. In *5ᵗʰ IIAI International Congress on Advanced Applied Informatics* (pp. 576-582). Kumamoto, Japan: IEEE.

Song, Q., Ni, J., & Wang, G. (2013). A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering, 25*(1), 1-14.

Sotoca, J. M., & Pla, F. (2010). Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition, 43*(6), 2068-2081.

Tang, R., Fong, S., Yang, X. S., & Deb, S. (2012). Wolf search algorithm with ephemeral memory. In *Seventh International Conference on Digital Information Management* (pp. 165-172). Macau, China: IEEE.

Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-Based Systems, 24*(7), 1024-1032.

Webber, C. G., Maria de Fátima, W., & Hepp, F. S. (2012). Testing phishing detection criteria and methods. In *Frontiers in Computer Education* (pp. 853-858). Berlin, Germany: Springer.

WEBSPAM-UK2006. (2006). *WEBSPAM-UK2006 (previous dataset).* Retrieved January 10, 2018, from http://chato.cl/webspam/datasets/uk2006/.

Xue, B., Zhang, M., & Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics, 4*3(6), 1656-1671.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. In *Feature extraction, construction and selection* (pp. 117-136). Boston, MA: Springer.

Yang, X. S. (2009). Firefly algorithms for multimodal optimization. In *International Symposium on Stochastic Algorithms* (pp. 169-178). Berlin, Germany: Springer.

Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In *Nature inspired cooperative strategies for optimization* (pp. 65-74). Berlin, Germany: Springer.

Yang, X. S., & Deb, S. (2009). Cuckoo search via Lévy flights. In *Nature and Biologically Inspired Computing* (pp. 210-214). Coimbatore, India: IEEE.

Yang, X. S., & He, X. (2013). Firefly algorithm: Recent advances and applications. *International Journal of Swarm Intelligence, 1*(1), 36-50.

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research, 5*(Oct), 1205-1224.

SCIENCE & TECHNOLOGY

PERTANIKA JOURNALS

Evaluation of Ball-Milling Process for the Production of Carbon Particles from Rice Straw Waste

Asep Bayu Dani Nandiyanto[1]*, Rosi Oktiani[1], Rena Zaen[1], Ari Arifin Danuwijaya[2], Ade Gafar Abdullah[3] and Nuria Haristiani[4]

[1]*Departemen Kimia, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*
[2]*Departemen Bahasa Inggris, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*
[3]*Departemen Teknik Elektro, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*
[4]*Departemen Bahasa Jepang, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*

ABSTRACT

The purpose of this study is to evaluate the ball-milling process for the production of carbon particles from rice straw waste. In the experimental method, carbon particles are prepared by adding heat treatment to the rice straw waste at the temperature of 250ºC. The heated rice straw waste is grinded using a conventional grinding method and put into the ball-milling process. Physicochemical properties of carbon particles are evaluated before and after the ball-milling process. The results indicate that sizes of carbon particles decrease significantly from micro (after the conventional grinding process) to submicron range (after additional ball milling process). Almost no changes in the chemical properties of carbon particles are observed. This result implies that the ball-milling process is effective to decrease the size of the prepared carbon particles. The collision phenomenon combined with shear stress among the stainless steel balls is the main reason for breaking carbon particles, making the particle sizes smaller.

Keywords: Activated carbon particles, ball-milling process, rice straw waste

INTRODUCTION

One of the best and popular methods for reducing particle size, known as a top-down method, is a ball-milling process. This process has been widely used for various materials, such as organo-vermiculite (Wang et al., 2011), Bi_3NbO_7 ceramics (Zhou, Wang, & Yao, 2007), nanocrystalline ZrO_2 (Chadwick, Pooley, Rammutla, Savin, & Rougier, 2003),

Pb $(Zr_{0.52}Ti_{0.48})O_3$ (Kong, Zhu, & Tan, 2000), magnetic nanoparticles (Chakka, Altuncevahir, Jin, Li, & Liu, 2006), WO_3/TiO_2 photocatalyst (Shifu, Lei, Shen, & Gengyu, 2005), and carbon nanotubes (Pierard, et al., 2001). This method is effective to reduce particle size down to submicron and further to nanometre range. Although the experimental procedures have been well-documented, reports on the use of ball-milling process to prepare carbon from rice straw waste are relatively rare.

Previous studies (Permatasari, Sucahya, & Nandiyanto, 2016; Nandiyanto et al., 2016a, 2016b, 2017a, 2017b, 2017c; Nandiyanto, Zaen, Oktiani, Abdullah, & Danuwijaya, 2018; Nandiyanto, Zaen, & Oktiani, in press), developed a method for the production of particles with controllable physicochemical properties. Control of sizes from nano to submicrometre range was also discussed.

Based on the previous works, the objective of this study is to evaluate the impact of ball-milling process on the changes of physicochemical properties of carbon particles gained from rice straw waste. To support the objective of this study, the physical appearance of the carbon particles prepared via the ball-milling process is analysed using a scanning electron microscope (SEM), whereas the chemical properties are evaluated using an X-Ray Diffraction (XRD).

Why Rice Straw Waste is an Excellent Candidate for Carbon Source?

Rice straw is used as a candidate for producing carbon particles (Karyasa, 2016; Permatasari et al., 2016; and Nandiyanto et al., 2016a). It offers many advantages, reducing serious environmental problems gained from rice straw waste. A large amount of rice straw increases as the production of rice continues. Although this abundant rice straw as by-products can be minimised by disposing and burning in the rice field, the burning process is not effective since the secondary product is still persisted (Nandiyanto, Zaen, & Oktiani, 2017d). Thus, understanding how to solve or reuse this material is inevitable.

Rice straw is rich in organic nutrients (such as cellulose (32-47%), hemicellulose (19-27%), lignin (5-24%), and ash (13-20%) that can be converted into carbon material (Nandiyanto et al., 2018). This is another main reason for the selection of rice straw waste as one the excellent candidates for carbon sources. Previous studies have reported the preparation of carbon from rice straw using chemical activation, pyrolysis, and temperature treatment. This information has been reviewed by Ioannidou and Zabaniotou (2007). Although their method is effective, there is no information how to reduce the carbon particle size using the ball-milling process. In fact, reducing particle size has a great impact on improving material performance, such as surface area.

Illustration for the Production of Carbon from Rice Straw Waste

Figure 1 shows an illustration mechanism for converting rice straw waste into carbon particles. The process involves two steps. The first is heat treatment, which is an effective method to convert an organic component of rice straw waste into carbon. The next step is ball-milling process, in which this method is used to decrease the particle size of the heated rice straw waste. Detailed information about the preparation of carbon from rice straw is described in the previous report (Nandiyanto, et al., 2017c).
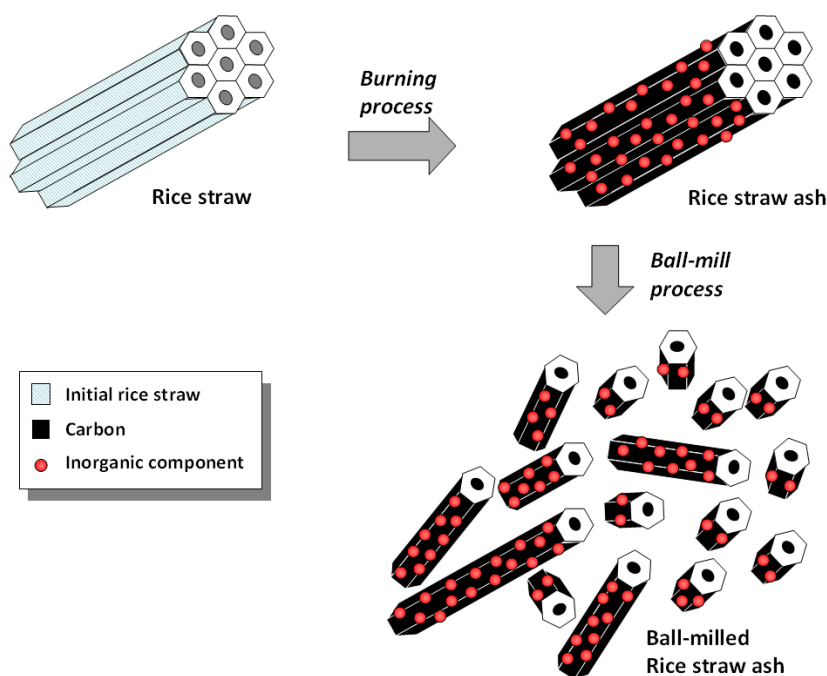
Figure 1. Schematic illustration of the formation of carbon porous particles from rice straw waste

Heat Treatment Process for Converting Rice Straw Waste into Carbon Material

Figure 2 shows the effect of additional heat treatment on the rice straw waste based on the previous studies. To confirm this impact, Nandiyanto et al. (2017c) reported the thermal analysis during the additional heat treatment on the rice straw waste. The analysis showed the detection of mass reduction during the heating process with existence of several exothermic and endothermic reactions. They concluded that some phenomena happened during the heating process (Nandiyanto et al., 2016a, 2017c):

(i)   Evaporation of physically adsorbed water at temperature of 20-100°C;

(ii) Carbonisation from cellulose and hemicellulose degradation process at temperature of 250-320°C;

(iii) Carbonisation from lignin degradation process at temperature of 330-470°C;

(iv) Carbon degradation due to oxidation at temperature of higher than 500°C.

According to above results, the effective temperature used in the heat treatment process of rice straw to produce carbon material is 250°C. This temperature is considered optimum to obtain carbon in the highest volume. If the temperature applied is higher than the optimum temperature condition, it can create two disadvantages; one is wasting thermal energy and the other is oxidation process that can deter carbon volume. Although some processes show excess temperature can increase the physicochemical properties of the prepared carbon due to its impact on activating carbon, the present study will not focus on the activation process.
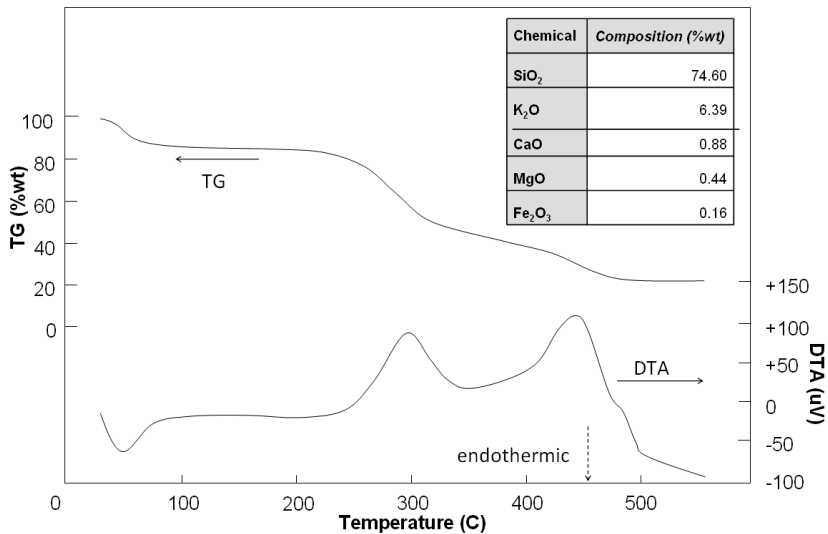
Figure 2. Thermal analysis results of rice straw waste. The insert table shows the chemical composition of rice straw. From Nandiyanto et al., 2016b

Technical Information about Ball-milling Process

When applying heat treatment to the rice straw waste, the prepared carbon is a replication of the initial size of the rice straw waste. Therefore, to create smaller particles, the rice straw waste must be sliced very thinly. However, cutting rice straw waste into smaller size is difficult since the rice straw is relatively pliant. Thus, some researchers add mechanical milling process after the burning rice straw, such as ball-milling process, for creating smaller carbon particles.

In short, illustration of the ball-milling process is shown in Figure 3. The ball-milling process is conducted by adding material and balls (as a material destroyer) into the milling vessel (see Figure 3(a)). During the ball-milling process (See Figure 3(b)), the balls inside the ball-milling machine cylinder vessel are rotated. During rotation, balls move and create shear stress. Then, when the ball reaches a certain point, it falls due gravitational force. The falling ball then overwrites other balls, creating collision process. Indeed, a combination between the "shear stress" and the "collision process" gives mechanical force and leads to the creation of particles with smaller sizes.
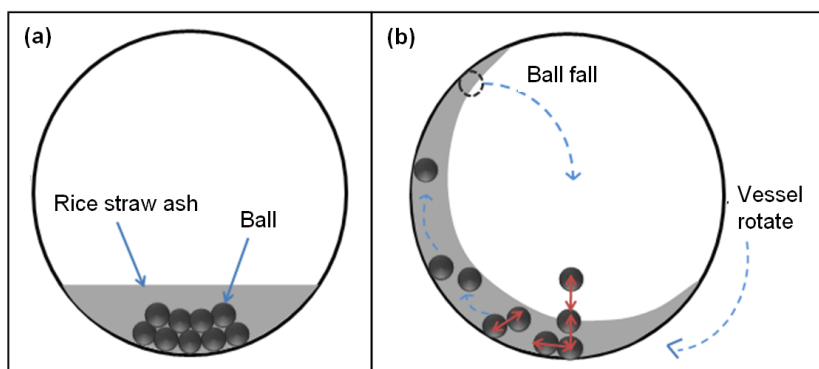
Figure 3. (a) The illustration of ball-mill processing in the initial condition and (b) after rotation process/ during ball-milling process

EXPERIMENTAL METHOD

The material used for the preparation of carbon material was rice straw waste gained from rice field in Kebumen, Indonesia. In the experimental procedure, the carbon particles were prepared using several steps: (i) heat treatment process; and (ii) ball-milling process.

Heat Treatment

Rice straw waste was washed, cut into sizes of about 1 cm, and dried at 100°C. The dried rice straw was then heat-treated at 250ºC in an electrical furnace. The heat-treated rice straw was grinded using a conventional grinding method to form carbon particles.

Ball-milling Process

The grinded carbon particles are put into a 1.42 L of batch-typed ball-milling vessel equipped with stainless steel balls (diameter of 8 mm). Ball-mill vessel is made from stainless steel with dimensions of 25 and 8.50 cm for length and diameter respectively. Ball-milling process is conducted at a room temperature with rotation speed of 70 rpm for 30 minutes. Illustration of the ball-milling process is shown in Figure 3.

Characterisation

The physical appearances of the produced carbon particles are analysed using an electron microscope (SEM; JSM-6360LA; JEOL Ltd., Japan). The chemical properties of the produced carbon particles are characterised using an X-ray diffraction (XRD; PANalytical X'PertPRO; Philips Corp., Netherland).

RESULTS AND DISCUSSION

Figure 4 shows the SEM analysis images of the prepared carbon particles before and after the ball-milling process. The differences in shape and size are observed. Carbon particles before ball-milling process have a homogeneous shape that is rectangular with a length of between

1800 and 3750 nm (Figure 4(a)). The size of the carbon particles is identical to the size of the initial rice straw waste, confirming the carbon particles are the replication of rice straw waste. Then, applying ball-milling process to the carbon resulted in the decreases in the particle size, in which the particles have more diverse and irregular shape with a length of between 125 and 625 nm (Figure 4(b)).

Based on the above results, we can conclude that the conventional grinding process is effective to create micrometer-sized particles. However, this method has limitations due to ineffectiveness of the production of particles with sizes of submicron or further nanometre range. The additional ball-milling process is decrease the particle size down to hundreds of nanometres. Differences in shape and size gainded from above methods (i.e. conventional grinding method and ball-milling process) are due to the existence of shear stress and collisions among stainless steel balls during the ball-milling process.
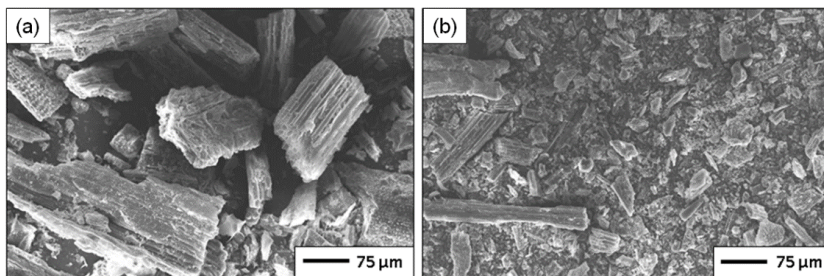
Figure 4. SEM images of carbon particles from rice straw waste (a) before and (b) after ball-milling process

In order understand the phenomenon that occurs during the ball-milling process, Figure 5 illustrates the size destruction mechanism during ball-milling process. In this figure, collision between balls is focused, while the shear stress is assumed to have less impact than the ball collision.

The collision between two stainless steel balls pressures the carbon to break into smaller parts. Since the colliding process continuously occurs, the carbon faces continuous collision, resulting in more size destruction to happen. Indeed, the final particle sizes become smaller.

The breaking characteristics of carbon particles depend on the position of the carbon material when the collision between the stainless steel balls takes place. If the collision occurs at the center of the carbon material, the carbon would break in the middle parts, permitting formation of two equal lengths and some smaller irregular parts of carbon particles (see Figure 5(a)). Indeed, when the collision occurs at the edge of the carbon material, the edge of the carbon turns into smaller irregular parts (see Figure 5(b)). If the collision occurs in the vertical position of the activated carbon, the carbon graphite breaks into some irregular parts (see Figure 5(c)). When the collission process continously occurs, the final particles have smaller sizes with various shapes. Indeed, the sizes and shapes of ball-milled carbon are different compared with that of originated carbon, as shown in Figure 4(b)).
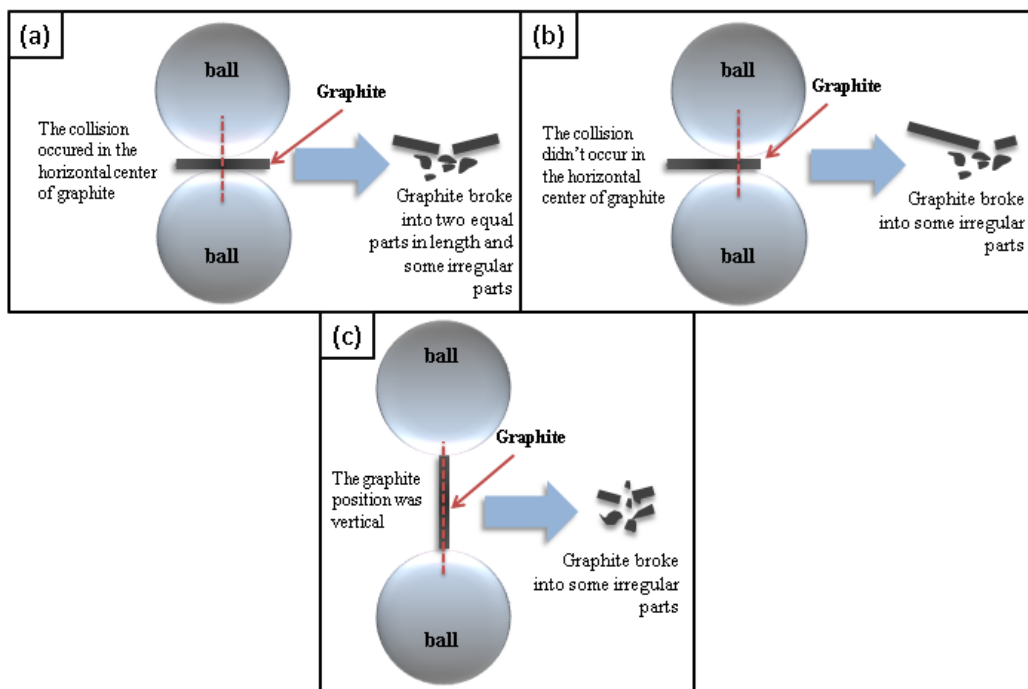
Figure 5. The process of rupture of graphite due to the collision of stainless steel balls in the ball milling process. (a) The collision occurring in the horizontal center, (b) does not occur in the horizontal center, and (c) when the graphite position is vertical

Figure 6 shows XRD analysis results of the prepared carbon particles. Three samples were analysed. The first sample was heat-treated rice straw waste. The next sample undergone additional conventional grinding process on the heated rice straw waste. Finally, the third sample undergone additional conventional grinding with ball-milling processes on the heated rice straw waste. All samples were tested based on the Joint Committee Powder Diffraction System (JCPDS) for graphite (JCPDS No. 75-1821), silica (JCPDS No. 27-1402), and calcium magnesium silicate ($CaMgSi_2O_6$) (JCPDS No. 11-854).

The results show all samples contain graphite, silica, and $CaMgSi_2O_6$. The content of silica and $CaMgSi_2O_6$ is probably derived from the initial rice straw as the previous described in our previous reports (Nandiyanto et al., 2016b) using atomic absorption spectrophotometry and gravimetric analyses. No additional pattern appeared in all samples, confirming no chemical reaction during the grinding and ball-milling process. However, the intensities of graphite, silica, and $CaMgSi_2O_6$ change during grinding and ball-milling process. When adding grinding and ball-milling processes, intensity of graphite decreases followed by the increasing intensities of silica and $CaMgSi_2O_6$.
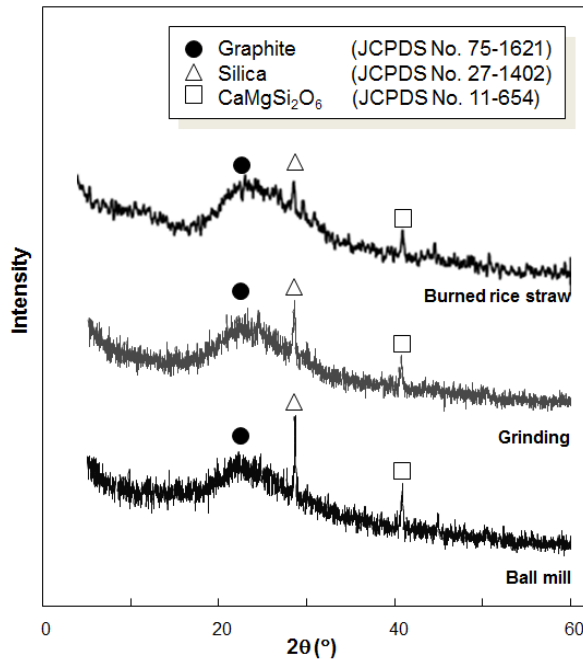
Figure 6. XRD analysis result of burnt rice straw, activated carbon by grinding, and activated carbon by ball milling

To confirm the analysis of the chemical component during the grinding and ball-milling process, component analysis based on pattern area in the XRD result was conducted (See Table 1). The result shows that graphite composition decreases from about 90 to about 84%. The decrease in graphite composition reached more than 6%. This decrease of the graphite composition was followed by the increases in the silica and $CaMgSi_2O_6$ of 4.91 and 1.38% respectively.

Two hypotheses can explain the change in the composition. The first hypothesis is due to the possibility of carbon attached to the balls and the ball-milling apparatus. However, this hypothesis cannot be accepted clearly. If the carbon is attached to the balls and the ball-milling apparatus, this should be followed by a decrease in yield and this will not be composition. The second hypothesis is formulated based on the existence of oxidation process during the ball-milling process. However, the present ball-milling process was conducted in batch process. No circulation of air happens inside the ball-milling process. Further, the process is fixed at room temperature, deterring the possibility of oxidation process in this temperature range.

In addition, the use of batch-typed ball-milling process is the best, compared with that of continuous-typed ball-milling process. The existence of oxygen can be minimised. Indeed, this will deter the possibility of the oxidation process as discussed in the second hypothesis. However, based on both hypotheses, further analyses must be conducted and will be done in our future work.

Table 1
Percentage (%) of intensity from rice straw prepared by burning, grinding, and ball-milling process

Component	Rice straw waste		
	Burned	Grinded	Ball-milled
Graphite	90,45%	88,36%	84,16%
Silica	7,54%	8,41%	12,45%
$CaMgSi_2O_6$	2,01%	3,23%	3,39%

CONCLUSION

The results show the ball-milling process is effective in producing smaller activated carbon particles. By applying the ball-milling process, the size of the carbon particles decreases from micrometer to sub-micrometre range. Interestingly, the particle sizes can be decreased to 125 nm. In addition, there are no different material compositions between samples with grinding and ball-milling process. This confirms no chemical reaction during the grinding and ball-milling process. The main reason for this successful production of sub-micrometre particles when applying ball-milling process is due to the existence of shear stress and collision phenomena among stainless steel balls and carbon material.

ACKNOWLEDGEMENT

REFERENCES

Chadwick, A. V., Pooley, M. J., Rammutla, K. E., Savin, S. L., & Rougier, A. (2003). A comparison of the extended x-ray absorption fine structure of nanocrystalline ZrO_2 prepared by high-energy ball milling and other methods. *Journal of Physics: Condensed Matter, 15*(3), 431-440.

Chakka, V. M., Altuncevahir, B., Jin, Z. Q., Li, Y., & Liu, J. P. (2006). Magnetic nanoparticles produced by surfactant-assisted ball milling. *Journal of Applied Physics, 99*(8), 08E912.

Ioannidou, O., & Zabaniotou, A. (2007). Agricultural residues as precursors for activated carbon production—A review. *Renewable and Sustainable Energy Reviews, 11*(9), 1966-2005.

Karyasa, I. W. (2016). Renewable silica-carbon nanocomposite and its use for reinforcing synthetic wood made of rice straw powders. *KnE Engineering, 1*(1), 1-6.

Kong, L. B., Zhu, W., & Tan, O. K. (2000). Preparation and characterization of $Pb(Zr_{0.52}Ti_{0.4}8)O_3$ ceramics from high-energy ball milling powders. *Materials Letters, 42*(4), 232-239.

Nandiyanto, A. B. D., Putra, Z. A., Andika, R., Bilad, M. R., Kurniawan, T., Zulhijah, R., & Hamidah, I. (2017c). Porous activated carbon particles from rice straw waste and their adsorption properties. *Journal of Engineering Science and Technology, 4*(2), 1-11.

Nandiyanto, A. B. D., Rahman, T., Fadhlulloh, M. A., Abdullah, A. G., Hamidah, I., & Mulyanti, B. (2016a). Synthesis of silica particles from rice straw waste using a simple extraction method. *IOP Conference Series: Materials Science and Engineering, 128*(1), 012040.

Nandiyanto, A. B. D., Sofiani, D., Permatasari, N., Sucahya, T. N., Wiryani, A. S., Purnamasari, A., & Prima, E. C. (2016b). Photodecomposition profile of organic material during the partial solar eclipse of 9 March 2016 and its correlation with organic material concentration and photocatalyst amount. *Indonesian Journal of Science and Technology, 1*(2), 132-155.

Nandiyanto, A. B. D., Wiryani, A. S., Rusli, A., Purnamasari, A., Abdullah, A. G., & Riza, L. S. (2017, March). Decomposition behavior of curcumin during solar irradiation when contact with inorganic particles. In *IOP Conference Series: Materials Science and Engineering* (Vol. 180, No. 1, p. 012135). IOP Publishing.

Nandiyanto, A. B. D., Wiryani, A. S., Rusli, A., Purnamasari, A., Abdullah, A. G., Widiaty, I., & Hurriyati, R. (2017, March). Extraction of curcumin pigment from Indonesian local turmeric with its infrared spectra and thermal decomposition properties. In *IOP Conference Series: Materials Science and Engineering* (Vol. 180, No. 1, p. 012136). IOP Publishing.

Nandiyanto, A. B. D., Zaen, R., & Oktiani, R. (2017d). Current utilizaton of rice straw waste: from the traditional method to the application of advanced technology for isolating silica and activated carbon materials. In A. Timayev & G. Kadyrov (Eds.), *Rice and rice straw production, cultivation, and uses* (pp.193-230). New York: Nova Science Publishers, Inc.

Nandiyanto, A. B. D., Zaen, R., & Oktiani, R. (In Press). Correlation between crystallite size and photocatalytic performance of micrometer-sized monoclinic WO$_3$ particles. *Arabian Journal of Chemistry*.

Nandiyanto, A. B. D., Zaen, R., Oktiani, R., Abdullah, A. G., & Danuwijaya, A. A. (2018). Monodispersed and size-controllable potassium silicate nanoparticles from rice straw waste produced using a flame-assisted spray pyrolysis. *Pertanika Journal of Science and Technology, 2*6(1), 391–408.

Permatasari, N., Sucahya, T. N., & Nandiyanto, A. B. D. (2016). Agricultural wastes as a source of silica material. *Indonesian Journal of Science and Technology, 1*(1), 82-106.

Pierard, N., Fonseca, A., Konya, Z., Willems, I., Van Tendeloo, G., & Nagy, J. B. (2001). Production of short carbon nanotubes with open tips by ball milling. *Chemical Physics Letters, 335*(1), 1-8.

Shifu, C., Lei, C., Shen, G., & Gengyu, C. (2005). The preparation of coupled WO$_3$/TiO$_2$ photocatalyst by ball milling. *Powder Technology, 16*0(3), 198-202.

Wang, L., Chen, Z., Wang, X., Yan, S., Wang, J., & Fan, Y. (2011). Preparations of organo-vermiculite with large interlayer space by hot solution and ball milling methods: A comparative study. *Applied Clay Science, 51*(1), 151-157.

Zhou, D., Wang, H., & Yao, X. (2007). Sintering behavior and dielectric properties of Bi$_3$NbO$_7$ ceramics prepared by mixed oxides and high-energy ball-milling methods. *Journal of the American Ceramic Society, 90*(1), 327-329.

Implementation of Markerless Augmented Reality Method to Visualise Philosophy of Batik Based on Android

Isma Widiaty[1]*, Ivan Yustiawan[2], Yudi Wibisono[2], Ade Gafar Abdullah[3], Cep Ubad Abdullah[4] and Lala Septem Riza[2]

[1]Departement of Home Economics, Universitas Pendidikan Indonesia, Jl.Dr.Setiabudhi, Bandung 40154, Indonesia
[2]Departement of Computer Science Education, Universitas Pendidikan Indonesia, Jl.Dr.Setiabudhi, Bandung 40154, Indonesia
[3]Departement of Electrical Engineering, Universitas Pendidikan Indonesia, Jl.Dr.Setiabudhi, Bandung 40154, Indonesia
[4]Departement of Physical Education, Universitas Pendidikan Indonesia, Jl.Dr.Setiabudhi, Bandung 40154, Indonesia

ABSTRACT

Indonesia has a variety of culture, tribes, customs, religion, and arts. One of the most popular cultural heritages in Indonesia is Batik. Each print or pattern has different meanings reflecting the culture of the area where the batik originates from. There are many ways to preserve batik; one of which is by giving information and understanding about the culture of batik through a technology called Augmented Reality (AR). In this study, the philosophy of batik patterns is visualised through AR application. Markerless's AR method is used so that the AR used to detect batik without using any markers. The application is implemented on android-based devices, either smartphones or tables, as they are very popular. Results of the visualisation are the image, name and philosophy of the batik detected. The application can detect at most around 50 cm and 30°C.

Keywords: Batik pattern, android, Markerless augmented reality method, philosophy

INTRODUCTION

Indonesia has many different tribes, culture, religion, and art. One of the most popular cultural heritages is batik. Batik is a fabric which has a unique pattern and processed in a certain way (Syahputra & Soesanti, 2016). Almost each area in Indonesia has batik with different patterns that have special philosophical meanings reflecting their culture.

The younger generation in Indonesia needs to preserve their culture; one of which is batik. One way of doing it is by understanding and teaching about batik and its philosophical meanings to the next generation using a technology namely augmented reality (AR) and android-based smartphones, which are very popular. Therefore, an AR application to visualise the information of the philosophy of batik patterns based on android using Markerless augmented reality is proposed, where the application detects the objects directly without any marker.

There are studies on applications to introduce batik patterns based on AR (Widiaty, Riza, Danuwijaya, Hurriyati, & Mubaroq, 2017), in which the application is able to detect batik as its reference. To detect the objects directly, the algorithm used in this study is ORB (Oriented FAST and Rotated BRIEF). Based on previous studies (Rublee, Rabaud, Konolige, & Bradski, 2011), ORB is proven to outperform SIFT (Scale Invariant Feature Transform) and SURF (Speed Up Robust Features). In matching feature, the algorithm used is hamming distance which measures the difference between two strings with the same length. The process of hamming distance is fast and efficient since it can be operated using XOR followed by a few instructions (Muhammad, 2015). To match key point, the study uses RANSAC which estimates the homography of picture perspective mapping.

AUGMENTED REALITY AND ITS TECHNOLOGY

Augmented Reality

Augmented Reality is a technology combining real world with virtual environment made by computers in real time (Azuma, 1997). It is actually a variation of virtual environment or popularly known as virtual reality.
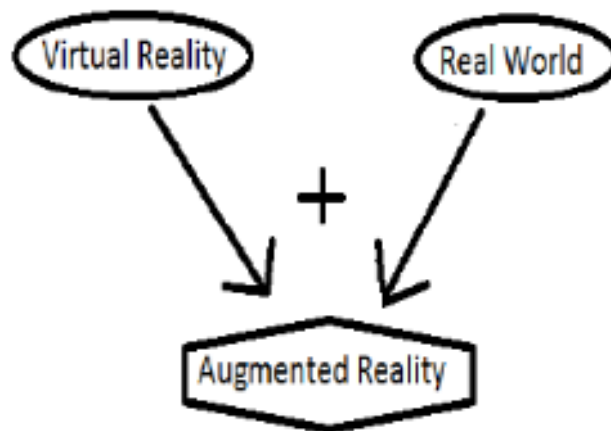
Figure 1. Augmented Reality as a combination of real world and virtual reality. From "A survey of augmented reality", by Shoaib and Jaffry, 2015

Almost every AR system available nowadays focuses mainly on the annotation of the real world objects (Neri, García, Barón, & Crespo, 2013). One of the AR applications developed is in mechanical repair, in which AR is used to give explanations on parts of the car machines identified by the users. Thus, the users can identify certain parts of the machines where the AR system shows lines and texts by computers describing components from the machines as the following Figure 2 shows.
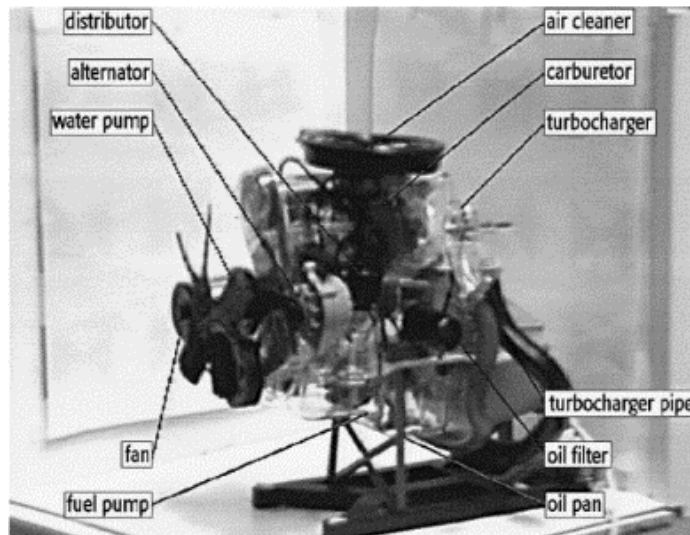
Figure 2. Annotation and visualisation on mechanics. From "A survey of augmented reality", by Azuma, 1997

Markerless Augmented Reality

Markerless augR method to detect objects directly without any marker. There are two techniques.

1. Pose Tracking

   Pose Tracking can be used when the device has a camera and moving screen. This technique can be used by utilising Global Positioning System (GPS), digital compass, and sensor.

2. Pattern Matching

   Pattern Matching is actually almost similar to marker-based AR, yet it does not use marker. Instead, it uses common pictures. This technique can be used to recognise pictures, such as book covers, paintings, bus windows, human faces, and others.

3. Object Recognition

   Object Recognition is a way to recognise or identify an object. The common process of recognising objects and tracking usually uses a lot of algorithms based on features. The tracking based on features is a way to find a certain feature of a picture. After tracking, matching takes place to see whether two objects are similar to one another.

In feature-based tracking, there are several algorithms used (Du, Miao, & Cen, 2014), such as SURF (Speed Up Robust Features), SIFT (Scale Invariant Feature Transform), and BRISK (Binary Robust Invariant Scalable Key points). This study uses ORB which will be discussed in the next sub-chapter. The scheme of object recognition is described in Figure 3.
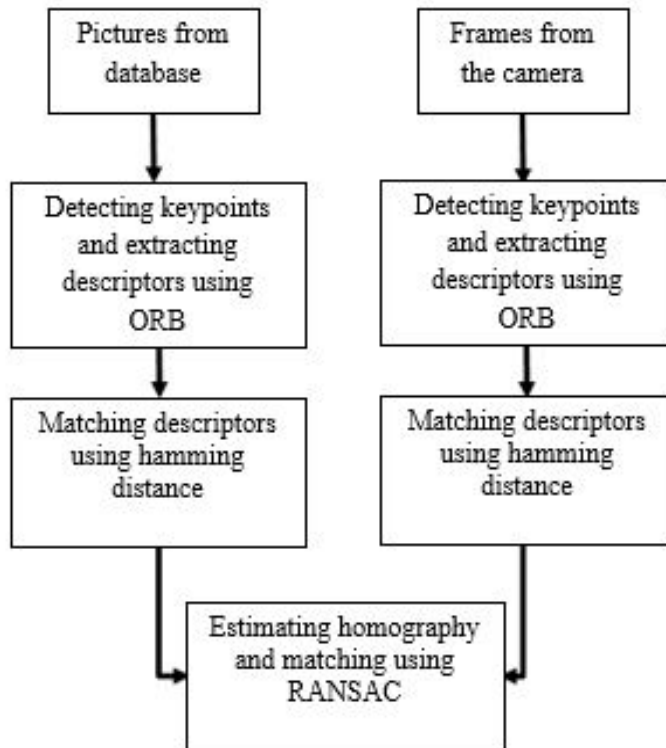
Figure 3. The scheme of object recognition

ORB (Oriented FAST and Rotated BRIEF) is a combination between oriented FAST feature point and rotated BRIEF description algorithms (Rublee et al., 2011). FAST algorithm is used to detect key points in pictures while BRIEF algorithm is used as a descriptor (Rublee et al., 2011).

FAST (Feature from Accelerated Segment test) is an algorithm to detect key points quickly. It is based on a parameter, namely threshold between the centre pixels and the other pixels. It does not have corner measurement. The measurement of Harrick corner is used to get the FAST key points. For the number of target of the key points, threshold is set quite low then done in accordance with Harris measure to get the highest number of key points. FAST also does not result if multi-scale feature, yet it uses picture scale pyramid and result it FAST feature in every level of pyramid.

An approach based on intensity centroid orientation uses simple yet effective measurement. The centroid intensity assumes that the corner intensity is corrector from the centre and that the vector that can be used to connect orientation is defined as follows.

$$m_{pq} = \sum_{xy} x^p y^q \, I(x,y)$$

Where m_{pq} represents the order moment $(p+q)^{th}$ of a picture whose intensity change $I(x,y)$ as a function of x and y coordinates in the figure. The centroid is then obtained as follows.

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right)$$

A vector is built from the centre to the centroid $\overrightarrow{OC}$ and the patch orientation becomes:

$$\theta = atan2(m_{01}, m_{10})$$

Where atan2 represents the quadrant version found from the arctan.

ORB also needs an additional component known as r-BRIEF. The descriptor of BRIEF is string bit description from various patches of pictures made from every set intensity of the binary test. For instance, a patch of a picture that is already softened is then identified as p and thus, the binary test and τ are defined as follows.

$$\tau(p; x, y) := \begin{cases} 1 : p(x) < p(y) \\ 0 : p(x) \geq p(y) \end{cases}$$

Where p(x) is the intensity of p in an x point. The feature of *f* is defined as a vector test of the *n* binary test

$$f_n(p) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i y_i)$$

The concept of hamming distance is basically measurement of two strings with the same length. It shows numbers between two binary strings. The process is fast and efficient since it can use machine instruction or XOR operation with a few numbers (Muhammad, 2015).

Table 1
Example of hamming distance

String 1	String 2	Hamming Distance
karolin	kathrin	3
karolin	kerstin	3
1011101	1001001	2
2173896	2233796	3

In this study, hamming distance is used to match the results of ORB descriptors. The descriptor with the shortest distance will be replaced with the right picture.

Homography is a transformational matrix used to project one picture over another based on match feature found by multiplying the pictures. The pictures will transform geometrically through translation, rotation, scaling, skew, shear, etc. The equation for homography is formulated as follows (Kriegman, 2007):

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

(x_2, y_2, z_2) is the posisiton of pixels as the results of transformation, while H is a homographical matrix sized 3 x 3 and (x_1, y_1, z_1)) is the position of the pixel in the original pictures.

The algorithm of RANSAC (Random Sample And Consensus) was first introduced as a method to estimate a certain parameter contaminated by outlier (average deviation point) in a large number (Fischler & Bolles, 1981). Outliers are a feature with more deviant values compared with other features. In this study, RANSAC is used to estimate the homography. The steps are as follows:

1. Choosing four correspondences randomly and calculating the homography as explained before.

2. Calculating the distance of $d_\perp$ for every assumed correspondence.

3. Calculating the number of matching inliers in H with the number of correspondence where $d_\perp$<t.

4. Choosing H with the biggest inliers.

5. Re-estimating H from all the correspondences arranged based on inliers. Determining correspondent key points using H estimated to define the search location of the changed point position.

OpenCV

OpenCV (Open Source Computer Vision) is a library computer vision open source written in C dan C++ (Bradski, Kaehler, & Pisarevsky, 2005). The algorithm in OpenCV can be used to detect and recognise faces, identify objects, clarify human movements in videos, search movements in cameras, detect moving objects, make 3D models of an object, result in 3D from stereo cameras, combine pictures to have high resolution, find similar pictures in the database, erase red eyes from pictures taken with flash, follow eye movements, recognise environment and build objects with augmented reality. In this study, the OpenCV is used as a tool and library in making android-based augmented reality which is the process of recognition. OpenCV used in this study is OpenCV for android that can be downloaded on http://opensource.org.

Android

Android is an open source platform designed for mobile devices bought by Google Inc from Android Inc (Gargenta & Nakamura, 2014). The six frameworks of augmented reality developed from android are ARLab, ARToolKit, D'Fusion, Vuforia, catchoom and metaio (Marneanu, Ebner, & Roessler, 2014). There is also an OpenCV which provides a library for augmented reality. In this study, OpenCV is used as a medium to make the application.

ARCHITECTURAL DESIGN OF AR

Figure 4 is the architectural design of AR proposed in this study.
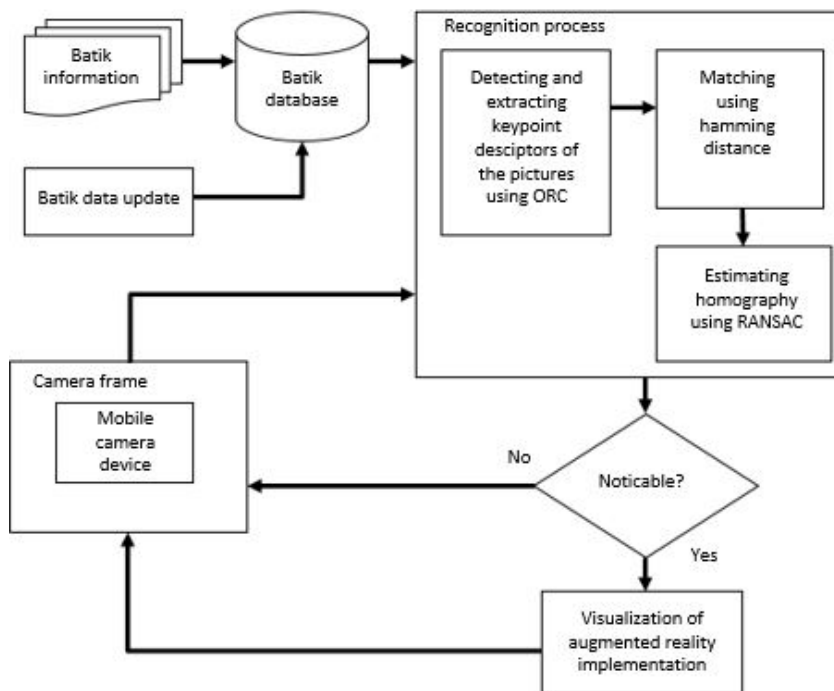
Figure 4. The architectural design of augmented reality

Information about Batik

Data about batik information is inserted into database namely db_batik di sever whose URL is http://batik.16mb.com/index.php/. The data input into the server include names of batik, philosophy of batik, and pictures of batik. One type of batik may have one pattern and others may have more.

Recognition Process

In recognition process, functions used are mostly those of OpenCV. Below is a series of recognition process.

1. Detecting and extracting picture key point descriptors using ORB.

    Before the pictures are processed to obtain key points, the pictures are transformed into grayscale. The results are a key point matrix from the detected pictures. The extraction leads to vector features. These steps are also done in the camera frame.

2. Matching using hamming distance

    Matching vector feature is done using hamming distance algorithm resulting in distance values. The minimum distance from the vector values is then calculated. After the minimum value is acquired, the vector values of the distance are selected. Only ones belonging to good match will be chosen if the distance values are smaller than the minimum distance values is tripled. The next step is obtaining key points of the good match on batik and frame from the camera based on the distance calculation.

3. Estimating homography using RANSAC

    After a good match of batik and frame of the camera are obtained, the next step is estimating homography using RANSAC by randomly taking at least four key point values. Homography is done if there is a picture or a frame with the same objects but with different perspectives. Meanwhile, RANSAC algorithm is used to match key points with batik pictures and frame of the camera.

Visualisation of AR

Visualisation is used to perform AR on the device display so that users will understand batik and its philosophy easily. In this phase, there is additional information about the detected batik. In this application, the visualisation is texts and pictures about the batik detected. The information displayed is from the database and the pictures are captured when the batik is detected.

Figure 5. Visualisation of AR

Update of Batik Data

In this phase, the data about batik in the database are updated. The process involves either adding or removing some information in the database.

EXPERIMENTAL DESIGN

Data Collection

The data about batik was obtained from a book and a respondent who was a lecturer at Faculty of Technology and Vocational Education, Universitas Pendidikan Indonesia through an interview about the philosophy of batik. Table 2 shows data about batik from the book and the interview.

Table 2
Data about batik
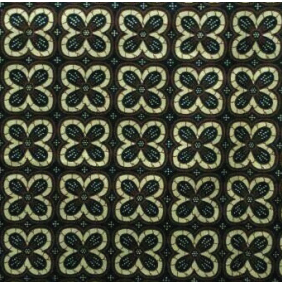
No	Name	Description	Picture
1	Cirendeu Pattern	Cirendeu pattern is a batik from Cimahi City. The pattern represents culture of Cirendeu village people whose staple food is cassava. Characteristics of Cassava are : a. It can be grown in the narrow land The philosophy is we have to be visionaries. b. Cassava is a versatile plant The philosophy is whatever our circumstances, we must remain useful for others. c. Simple food The philosophy is living a simple life, far from being consumerism. d. Cassava tubers are most sought-after even though it grows underground. The philosophy is your goal come from the bottom if you have intelligence or ability.	
2	Kawung Pattern	Kawung pattern is inspired by sugar palm fruit from palm tree trunks. All parts of the palm tree from the top (tip of the leaf) until the roots have benefits for humans. The symbol means humans must be useful to others. Visually, kawung pattern (sugar palm fruit) is a symbol of unity, which is four dots in a line of fraternity. Four circle or square is a form of unification of elements which are aligned. Kawung name means that life will return to Suwung. The Batik kawung pattern at the time of yore is often used to cover the dead. Kawung pattern is inspired by the tree which means people always remember their origins. Kawung pattern describes leaders who with abstinence can fortify their conscience so that there is harmony in the conduct of human life.	
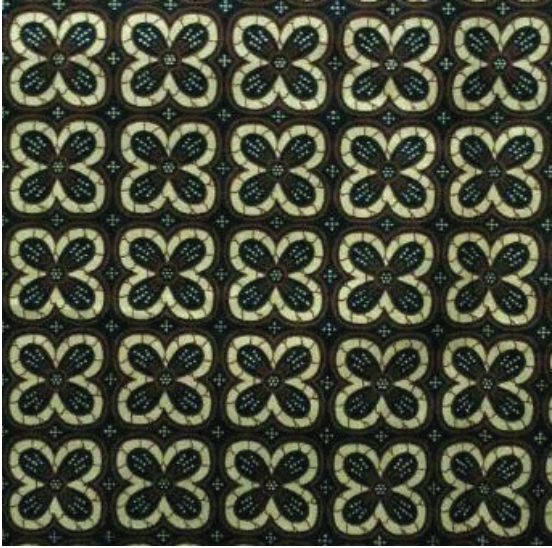
Table 2 *(continue)*

No	Name	Description	Picture
3	SPBS (Sumedang Puseur Budaya Sunda) Pattern	SPBS pattern has two explanations. First, "DINA BUDAYA URANG NAPAK", that means Sumedang people have willpower and strong commitment to carry out the preservation and development of Sunda culture. Second, "TINA BUDAYA URANG NGAPAK ", that means Sumedang people will utilise Sunda cultural resource of theirs as effective media to realise Sumedanag vision to be prosperous, religious, and democratic (SUMEDANG SEHATI). The SPBS logo is inspired by Cangkok Wijaya Kusumah flower which shows that in order to achieve lofty ideas, wisdom is needed. a wisdom in will's structure of "Silih Asah" from domain "Resi" that is public figures in the province. Kujang means in order to achieve lofty ideas, = wisdom is needed and represented by Silih Asah" from domain "Rama" that is that is a public figure in the province. Manuk Julang represents that in order to achieve lofty ideas and realise success in life, struggle and sacrifice are essential.	
4	Isola Park Pattern	Isola Park pattern represents the early history of the university. Visually, there are a Isola building, a banyan tree, and inscriptions of establishment dating back to 1954. Repetition of pattern is done to form rereng. The philosophy of this pattern represents the early history of the university. A solid Isola building represents a place to gain knowledge for people. A tip of banyan tree represents a will and a hope that always grows, a tip of banyan represented symmetrically to left and to right means a balance in various aspects of life. Inscriptions establishment PTPG 1954 means that we must inget kana purwadaksi and a past experience as the best teacher. A use of analogous colours means a harmony.	

Scenarios

Experiment was done using batik materials with the size of 60 cm x 50 cm. The tested batik is shown in Table 3.

Table 3
The tested batik

No	Name	Object
1	Cirendeu pattern	
2	Kawung pattern	

Meanwhile, the device used was a tablet with the following specifications: 5-megapixel camera, 1024 x 576 camera resolution, Jelly Bean 4.2.2 operating system, A 60-cm ruler, A protractor. There were two conditions in the experiment: The measurement distance of detection of this experiment is ±10 cm, ±20 cm, ±30 cm, ±40 cm, and ±50 cm. This is done to find out how far the application can detect batik. The slope angle is this experiment is ±0 degree, ±15 degree, ±30 degree, and ±45 degree. This is also done to find out how far the application can detect batik in different positions.
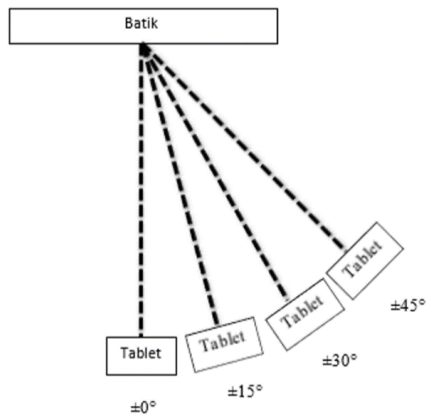
Figure 6. Angle position from above

Figure 7. The experimental test

RESULTS AND DISCUSSION

Results

Table 4 contains the results of the experiment which record time, frame per second, and detection (whether it is detected or not).

Table 4
Experimental results

No	Name	Object	Distance (cm)	Slope Angle (degree)	Time (second)	FPS	Detection
1	Kawung pattern		±10	±0	-	3,58	Undetected
			±10	±15	-	3,58	Undetected
			±10	±30	-	3,58	Undetected
			±10	±45	-	3,58	Undetected
			±20	±0	5,8	3,58	Detected
			±20	±15	5,6	3,58	Detected
			±20	±30	4,4	3,58	Detected
			±20	±45	-	3,58	Undetected
			±30	±0	2	3,58	Detected
			±30	±15	3,3	3,58	Detected
			±30	±30	3	3,58	Detected
			±30	±45	-	3,58	Undetected
			±40	±0	2,1	3,58	Detected
			±40	±15	2,2	3,58	Detected
			±40	±30	1,6	3,58	Detected
			±40	±45	-	3,58	Undetected
			±50	±0	2,1	3,58	Detected
			±50	±15	2,9	3,58	Detected
			±50	±30	2,5	3,58	Detected
			±50	±45	-	3,58	Undetected
2	Kawung pattern		±10	±0	1	3,58	Detected
			±10	±15	5	3,58	Detected
			±10	±30	4,3	3,58	Detected
			±10	±45	-	3,58	Undetected
			±20	±0	-	3,58	Undetected
			±20	±15	-	3,58	Undetected
			±20	±30	-	3,58	Undetected
			±20	±45	-	3,58	Undetected
			±30	±0	-	3,58	Undetected
			±30	±15	-	3,58	Undetected
			±30	±30	-	3,58	Undetected
			±30	±45	-	3,58	Undetected
			±40	±0	-	3,58	Undetected
			±40	±15	-	3,58	Undetected
			±40	±30	-	3,58	Undetected
			±40	±45	-	3,58	Undetected
			±50	±0	-	3,58	Undetected
			±50	±15	-	3,58	Undetected

Table 4 *(continue)*
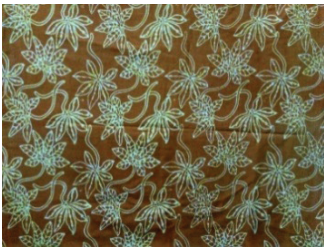
No	Name	Object	Distance (cm)	Slope Angle (degree)	Time (second)	FPS	Detection
			±50	±30	-	3,58	Undetected
			±50	±45	-	3,58	Undetected
3	Motif Cirendeu		±10	±0	-	3,58	Undetected
			±10	±15	-	3,58	Undetected
			±10	±30	-	3,58	Undetected
			±10	±45	-	3,58	Undetected
			±20	±0	-	3,58	Undetected
			±20	±15	-	3,58	Undetected
			±20	±30	-	3,58	Undetected
			±20	±45	-	3,58	Undetected
			±30	±0	-	3,58	Undetected
			±30	±15	-	3,58	Undetected
			±30	±30	-	3,58	Undetected
			±30	±45	-	3,58	Undetected
			±40	±0	5	3,58	Detected
			±40	±15	2,6	3,58	Detected
			±40	±30	5,4	3,58	Detected
			±40	±45	-	3,58	Undetected
			±50	±0	4,1	3,58	Detected
			±50	±15	6,9	3,58	Detected
			±50	±30	8,3	3,58	Detected
			±50	±45	-	3,58	Undetected
4	Motif Cirendeu		±10	±0	-	3,58	Undetected
			±10	±15	-	3,58	Undetected
			±10	±30	-	3,58	Undetected
			±10	±45	-	3,58	Undetected
			±20	±0	9,4	3,58	Detected
			±20	±15	7,4	3,58	Detected
			±20	±30	7,6	3,58	Detected
			±20	±45	-	3,58	Undetected
			±30	±0	8	3,58	Detected
			±30	±15	7,5	3,58	Detected
			±30	±30	8.8	3,58	Detected
			±30	±45	-	3,58	Undetected
			±40	±0	-	3,58	Undetected
			±40	±15	-	3,58	Undetected
			±40	±30	-	3,58	Undetected
			±40	±45	-	3,58	Undetected
			±50	±0	-	3,58	Undetected

Table 4 *(continue)*

No	Name	Object	Distance (cm)	Slope Angle (degree)	Time (second)	FPS	Detection
			±50	±15	-	3,58	Undetected
			±50	±30	-	3,58	Undetected
			±50	±45	-	3,58	Undetected
5	Cirendeu pattern		±10	±0	-	3,58	Undetected
			±10	±15	-	3,58	Undetected
			±10	±30	-	3,58	Undetected
			±10	±45	-	3,58	Undetected
			±20	±0	6,3	3,58	Detected
			±20	±15	5,7	3,58	Detected
			±20	±30	5	3,58	Detected
			±20	±45	-	3,58	Undetected
			±30	±0	7	3,58	Detected
			±30	±15	7,8	3,58	Detected
			±30	±30	8	3,58	Detected
			±30	±45	-	3,58	Undetected
			±40	±0	-	3,58	Undetected
			±40	±15	-	3,58	Undetected
			±40	±30	-	3,58	Undetected
			±40	±45	-	3,58	Undetected
			±50	±0	-	3,58	Undetected
			±50	±15	-	3,58	Undetected
			±50	±30	-	3,58	Undetected
			±50	±45	-	3,58	Undetected

The number of experiments conducted was 100. The number of detected and undetected experiments were 33 and 67 respectively.

The curation calculation in this experiment is as follows.

$$whole\ accuracy = \frac{The\ number\ of\ detected\ experiments}{The\ number\ of\ total\ experiments}\ x\ 100\%$$

Thus, $whole\ accuracy = \frac{33}{100}\ x\ 100\% = 33$

The average accuracy on the distance of ±10 cm from every slop angle is:

$$\frac{3}{20}\ x\ 100\% = 15\%$$

The average accuracy on the distance of ±20 cm from every slop angle is:

$$\frac{9}{20} \, x \, 100\% \, = 45\%$$

The average accuracy on the distance of ±30 cm from every slop angle is:

$$\frac{9}{20} \, x \, 100\% \, = 45\%$$

The average accuracy on the distance of ±40 cm from every slop angle is:

$$\frac{6}{20} \, x \, 100\% \, = 30\%$$

The average accuracy on the distance of ±50 cm from every slop angle is:

$$\frac{6}{20} \, x \, 100\% \, = 30\%$$

The average accuracy on the angle of ±0 degree on every distance is:

$$\frac{11}{20} \, x \, 100\% \, = 44\%$$

The average accuracy on the angle of ±15 degree on every distance is:

$$\frac{11}{20} \, x \, 100\% \, = 44\%$$

The average accuracy on the angle of ±30 degree on every distance is:

$$\frac{11}{20} \, x \, 100\% \, = 44\%$$

The average accuracy on the angle of ±45 degree on every distance is:

$$\frac{0}{20} \, x \, 100\% \, = 0\%$$

DISCUSSION

A series of experiments with different distances and angles were conducted in this study. The FPS in the detection was 3.58 meaning that it is considered slow since the camera usually is at 15.78 in the normal condition. This is due to the fact that every frame has to be processed and matched with every picture in the database. The data shift process in Java with library used is also influential in the FPS work since the library used is JNI based on C++.

The experiment has 33% accuracy . The calculation of the accuracy is done in every distance ranging from 10 cm with the accuracy of 15%, 20% with 45% accuracy, 30 cm with 45% accuracy, 40 cm with 30% accuracy, and 50% with 30% accuracy. Not every object with

slope angle of ±45 degree are detectable since homography is not able to handle the object perspectives at the that degree. The objects could only be detected at ±30 degree. At ±0 degree, the accuracy is 44%, at ±15 degree, the accuracy is 44%, and at ±30 degree, the accuracy is 44%.

The discussion of the experimental results of the study is listed below.

1. Object 1

   Object 1 detects all Kawung patterns. Experiments at ±10 cm could not detect the batik since the frame could not reveal the whole object picture. However, at ±20 cm, ±30 cm, ±40 cm, and ±50 cm, the object could be detected.

2. Object 2

   Object 2 detects Kawung pattern. Experiments at ±10 cm could be detected. Objects at ±20 cm, ±30 cm, ±40 cm, and ±50 cm could not be detected since the distance is too far.

3. Object 3

   Object 3 detects all Cireundeu patterns. Experiments at ±10 cm, ±20 cm, and 3±0 cm could not detect the batik due to the inability of the camera frame to capture all the objects. However, experiments at ±40 cm and ±50 cm could be detected.

4. Object 4

   Object 4 detect Cireundeu pattern. Experiments at ±10 cm could not detect the batik since the frame could not capture the whole picture. Experiments at ±40 cm and ±50 cm could not also detect it due to the same reason. Experiments at ±20 cm and ±30 cm are proven to be able to detect the batik.

5. Object 5

   Object 5 also detects Cireundeu patterns. Experiments at ±10 cm could not detect it since the frame could not capture the whole pattern. Experiments at ±40 cm and ±50 cm also failed due to the far distance. However, experiments at ±20 cm and ±30 cm are successful.

CONCLUSION

Below is a summary of the implementation of augmented reality based on android to visualise the philosophy and patterns of batik:

1. The implementation involves ORB algorithm, hamming distance, homography, and RANSAC. ORB algorithm is used to detect and describe the picture key points. Hamming distance is used to match the results of ORB from the database and the camera frame. Homography is used to overcome if there is any picture and frame with the same object but different perspectives. RANSAC algorithm is used to math key points of the batik picture and the camera frame.

2. The results of the application using a mobile device could detect five patterns out of two batik objects with the accuracy of 33% out of 100 experiments. However, it had been proven that the FPS camera worked slowly due to the distance and angle of the detection. This affect the time length of the detection.

REFERENCES

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385.

Bradski, G., Kaehler, A., & Pisarevsky, V. (2005). Learning-based computer vision with intel's open source computer vision library. *Intel Technology Journal, 9*(2), 119-130. doi: 10.1535/itj.0902.03

Du, Y., Miao, Z., & Cen, Y. (2014). Markless augmented reality registration algorithm based on ORB. In Y. Baozong, R. Quiqi, & T. Xiaofang (Eds.), *Proceedings of the 12th International Conference on Signal Processing* (pp. 1236–1240). HangZhou, China: Institute of Electrical and Electronics Engineers, Inc.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381-395.

Gargenta, M., & Nakamura, M. (2014). *Learning android: Develop Mobile apps using java and eclipse* (2nd Ed.). USA: O'Reilly Media, Inc.

Kriegman, D. (2007). *Homography estimation.* Retrieved from https://cseweb.ucsd.edu/classes/wi07/cse252a/homography_estimation/homography_estimation.pdf

Marneanu, I., Ebner, M., & Roessler, T. (2014). Evaluation of augmented reality frameworks for android Development. *International Journal of Interactive Mobile Technologies, 8*(4), 37–44.

Muhammad, A. (2015). *OpenCV android programming by example.* Birmingham, UK: Packt Publishing Ltd.

Neri, R. B., García, G. M., Barón, H. D. B., & Crespo, R. G. (2013). Annotation and visualization in android: An application for education and real time information. *IJIMAI, 2*(2), 7–12.

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In S. Lin, B. Schiele, S. Soatto, & P. Sturm (Eds.), *Proceedings of the International Conference Computer Vision* (pp. 2564–2571). Barcelona, Spain: Institute of Electrical and Electronics Engineers, Inc. Retrieved from https://ieeexplore.ieee.org/document/6126544/

Shoaib, H., & Jaffry, S. W. (2015). A survey of augmented reality. In *International Conference on Virtual and Augmented Reality (ICV AR 2015)* (p. 34). Macquarie University, Sydney, Australia.

Syahputra, R., & Soesanti, I. (2016). Application of green energy for batik production process. *Journal of Theoretical and Applied Information Technology, 91*(2), 249-256.

Widiaty, I., Riza, L. S., Danuwijaya, A. A., Hurriyati, R., & Mubaroq, S. R. (2017). Mobile-based augmented reality for learning 3-dimensional spatial batik-based objects. *Journal of Engineering Science and Technology, 12*, 12-22.

MetaheuristicOpt: An R Package for Optimisation Based on Meta-Heuristics Algorithms

Lala Septem Riza*, Iip, Eddy Prasetyo Nugroho and Munir

Department of Computer Science Education, Universitas Pendidikan Indonesia, JL. Setiabudhi, 40154, Bandung, Indonesia

ABSTRACT

Optimisation, which is a method to obtain optimal or near-optimal values of objective functions, has been widely used to make a decision in many problem domains, such as engineering, chemical, business, etc. This research is aimed to build an R package that implements 11 methods based on meta-heuristics methods that are inspired by natural phenomena and animal behaviours. Here, R programming language is considered since it is a popular programming language for data science. In this version of the package, 11 meta-heuristic algorithms are implemented, namely particle swarm optimisation (PSO), ant lion optimizer (ALO), grey wolf optimizer (GWO), dragonfly algorithm (DA), firefly algorithm (FFA), genetic algorithm (GA), grasshopper optimisation algorithm (GOA), moth flame optimizer (MFO), sine cosine algorithm (SCA), whale optimisation algorithm (WOA), and harmony search (HS). The methods have proven to be reliable and stable. To validate the package, the study presents 13 benchmarking functions in our experiments such as sphere model, Schwefel's Problem 2.22, Generalised Rosenbrock's Function and Step Function. Based on the experiments, package metaheuristicOpt produces optimal solutions as indicated by references proposing respective algorithms.

Keywords: Meta-heuristics algorithm, optimisation, R programming language, software library, Swarm intelligence

INTRODUCTION

Humans make decision every day based on its optimal output. For example, to visit a specific place from our current location, we choose the best by considering many factors, such as distance, width of street, traffic condition, type of transportation, safety, and many others. Many complicated problems are solved by optimisation methods, such as molecular biology (Festa, 2007), electrics measurements

and electrical engineering (Sergeyev, Daponte, Grimaldi, & Molinaro, 1999), optimal solar sail steering (Dachwald, 2004).

Additionally, to solve various real-world problems, many researchers have proposed many methods. These algorithms can be divided into two groups according to the precision of solutions, such as exact and approximate methods (Talbi, 2009). Algorithms need a software library so that scientists and engineers can utilise them easily. These packages are very helpful for others since not every scientist and engineer has the capability to make a program and sometimes they just need to solve their problems without needing to know the methods in detail.

Therefore, this research is aimed at developing a software library (i.e., a package) that implements 11 algorithms for dealing with optimisation tasks. These algorithms are included in the approximate method, which is population based meta-heuristics. The following algorithms are considered in this paper: particle swarm optimisation (PSO), ant lion optimizer (ALO), grey wolf optimizer (GWO), dragonfly algorithm (DA), firefly algorithm (FA), genetic algorithm (GA), grasshopper optimisation algorithm (GOA), moth flame optimizer (MFO), sine cosine algorithm (SCA), whale optimisation algorithm (WOA), and harmony search (HS).

The package developed was written in R programming language (Ihaka & Gentleman, 1996). It is not only programming language but also an ecosystem that provides over 8000 packages for implementing many methods such as machine learning, natural language processing, optimisation, etc. For example, in machine-learning domain, we can find the following packages: frbs (Riza, Bergmeir, Herrera, & Benítez Sánchez, 2015), RoughSets (Riza et al., 2014), and gradDescent (Riza, Nasrulloh, Junaeti, Zain, & Nandiyanto, 2016a). For dealing with the optimisation task, we can find several packages, such as Rmalschains and DEoptim. Additionally, R programming language is considered in this research since according to a survey conducted by KDnuggets (Piatetsky, 2017), it is most popular programming language for data science in 2013, 2014, 2015, and 2016. It means the package can be possibly used by many users in data science.

The paper is organised as follows. Population based meta-heuristics algorithms is discussed in Section 2 while Section 3 explains the package architecture developed in this research. The experimental design to validate the proposed package is shown in Section 4 while Section 5 discusses results and Section 6 summarises and concludes the paper.

METHODS

Population Based Meta-Heuristics Algorithms for Optimisation

Basically, optimisation is a process to find best solutions from sets of alternatives. In other words, given an objective function (f) to be minimised (i.e., a minimisation problem) or maximised (i.e., a maximisation problem), we need to find x_0 such that $f(x_0) \leq f(x)$ for all x in sets of solutions for minimisation or such that $f(x_0) \geq f(x)$ for all x in sets of solutions in the search space for maximisation. Detailed information regarding an introduction to optimisation can be found in (Pedregal, 2006). Furthermore, there are many approaches that can be used for dealing with optimisation tasks. According to the research conducted by Talbi (2009), these methods can be divided into two kinds: exact methods and approximate methods. The approximate methods mean that optimal solutions are not guaranteed to be obtained as in

the exact algorithms. In this research, the focus is on approximate methods, especially on population based meta-heuristics.

Recently, many methods related to population based meta-heuristics have been introduced. A method that utilises meta-heuristics and selects a solution from a population can be stated as population based meta-heuristics. They are mostly inspired by natural phenomena. Figure 1 shows general pseudo code applied in population based meta-heuristic methods.

Input: Objective function (f(x))

Output: Best solution

Algorithm:

    1. Generate initial population (P_0)

    2. Evaluate each candidate solution in P_0

    3. Select the best solution

    4. **while** $t < maxIteration$

        a. Update new population (P_t) by considering defined operators

        b. Evaluate each candidate solution in P_t

        c. Update the best solution

        d. $t = t + 1$

    5. **end while**

Figure 1. General pseudo code in population meta-heuristics methods

This research considers 11 algorithms as follows:

1. Particle Swarm Optimisation (PSO) (Poli, Kennedy, & Blackwell, 2007): It was proposed by Kennedy and Eberhart in 1995. In order to update a new population, the method uses equations for updating on the velocity and position of particles. An example of PSO implementations is for determining pressure distribution on water pipeline networks (Riza, Azmi, Rahman, & Sidarto, 2016b).

2. Ant Lion Optimiser (ALO) (Mirjalili, 2015a): It is inspired by smart behaviors of antlion (i.e., Myrmeleontidae) when hunting ants. In this method, random walk is used to update a new position, which is a new population.

3. Grey Wolf Optimiser (GWO) (Mirjalili, Mirjalili, & Lewis, 2014): It is inspired by behaviors of grey wolves (i.e., Canis lupus) in hunting techniques: searching, encircling, and attacking. The wolves involved are the leaders (i.e., alpha), second level (i.e., beta), and the lowest level (i.e., omega).

4. Dragonfly Algorithm (DA) (Mirjalili, 2016a): In this method, the following concepts are considered, as follows: separation, alignment, cohesion, resources, and predator, to update a new population.

5. Firefly Algorithm (FFA) (Yang, 2009): It is inspired by tropical firefly. Two essential rules are involved in these methods: to calculate attractiveness s and to evaluate new solutions and update light intensity/brightness. These methods have been used for dealing with many optimisation tasks, such as in calculating pressure distribution on water pipeline networks (Riza, Kusnendar, Hays, & Sidarto, 2016c).

6. Genetic Algorithm (GA) (Goldberg & Holland, 1988): It is a method based on genetic evolution that has several operators: mutation, crossover, and selection processes.

7. Grasshopper Optimisation Algorithm (GOA) (Saremi, Mirjalili, & Lewis, 2017): It is inspired by Grasshoppers' behaviours, such as attraction and repulsion. These behaviors are expressed by two equations: social forces and updating position.

8. Moth Flame Optimizer (MFO) (Mirjalili, 2015b): It is inspired by the moth's natural navigation techniques as seen in nature called transverse orientation. It is used to design a mathematical model of spiral flying path of moths around artificial lights (flames).

9. Sine Cosine Algorithm (SCA) (Mirjalili, 2016b): It is based on characteristics of profiles of sinus and cosines functions. So, to explore and exploit and then to generate a new population, two functions in sinus and cosines are used.

10. Whale Optimisation Algorithm (WOA) (Mirjalili & Lewis, 2016): It is inspired by the social behaviors of humpback whales, especially on the bubble-net hunting strategy. There are some concepts involved to obtain best solution, as follows: encircling prey, bubble-net attacking method (exploitation phase), and search for prey (exploration phase).

11. Harmony Search (HS) (Geem, Kim, & Loganathan, 2001): It is inspired by the improvisation of music players to produce harmony in music.

These algorithms are considered because they have solved many tasks. Additionally, these algorithms are state of the art on optimisation methods.

Metaheuristicopt: Package Architecture, User Guide, and Example

The package proposed to implement 11 methods based on meta-heuristics methods for dealing with optimisation problems is called metaheuristicOpt. It is submitted into the Comprehensive R Archive Network (CRAN), and retrieved at https://cran.r-project.org/package=metaheuristicOpt. Additionally, users can download it free. The package contains several functions where each name is defined as the abbreviation of the methods as shown in Figure 2. It can be seen there are 11 functions representing the considered methods and 1 main function, which is `metaOpt()`, as the highest level of function. It has the following signature:

```
metaOpt(FUN, optimType = "MIN", algorithm = "PSO",
numVar, rangeVar, control = list(), seed = NULL)
```
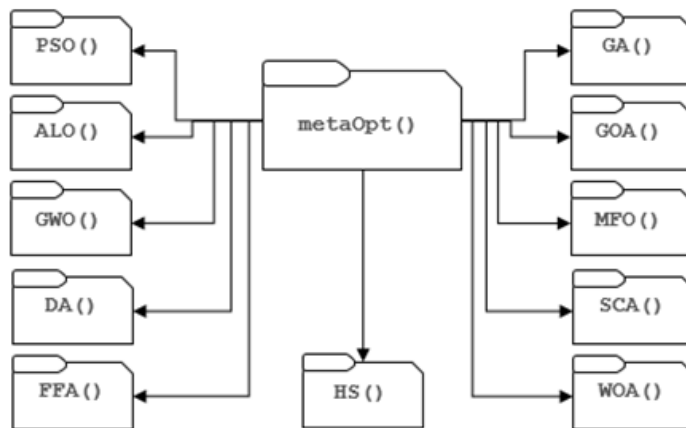
Figure 2. Functions included in the metaheuristicOpt package

So, the function has several arguments as follows:

• `FUN`: an objective function or cost function. It can be seen that the package can be used for continue problems with a single objective function.

• `optimType`: a string value that represents one of two optimisation types: "`MIN`" and "`MAX`".

• `algorithm`: a vector or single string value representing one of 11 implemented algorithms, such as "`PSO`" for Particle Swarm Optimization.

• `numVar`: a positive integer to determine the number of variables.

• `rangeVar`: a matrix (`2 x n`) containing the range of variables.

• `control`: a list containing all specific arguments of the chosen algorithm.

Detailed description regarding the arguments can be found in the manual of the package at the website.

In order to use the package metaheuristicOpt, we firstly need to install it from CRAN, and then we load it as explained in (Riza et al., 2014; Riza et al., 2015; Riza et al., 2016a). The following is an example of a user guide to run the package. For example, we need to minimise the McCormic function defined as (Surjanovic & Bingham, 2017):

Minimise: $f(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$
with $-1.5 \leq x_1 \leq 4$ and $-3 \leq x_2 \leq 4$

Figure 3 shows the code showing how to execute the function `metaOpt()` and its arguments. In this case, we use GWO algorithm. The results show that the optimum value is -1.913223 at -0.5473604 and -1.54727 for x_1 and x_2.

```
R> mcCormic <- function(x){
+   F <- sin(x[1]+x[2])+(x[1]-x[2])^2-
1.5*x[1]+2.5*x[2]+1
+   return(F)
+   }
R> optimType <- "MIN"
R> algorithm <- "GWO"
R> numVar <- 2
R> rangeVar <- matrix(c(-1.5, 4, -3, 4), nrow=2)
R> control <- list(numPopulation=30, maxIter=300)
R> result <- metaOpt(mcCormic, optimType, algorithm,
numVar, rangeVar, control, seed=1)
# to show the result
R> result
$result
            var1      var2
GWO -0.5473604 -1.54727


$optimumValue
    optimum_value
GWO     -1.913223


$timeElapsed
    user system elapsed
GWO 1.31      0    1.31
```

Figure 3. An example to use the package metaheuristicOpt

Experimental Design

To validate the package, in this research we did experimentations that involved 13 benchmark functions on optimisation. These functions can be classified into two types of optimisations: unimodal and multimodal, for minimisation tasks. The following is a list of these functions:

- Unimodal: sphere model (F1), Schwefel's problem 2.22 (F2), Schwefel's problem 1.2 (F3), Schwefel's problem 2.21 (F4), generalised Rosenbrock's (F5), step function (F6), and quartic function with noise (F7).

- Multimodal: generalised Schwefel's problem 2.26 (F8), generalised Rastrigin's function (F9), Ackley's function (F10), generalised Griewank function (F11), generalised penalized function 1 (F12), and generalised penalized function 2 (F13).

All these functions can be found in Yao, Liu, & Lin, (1999). For example, F8 to F13 are shown in Figure 4, and we can see these functions are complicated and have many solutions.
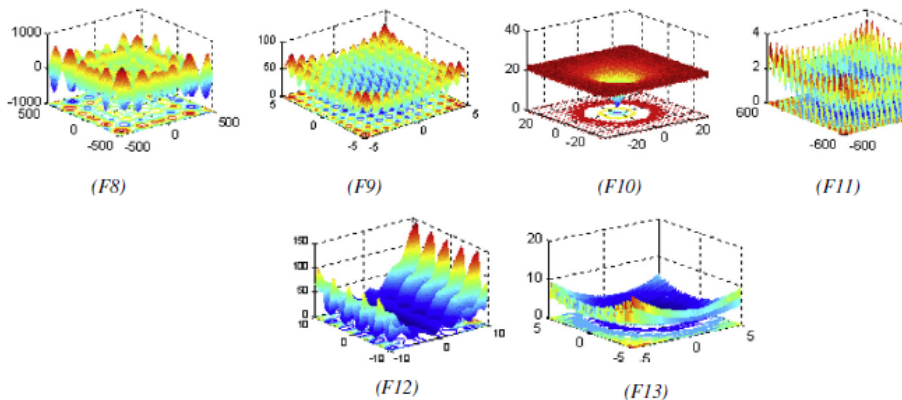
Figure 4. Multimodal benchmark function in 2D. From "Evolutionary programming made faster", by X. Yao, Y. Liu, & G. Lin, 1999, *IEEE Transactions on Evolutionary Computation, 3*(2), pp. 82-102. Copyright 1999 by IEEE Computational Intelligence Society

Furthermore, before performing simulations over all benchmark functions, we need to define all parameters required on each algorithm. In these simulations, we make some sensitivity analysis on the population size (i.e., `numPopulation`) and maximum iteration (i.e., `maxIter`) with the following values: {10, 30, 50, and 100} and {100, 500, and 1000} respectively. For other parameters, we assign the values with the default ones. So, it is obvious that we simulate 1716 times for all combinations of the algorithms, the benchmark functions and parameter values.

RESULTS AND DISCUSSION

By using scenarios based on the previous section, 1716 simulations were performed. The best solution for of each function on all the objective functions were compared and shown in Table 1. The best solution refers to values of variables that provide minimum values of the objective functions. It can be seen that PSO provides best solutions on the function F5, F6, F12, and F13 while the best solutions for F3, F4, F9, and F11 are by GWO. Additionally, best solutions of F1, F2, F8, F9, and F10 are presented by WOA and the rest are obtained by HS. In short, the algorithms mostly provide reasonable results for all benchmark functions. It should be noted that the experiments are not intended to test the algorithms, but to verify and validate that the package has been run very well when simulating the benchmarking functions. To validate the algorithms, interested readers can refer the articles proposing the corresponding algorithms.

Furthermore, comparison between the package metaheuristics and other software libraries available in CRAN is shown in Table 2. According to the table, we can state the package metaheuristicOpt offers more numbers of methods than the others. From the perspective of documentation, we have provided a comprehensive manual for common users along with many examples in the CRAN website. Lastly, the package uploaded in CRAN has the license of the GNU General Public License, so R users can also modify and improve the package freely.

Table 1
Comparison on minimum values of the objective functions of all algorithms

Benchmark Function	Algorithms										
	PSO	ALO	GWO	DA	FFA	GA	GOA	MFO	SCA	WOA	HS
F1	7.31E-25	7.06E-09	2.84E-125	7.05E-02	6.68E+02	3.26E+00	6.38E-06	4.56E-19	5.35E-16	2.37E-134	3.53E-05
F2	5.50E-14	6.68E+00	2.14E-71	2.48E+00	7.72E-01	4.22E-01	1.13E-03	3.76E-13	1.07E-08	3.29E-75	1.24E-02
F3	1.80E-17	2.25E-03	1.17E-62	3.14E+02	1.44E+03	9.96E+02	8.81E-02	5.32E-03	1.04E-04	1.23E+02	1.78E+02
F4	6.07E-11	3.10E+00	1.86E-41	2.05E+01	4.82E+01	4.52E+00	1.90E-01	7.24E+00	1.46E-05	1.98E+01	1.60E+00
F5	8.20E-02	1.63E+01	7.19E+00	2.59E+03	1.43E+03	5.92E+01	2.86E+01	7.21E-01	7.37E+00	7.18E+00	2.42E-01
F6	3.13E-25	2.76E-09	1.09E-06	3.80E+01	7.07E+02	9.82E+00	2.43E-06	9.04E-19	4.29E-01	7.74E-04	3.28E-05
F7	6.56E-01	6.19E-01	7.58E-01	9.21E-01	1.03E+00	1.03E+00	9.90E-01	9.44E-01	4.94E-01	3.54E-01	2.55E-01
F8	-1614.01	-2285.15	-2233.52	-2728.70	-2610.03	-4165.29	-2985.69	-3479.19	-2249.93	-4188.70	-4188.68
F9	6.96E+00	3.08E+01	0.00E+00	4.14E+01	7.10E+01	1.70E+00	6.77E+01	4.88E+01	2.56E-09	0.00E+00	5.34E-03
F10	1.11E-14	2.58E+00	7.55E-15	3.03E+00	8.20E+00	9.90E-02	2.32E+00	2.36E-10	2.90E-08	4.44E-16	6.34E-03
F11	8.03E-01	2.29E-01	2.45E-02	1.14E-01	1.20E+02	1.05E+00	2.23E-01	8.36E-02	4.51E-02	1.46E-01	1.18E-01
F12	1.21E-26	8.73E+00	1.57E-07	4.53E-01	2.69E+01	3.93E-01	2.58E-01	1.17E-21	1.57E-01	1.92E-02	1.05E-06
F13	1.45E-27	7.58E-08	1.54E-06	2.59E+04	2.34E+01	2.27E-01	2.11E-02	1.10E-02	5.68E-01	2.20E-02	1.60E-02

Table 2
Comparison metaheuristicOpt with other software libraries

No	Software Libraries	Numbers of Method	Algorithms	Support on Parallel Computing	References
1	MetaheuristicOpt	11	PSO, ALO, GWO), DA, FFA, GA, GOA, MFO, SCA, WOA, and HS	No	-
2	DEoptim	1	Differential Evolution	Yes	(Mullen, Ardia, Gil, Windover, & Cline, 2011)
3	hydroPSO	5	PSO and its variants: Standard PSO 2011, Standard PSO 2007, ipso, fips, and canonical PSO	Yes	(Zambrano-Bigiarini & Rojas, 2013)
4	Rmalschains	4	CMA-ES The Covariance Matrix Adaptation Evolution Strategy, SW A Solis Wets solver, SSW Subgrouping Solis Wets, Simplex	No	(Bergmeir, Molina Cabrera, & Benítez Sánchez, 2016)
5	NMOF	5	Differential Evolution, Genetic Algorithm, Local Search, Particle Swarm Optimisation, Threshold Accepting	Yes	(Gilli, Maringer, & Schumann, 2011)

CONCLUSION

The package metaheuristicOpt, providing the implementation of 11 algorithms included in population based meta-heuristics for optimisation, has been developed. It can be downloaded from CRAN at https://cran.r-project.org/package=metaheuristicOpt. Moreover, 13 benchmark functions in optimisation are used to test the package. The results show that all solutions obtained are reasonable. Therefore, the package can be used as an alternative software library for dealing with optimisation tasks. Moreover, in the future, the researchers plan to improve the package by adding other methods. Supporting parallel computing can be also considered in the next research.

REFERENCES

Bergmeir, C. N., Molina Cabrera, D., & Benítez Sánchez, J. M. (2016). Memetic algorithms with local search chains in R: The Rmalschains package. *Journal of Statistical Software, 75*(4), 1-33.

Dachwald, B. (2004). Optimization of interplanetary solar sailcraft trajectories using evolutionary neurocontrol. *Journal of Guidance, Control, and Dynamics, 27*(1), 66-72.

Festa, P. (2007). On some optimization problems in molecular biology. *Mathematical Biosciences, 207*(2), 219-234.

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation, 76*(2), 60-68.

Gilli, M., Maringer, D., & Schumann, E. (2011). *Numerical methods and optimization in finance.* USA: Academic Press is an imprint of Elsevier.

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning, 3*(2), 95-99.

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299-314.

Mirjalili, S. (2015a). The ant lion optimizer. *Advances in Engineering Software, 83*, 80-98.

Mirjalili, S. (2015b). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems, 89*, 228-249.

Mirjalili, S. (2016a). Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Computing and Applications, 27*(4), 1053-1073.

Mirjalili, S. (2016b). SCA: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems, 96*, 120-133.

Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software, 95*, 51-67.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software, 69*, 46-61.

Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). 'DEoptim': An R package for global optimization by differential evolution. *Journal of Statistical Software, 40*(6), 1-26.

Pedregal, P. (2006). *Introduction to optimization.* New York, NY: Springer Science and Business Media.

Piatetsky, G. (2017). *New leader, trends, and surprises in analytics, data science, machine learning software poll*. KDnuggets™. Retrieved from http://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html.

Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm Intelligence, 1*(1), 33-57.

Riza, L. S., Azmi, A. F., Rahman, E. F., & Sidarto, K. A. (2016b). Particle swarm optimization for calculating pressure on water distribution systems. In Y. Tan, Y. Shi, & B. Niu (Eds.), *Proceedings of the International Conference in Swarm Intelligence* (pp. 381-391). Bali, Indonesia: Springer International Publishing.

Riza, L. S., Bergmeir, C. N., Herrera, F., & Benítez Sánchez, J. M. (2015). FRBS: Fuzzy rule-based systems for classification and regression in R. *Journal of Statistical Software. 65*(6), 1-30.

Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Ślezak, D., & Benítez, J. M. (2014). Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets". *Information Sciences, 287*, 68-89.

Riza, L. S., Kusnendar, J., Hays, R. N., & Sidarto, K. A. (2016c). Determining the pressure distribution on water pipeline networks using the firefly algorithm. In D. Al-Dabass, T. Achalakul, S. Prom-On, & R. Sarochawikasi (Eds.), *Proceedings of the 7th International Conference on Intelligent Systems, Modelling and Simulation* (pp. 31-36). Bangkok, Thailand: Institute of Electrical and Electronics Engineers, Inc.

Riza, L. S., Nasrulloh, I. F., Junaeti, E., Zain, R., & Nandiyanto, A. B. D. (2016a). gradDescentR: An R package implementing gradient descent and its variants for regression tasks. In *Proceedings of the International Conference on Information Technology, Information Systems and Electrical Engineering* (pp. 125-129). Yogyakarta, Indonesia: Institute of Electrical and Electronics Engineers, Inc.

Saremi, S., Mirjalili, S., & Lewis, A. (2017). Grasshopper optimisation algorithm: Theory and application. *Advances in Engineering Software, 105*, 30-47.

Sergeyev, Y. D., Daponte, P., Grimaldi, D., & Molinaro, A. (1999). Two methods for solving optimization problems arising in electronic measurements and electrical engineering. *SIAM Journal on Optimization, 10*(1), 1-21.

Surjanovic, S., & Bingham, D. (2015). *McCormick function.* Retrieved August 1, 2017, from https://www.sfu.ca/~ssurjano/mccorm.html.

Talbi, E. G. (2009). *Metaheuristics: From design to implementation.* Lille, Prancis: John Wiley & Sons.

Yang, X. S. (2009). Firefly algorithms for multimodal optimization. In O. Watanabe & T. Zeugmann (Eds.), International symposium on stochastic algorithms (pp. 169-178). Sapporo, Japan: Springer.

Yao, X., Liu, Y., & Lin, G. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation, 3*(2), 82-102.

Zambrano-Bigiarini, M., & Rojas, R. (2013). A model-independent particle swarm optimisation software for model calibration. *Environmental Modelling and Software, 43*, 5-25.

SCIENCE & TECHNOLOGY

Natural Circulation System for Advanced Fast Reactor with Lead-Bismut as a Coolant

Ade Gafar Abdullah[1]*, Zaki Su'ud[2] and Asep Bayu Dani Nandiyanto[3]

[1]*Department of Electrical Engineering Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi, Bandung, West Java, Indonesia*
[2]*Department of Physic, Institut Teknologi Bandung, Jl. Ganesha 10 Bandung, West Java, Indonesia*
[3]*Department of Chemistry Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi 229 Bandung, West Java, Indonesia*

ABSTRACT

This paper is an outome of a study on the development of the conceptual design of advanced fast reactors that have passive and inherent abilities. Safety performance indicators are evidenced by the ability of the reactor to remain under control and can operate on a new equilibrium shortly after an accident. Simulations are performed when an accident due to unprotected loss of flow (ULOF) and unprotected loss of heat sink (ULOHS) occurs at the reactor. The passive safety system is realised by the reactor vessel auxiliary cooling systems (RVACS), before its performance is evaluated. The simulation results show that the current power reactors can survive the ULOF and ULOHS accident. The optimisation of the reactor and the RVACS results in higher natural circulation level, indicating the reactors satisfy the inherent or passive safety standards. The results from the 'accident' show there is a large safety margin to the maximum temperature within the fuel, cladding, and coolant.

Keywords: Accident analysis, advance reactor, heat sink system, NPP, RVACS, ULOF, ULOHS

INTRODUCTION

One of the most important aspects in the nuclear reactor design process is the safety aspect. Advanced and accurate safety simulation must be performed before it can be constructed. Complex calculation stages are needed to create a reliable simulation model, begins with the calculation of multi group diffusion equations (Ceolin, Schramm, Vilhena, & Bodmann, 2015; Ayyoubzadeh, Vosoughi, & Ayyoubzadeh, 2012; Quintero-

Leyva, 2012), reactor kinetic (Vyskocil & Macek, 2014; Xia, Jiang, Javidnia, & Luxat, 2012) and thermal hydraulic analysis (Giannetti, Di Maio, Naviglio, & Caruso, 2016; Liu, Zhang, Lu, Wang, & Qiu, 2016; Sun, Wang, Zhang, & Su, 2016; Mesquita & Gomes Do Prado Souza, 2014). Increased safety aspects of the advanced reactors can be realized by reducing the variety of active devices. The role of natural circulation that does not depend on electrical power becomes very important (Guo, Sun, Wang, & Yu, 2016; Wang et al., 2016; Zhao, Li, Chen, Zheng, & Chen, 2015). Two commonly used terminology associated with natural circulation are passive safety systems (PSS) and inherent safety systems (ISS) (Bochkarev, Alekseev, Korsun, & Kharitonov, 2016).

The research began with designing a reactor by adopting a two-dimensional geometry r-z of SPINNOR reactor (Small Pb-Bi Cooled Non-Refueling Nuclear Reactors), with 300 MWth powers, Pb-Bi liquid metal coolant, mixed oxide fuel (UO_2-PuO_2), as well as satisfying passive and inherent safety criteria in accordance with the standard of generation IV reactor designs (Yan & Sekimoto, 2008). The high fuel temperature safety margin as well as high coolant boiling point of Pb-Bi (~ 1700°C) made it possible to design reactors with inherent safety capabilities. The details of reactor design specifications are shown in Table 1. Preference of MOX fuel (UO_2-PuO_2) on this design is based upon the consideration that the fuel has a high Doppler coefficient, a high melting point and the availability of adequate in nature. MOX fuel can easily be applied in large power levels. At low power levels, MOX fuel requires a fraction of the volume of the fuel is high. The rationales for selecting of materials Pb-Bi as coolant because it has the characteristics of a good nuclear analysis in neutronic, void reactivity, and the breeding ratio. It is possible to design a reactor core with a long operating time without refueling (Su'ud & Sekimoto, 1995).

The design of advanced reactors is a conceptual design researched and developed. The reactor operates at a higher temperature than the previous generation, even several designs are intended for the production of hydrogen. The design is expected to represent excellence in enhancing the safety of nuclear reactors, improve resistance to proliferation, minimize waste and use of natural resources, and reduce costs through the development and operation (Su'ud, 2007). The advanced reactor must have the following criteria: innovative design that is compact, simple, modular, and more dependent on passive components (Butt, Ilyas, Ahmad, & Aydogan, 2016).

Reactor accident analysis can be realized by modifying the RETRAN code, in the case of accidents due to the increase in reactivity (Kim & Yang, 2005). The ULOF and UTOF accident simulation, also experimented on a small reactor with Pb-Bi as coolant, metal uranium and MOX as fuel. Natural circulation has succeeded in cooling the reactor. Changes in coolant temperature, fuel temperature, reactivity feedback, play an important role to achieve the capabilities' inherent safety (Su'ud, 2008). Numerically, the simulation of ULOF accident began with calculating of points kinetic equation and analyzing one phase thermal hydraulic and heat conduction calculation (Kazeminejad, 2008). Several conceptual designs that have

implemented the concept of passive safety include: Material Testing Reactors (MTR), High-Temperature Gas Cooled Reactor (HTGR), and CANDLE fast reactor (Gaheen, Elaraby, Aly, & Nagy, 2007; Hossain, Buck, Said, Bernnat, & Lohnert, 2008; Yan & Sekimoto, 2008).

Other reactors that have implemented the concept of PSS are fast reactors namely Flexible Conversion Reactor (FCR). This reactor can be described as a version of the actinide burner reactor (ABR), or minor actinide burner reactor (MABR), or actinide burner reactor with Thorium (ABRT). Each type differed only in the configuration of the reactor core and the number of control rods. The overall design of the reactor is Lead coolant and has a power 700MWth and 300MWe (Nikiforova, Hejzlar, & Todreas, 2009). The coolant of this reactor, contained completely in the reactor vessel, and the decay heat removal system using RVACS, equipped with perforated plates to improve heat transfer. Design is not equipped with a seismic isolation system, then to handle the weight of the coolant during the earthquake, the reactor vessel is thickened to 25 cm. Thickening the walls of the reactor vessel will cause a reduction in heat dissipation rate, this is corrected by adding the liquid metal in the gap between the reactor vessel and the walls of the reactor vessel. Others are related to reactor designs that implement PSS is Korean Lead-cooled fast reactor (KLFR). KLFR has a power rating 900 MWth, 372 MWe, and using the RVACS design modifications for decay heat removal systems. KLFR using the liquid metal is Lead, as the gap filler material between the reactor vessel and the walls of the reactor vessel. Although the reactor heat removal system is the same as the reactor FCR, the selection of a liquid metal of lead is still widely questioned. The PSS performed on PWR reactors, 300 MW power by using software RELAPS5 / MOD.3.4. The focus from the study is to conduct a comparison type of water and the air coolant. Two types of coolant successfully made stable natural circulation system, decay heat can effectively be removed through the primary loop. For long-term cooling, the air conditioning is better than water cooling. The heat-transfer capability of a water coolant is stronger from the beginning of time (Wang, Tian, Qiu, Su, & Zhang, 2013).

The PSS on BWR reactor can be done without operator assistance, AC power supply, and the injection pump. By utilising the effect of gravity, the components of the gravity-driven cooling system (GDCS) inject water into the reactor and the reactor core. Component's isolation condenser system (ICS), and passive containment cooling system (PCCS) are used to eliminate the residual heat. Analysis of the passive safety system response to the situation loss of coolant accidents (LOCA) can be performed by simulation (Lim et al., 2014). Thermal hydraulic analysis of LOCA accident on advanced heavy water reactor (AHWR), 920 MWT power, is also conducted. This study discusses the impact of high pressure injection in the process of cooling the reactor after the accident. Passive safety features through the natural circulation occur in primary coolant with no pump. This reactor is equipped with the emergency core cooling system (ECCS) and isolation condensers (ICs) to remove the decay heat. This ECCS provides cooling for the fuel in passive mode for 15 minutes after the accident. Cooling is then performed within three days through a gravity-driven water pool (GDWP) (Tyagi,

Kumar, Lele, & Munshi, 2013). Inherent and passive safety system is an innovative design to improve the reliability of nuclear power plants. CPR1000 design is equipped with a secondary cooling system of passive residual heat removal system (PRHRS) and system core makeup tank (CMT). The PRHRS effectively manages to eliminate the residual heat while the CMT is used to inject Boron water into the reactor core during an LOCA accident (Wang et al., 2013).

The concepts of integrated PSS are developed to perform various functions of passive safety. This design has a variety of functions, including: system of residual decay heats sinks, the injection of PSS, passive cooling containment system, cavitation system, passive system retention vessel, and filtering through the ventilation system to control the pressure. The safety systems are enhanced in cases when the situation reactor experiences a blackout (Chang, Kim, & Choi, 2013). In the transient state, the PSS is designed to cool down the reactor coolant from the normal state to the shutdown conditions through a natural circulation system (Dai, Jin, Wang, & Chen, 2013). Thermal-hydraulic research advances have significantly improved operating performance of nuclear reactors. Research and development aspects of thermal hydraulics are carried out using experimental and computational approaches. The use of fine-scale computational models of multi-physics aims to increase the computing power (Saha et al., 2013).

The new concept of offshore nuclear reactor enhances security features by implementing security structure based gravity (gravity-based structures). This structure is applied to the APR1400 reactor, which is the latest model of the nuclear power plant in the Republic of Korea. There are new features proposed and can be directly applied to a large-scale nuclear power plant. Such features include emergency passive containment cooling system (EPCCS) and emergency passive reactor-vessel cooling system (EPRVCS). Both features use the seawater as a coolant and are also equipped with security features to withstand earthquakes, tsunamis, storms and collisions at sea. The offshore NPP is more powerful than a nuclear power plant, and conventional land has the strong potential to provide opportunities in the nuclear-power industry (Lee, Lee, Lee, Jeong, & Lee, 2013).

This study aims to provide a comprehensive safety analysis of NPP. Simulations are performed on advanced liquid metal-cooled reactor. The focus is on how to improve the reactor design performance and increase natural circulation so that it meets the criteria of passive safety. The object of study is a small reactor that has a 225 MWth; Pb-Bi is a coolant, and mixed oxide fuel (UO2- PuO2). Pb-Bi chosen because it has a heat transfer coefficient and since its boiling point is very high, it is possible to design a reactor with inherent safety capabilities.

The simulation of the NPP accident was done using quasi-static method (Wei, Sui, Ma, Jing, & Shan, 2014). The accident was caused by the failure of the technical components in the reactor. Simulations are done on unprotected loss of flow (ULOF) and unprotected loss of a heat sink (ULOHS).

METHODS

Table 1
Specification of fast reactor design

Parameters		Specification
Thermal power		225 MWth
Coolant		Pb-Bi (Pb 45%-Bu 55%)
Shielding materials		B_4C + *Stainless steel*
Fuel		MOX (UO_2-PUO_2)
Fuel enrichment		10%~13% (PuO_2)
Reactivity Swing		Max. 0.002 (0.5$)
Reactor Core	Diameter of fuel pin	1.0 cm
	Cladding Thickness	0.05 cm
	Pin pitch of fuel	1.2 cm
	Average of temperature input	400°C
	Average of temperature output	800°C
	Volume fraction : Fuel	60%
	Structure	10%
	Coolant	35%
Steam Generator	Height of SG	4.0 m
	Pipe diameter (*inner/outer*)	1.7 / 2.0 cm
	Pitch of pipe	3 cm
	Primary flowrate	8000kg/s
	Secondary flowrate	100 kg/s
	Water temperature	225°C

The reactor designs used as research subjects are based on the basic concept of the reactor SPINNOR (Small Pb-Bi cooled No on-site refueling Nuclear-Power Reactors). The SPINNOR aresmall-sized nuclear reactors with a fast neutron spectrum that can be operated more than 15 years without replacing the fuel. The SPINNOR has a long life shelf live and uses little power. It was developed in Indonesia in 1990 in collaboration with the Research Laboratory for Nuclear Reactors of the Tokyo Institute of Technology. This reactor is suitable in isolated areas and small islands. It is easy to operate and maintain, transport, and meet the inherent safety/ passive, and resistant to nuclear proliferation (Su'ud, 2008). The SPINNOR reactor design adds a component RVACS to protect and enhance the reactor cooling system that operates naturally.
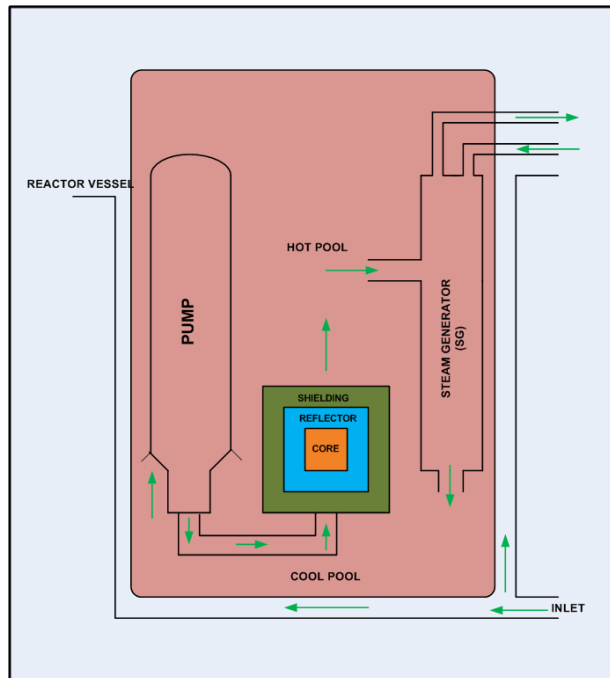
Figure 1. A schematic diagram of SPINNOR

A schematic diagram of the reactor SPINNOR is shown in Figure 1. The system consists of a reactor core, hot pool, cool pool, steam generators and pumps. In this system, there is no intermediate heat exchanger (IHX), so the flow heat (green arrow) from the primary cooling system was directly transferred to the steam generator. The working principle of the heat flow reactor can be explained as follows: the coolant flow passes through the reactor core to take the heat from the reactor core, after that moves up into the hot-pool. From the hot-pool, the coolant flow enters the generator steam and heat transfer, and then down to the cool pool. Cool-pool of coolant is pumped back into the reactor core. The thermal hydraulics system of SPINNOR has two flow circulation cooling system, which is the primary coolant and secondary coolant. Figure 2 shows a schematic flow of coolant. There is a pump that is used to provide forced circulation to cool the reactor core. When the reactor faces loss of cooling capability, its pump power driving cooling becomes inhibited.
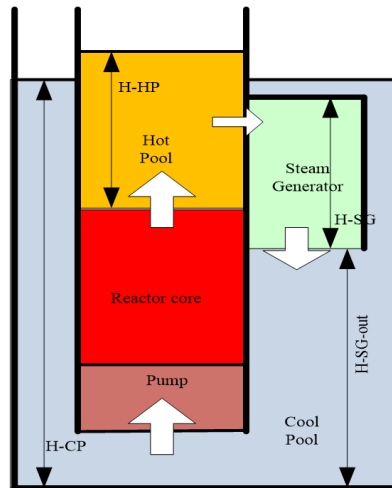
Figure 2. A scheme of coolant flow rate. From "Safety analysis of small long life CANDLE fast reactor", by Yan and Sekimoto, 2008, *Annals of Nuclear Energy, 35*(5), pp. 813-828. Copyright 2008 by Elsevier

The reactor design processes begins with neutronic analysis to get the neutron flux distribution and power distribution in a steady-state nuclear reactor. The calculation is performed using a multigroup diffusion equation. The discretisation processing diffusion equation is carried out using the finite difference method, and then solved using SOR (Successive Over Relaxation) method and inner-outer iteration scheme (Yulianti, Su'ud, Waris, & Khotimah, 2009). The calculation result of this neutronic analysis would provide information about reactor criticality level and neutron flux distribution.

The dynamics of neutron population in the core during the transient process is determined by using nuclear reactor kinetics. The kinetics calculations give time dependent power distributions pattern during some transient event of the reactors. Then, the calculations of the thermal hydraulic are done, including temperature distribution in each part of the reactor consisting of a reactor core and steam generator, i.e. fuel temperature, coolant temperature, cladding temperature, and all aspects of coolant circulation in the reactor.

The ULOF accident simulations are performed to investigate inherent safety performance with the following processes: a reduction of flow rates due to the loss of the pumping power. This then causes coolant temperature to increase due to an imbalance between the power level and the coolant flow rate. The increase of fuel and coolant temperature triggers negative feedback that forces the decrease in the power level. The power reduction causes a decline in fuel temperature which then triggers positive feedback, and the system would reach a new equilibrium if the absolute value of negative reactivity feedback due to coolant temperature rise equals to the positive reactivity feedback due to the decrease in fuel temperature. The final equilibrium temperatures depend on the coefficient of feedback reactivity as well as its thermal characteristics, especially heat capacity of the coolant and thermal conductivity of the fuel.

The ULOHS accident simulations are performed initially by investigating RVACS (Reactor Vessel Auxiliary Cooling System) performance and its optimisation. Passive safety capability during ULOHS accident strongly depends on the portion of heat which can be removed through RVACS during the accidents. The analysis includes heat transfer i.e. radiation, convection, and conduction. The first step begins with the design process of RVACS construction and numeric calculations of RVACS performance optimisation at steady state using Newton-Raphson method. Indicators of passive safety performance are shown in percentage quantity of RVACS system in removing heat from the reactor vessel and vessel wall. The RVACS system is expected to remove the heat using natural circulation of the atmosphere around the reactor vessel at a steady state. The systems of heat transfer involved in the analysis are radiation, conduction and natural convection. The heat transfer calculations ae conducted on the reactor vessel elements, outside wall of the reactor vessel, and the separator plate (Abdullah & Nandiyanto, 2016).

The current reactor designs eliminate IHX (intermediate heat exchanger) so that the heat is transferred directly from the core to the steam generator loop by a Pb-Bi coolant. The ULOHS accident is triggered by the loss of cooling capability of steam generators. The accident sequence begins with the loss of the reactor's ability to remove heat from the secondary cooling loop. During the accident, the heat dissipation rate decreases whereas the coolant inlet (flow input) temperatures increase till a new equilibrium level is reached.

RESULTS AND DISCUSSION

Analysis of RVACS Performance

The paper analyses the performance of reactor designs in removing decay heat. It also evaluates the performance of reactor design as a result of accidents that leads to cooling capability loss. Reactor accident simulation analysis is applied in the case ULOF and ULOHS accident. The calculation of the steady-state, such as the calculation of neutron flux distribution, power distribution and reactor multiplication factor are not discussed in this paper. Calculation of RVACS performance is discussed in detail as it is used to prove the hypothesis that passive safety system in this SPINNOR reactor design will run properly.

Before the simulation is done, it is necessary to arrange the total decay heat as initial assumptions to prepare the PSS; the latter is developed by utilising the natural circulation of the atmosphere surrounding the reactor by using RVACS. The RVACS design analysis is conducted to obtain information about how much residual heat can be removed by RVACS. The optimisation is realised to receive a large maximum heat transfer so that it can boost the performance of RVACS in removing heat. Two aspects are examined: First, the characteristic aspects of thermal emissivity covering the walls of the reactor vessel to the separator plates, as well as the effect of viscosity to the conductivity of the air. Second, optimising the geometry aspects which include the impact of changes RVACS height, diameter and channel gap. The reference design is used as shown in Table 2. The analysis only covers most of the coolant temperature. If the analysis produces a cooling mass below the maximum temperature of the cladding, then the design will be considered successful, because the power level of decay across the cladding is very small and there is little thermal resistance between the cladding and the primary coolant.

Table 2
Main parameters of RVACS design

High of RVACS	40 m
Downcomer gap	5,86 m
Radius of reactor vessel	4,62 m
The thickness of reactor vessel	0,4 m
The thickness of guard vessel	0,1 m

The simulation begins with calculating the total heat transfer in the reference design. The RVACS can remove residual decay heat of 4.39 MW, or approximately 1.95% of the thermal power reactors. Decay heat removal is obtained through a process of natural circulation. Simulations are performed when the reactor is a steady state. Heat-transfer coefficient and friction factor are assumed in the case of a turbulent flow.

(a). The effect of changes in emissivity value of guard vessel and the perforated plate

(b). The effect of changes in viscosity and air conductivity

(c). The change height effects of RVACS

(d). The change diameter effect of RVACS

(e). The change channel gap effect

(f). Heat removal capabilities of RVACS
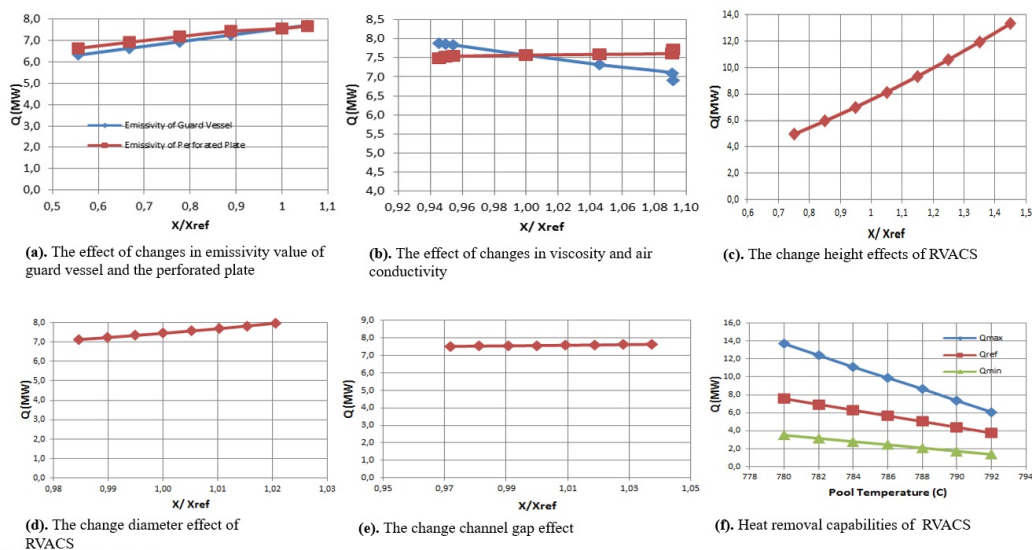
Figure 3. The result of optimisation of RVACS performance

Figure 3(a) and Figure 3(b) show RVACS performance in optimising emissivity of the reactor vessel wall and separator plates, and optimisation aspects of viscosity and air conductivity. Optimising emissivity is intended to clear the effect on the ratio in the energy radiated by the walls of the reactor vessel and the perforated plate with the energy radiated by a black body at the same temperature. This emissivity value shows the ability of the reactor vessel material and perforated plate to radiate absorbed energy. Optimising the emissivity value shows an increase in performance RVACS if the value ends at 1, while optimising the amount of viscosity and conductivity of the air does not contribute significantly to improving the heat-transfer of RVACS. The viscosity and conductivity of the atmosphere around RVACS is relatively unchangeable and is at room temperature. The simulation results will provide different conclusion if RVACS applied to the reactor is placed at the location that has an extreme climate.

Figures 3(c), 3(d) and 3(e) show the results from the optimisation of the geometry aspects. Our simulation results show that the factor altitude will affect the RVACS performance. Optimising the diameter and width of the channel gap does not alter the performance of RVACS. The heat-transfer is similar with the simulation results of reference design. Although with changing altitude of RVACS will give excellent results, a maximum limit must be determined, because it is limited by the ability of the construction of the reactor. The width of the channel gap cannot be extended more generally related to the fluid nature of gap fillers that require a very fast time to melting.

Figure 3(f) shows a result summary for the optimisation of minimum design, reference designs and a maximum design. The simulation results conclude the RVACS Design can remove the heat between 0.66% to 2.12% of the reactor's thermal power. Obtaining value of heat transfer is a result of the role of natural circulation of air around. If an accident occurs at the reactor, the PSS can remove the heat. The calculations show that the heat dissipation is at sufficient level to prevent the high temperatures provided RVACS is reliable and efficient.

Analysis of ULOF Accident

The ULOF accident occurs due to a failure in the primary loop triggering failure; the loss of pump power (coast down) P (t) is calculated using the formula (Yan & Sekimoto, 2008):

$$P(t) = \frac{P_o}{\left(1 + t/\tau\right)}$$

Where

t : time after accident (s).

P_o : steady state pump power.

τ : pump coast down halving time (s)

To evaluated to effect of halving time τ, investigated the two experiments on τ=20 s (case ULOF A) and τ=12 s (case ULOF B).

Figure 4 shows the change in the mass flow rate coolant through the reactor core for both cases. In the first 30 seconds after the accident, the decrease in coolant flow rate is very drastic. At 40 second to the end, the mass flow rate slowly declines to almost constant. In the transient state, the case ULOF-A has a mass flow rate is greater than in the case ULOF-B. When the pump power decreases from steady-state value, mass flow rate in reactor core still survives in the range of 32.5% up to 37.25%. This is due to natural circulation factor.
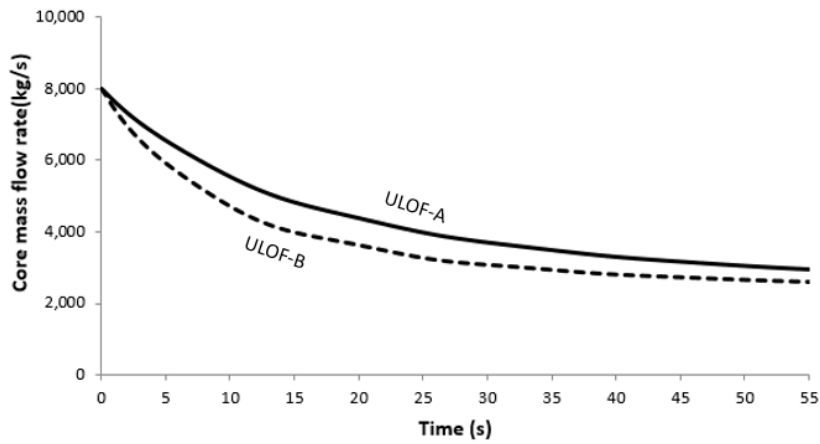
Figure 4. Change of mass flow rate at reactor core in ULOF-A and ULOF-B
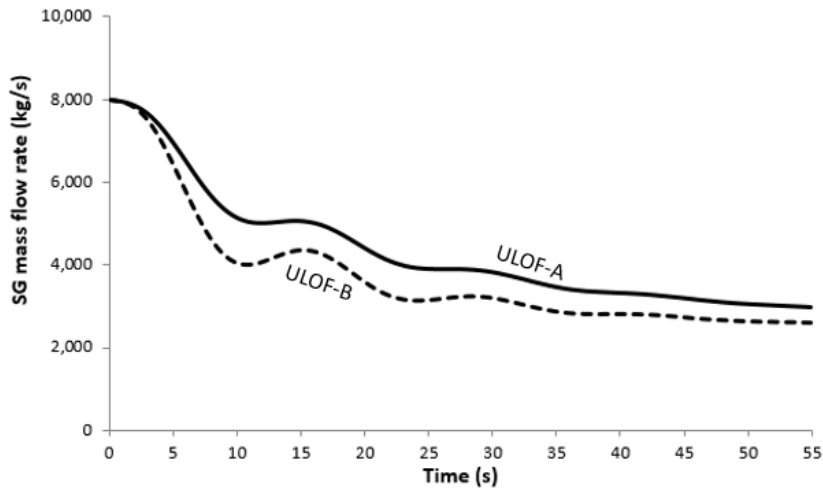
Figure 5. Change of mass flow rate at steam generator in ULOF-A and ULOF-B

Figure 5 shows the change of mass flow rate coolant in the steam generator (SG). The simulation results are almost identical to changes in mass flow rate coolant through the reactor core, but the flow rate of coolant in SG experiences a small oscillation that cause degeneration in the flow of the main pump that controls the level and coolant temperature. Natural circulation can increase the level of coolant circulation flow and the mass flow rate to survive the value of 32.25% up to 37% of the normal operating state.
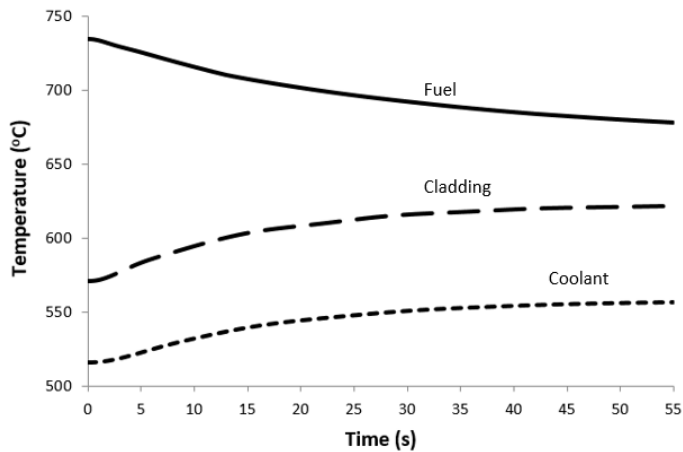
Figure 6. Hot spot temperature from fuel, cladding and coolant of case ULOH-A
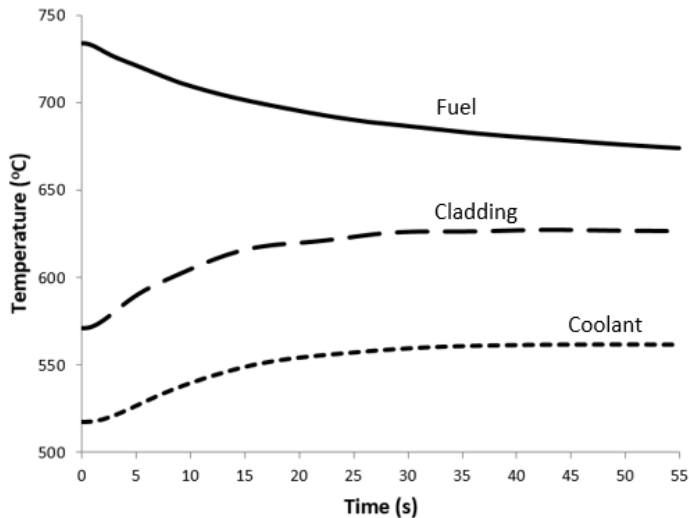
Figure 7. Hot spot temperature from fuel, cladding and coolant of case ULOH-B

Figure 6 shows the temperature profile of fuel, cladding and coolant in ULOF-A. In this case, the maximum temperature of the fuel gained 734.10°C, then declined slowly towards a constant value 678.20°C. Cladding temperature and coolant temperature respectively reach 621,64°C and 556,19°C. The temperature of the fuel, cladding, and cooling in the case or ULOF-A and ULOF-B still has a very large margin, so the reactor can survive the conditions of this accident. The threshold allows melting point of the fuel, cladding and cooling respectively at 2865°C, 2519°C and 1700°C. Usually the larger accident is caused by a high reactivity value, due to the failure of the control rod operation.

The simulation of results shows that when the flow rate of cooling declines, the coolant temperature average will increase, while the fuel temperature average declines. This occurs because of an imbalance between power and a coolant material. However, the temperature difference between a cooling material with the fuel decreases. This situation raises negative feedback causing a decrease in fuel temperature. The system will re-balance when the reactivity of the negative feedback due to the rise in coolant temperature is compensated by a positive reactivity feedback.
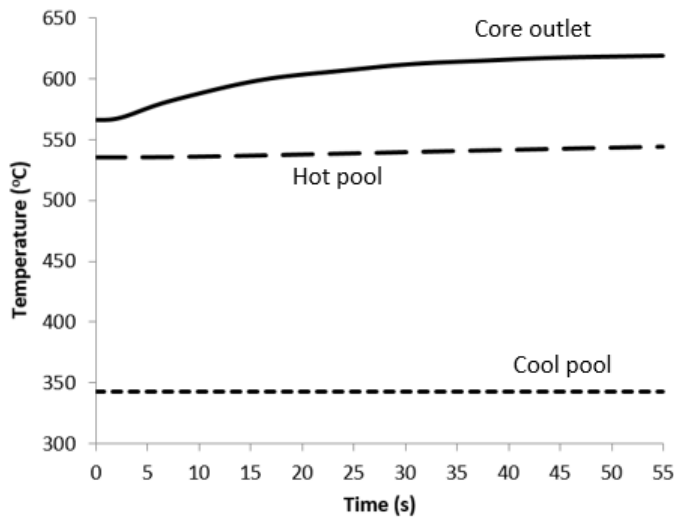
Figure 8. Temperature of hot pool, cool pool and core outlet of ULOF-A
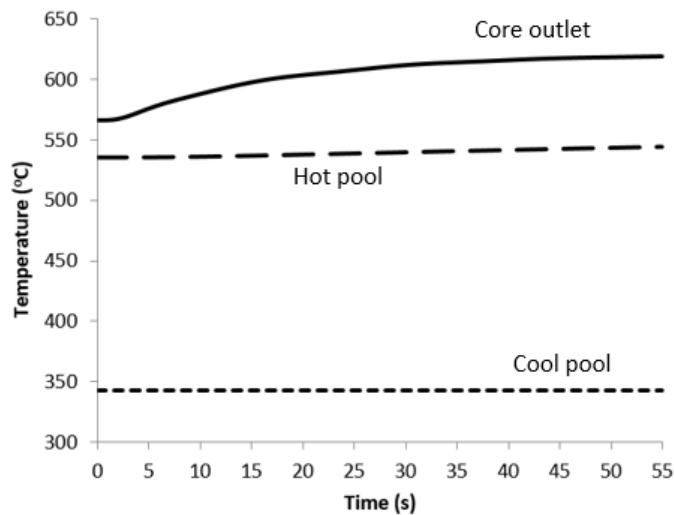
Figure 9. Temperature hot pool, cool pool and core outlet of ULOF-B

Figure 8 and Figure 9 show a thermal reactor system behaviour that occurs in the hot pool, cool pool, and core outlet in case ULOF-A and ULOF-B. The figures show the changes in the value of τ affects the transient process, while the temperature of the system is always convergent at the same value. Core outlet temperature changes tend to be faster and more intense than the temperature of the hot pool and a cool pool. Temperatures at a cool pool is relatively stable because the steam generator continues to work normally.
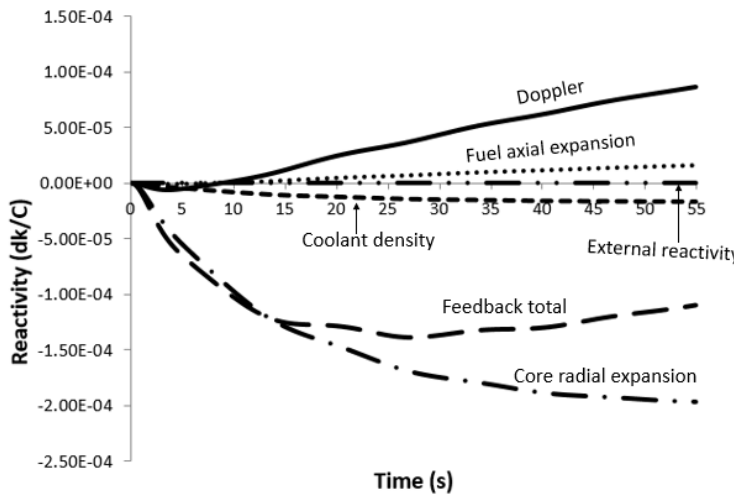
Figure 10. Reactivity change in case ULOF-A
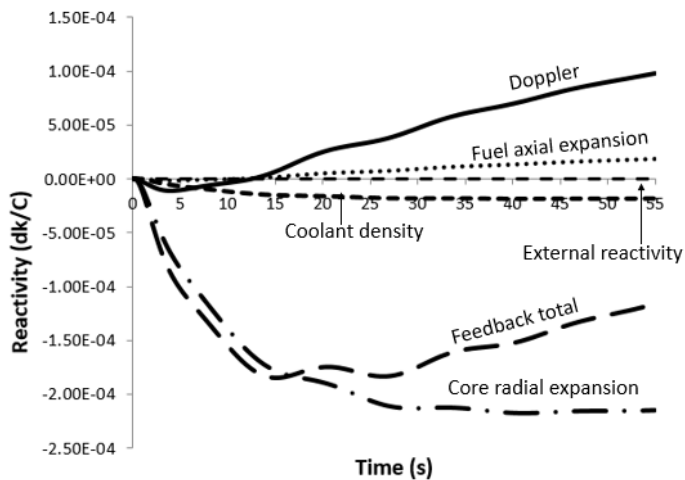
Figure 11. Reactivity change of case ULOF-B

Figure 10 and Figure 11 show the change in reactivity. Reactivity axial expansion provides the largest contribution, while reactivity contributes the smallest radial expansion. At the end of the accident, they will compensate positive reactivity of the Doppler effect and changes in coolant density. Doppler positive effect comes from the decreasing temperature cool pool.
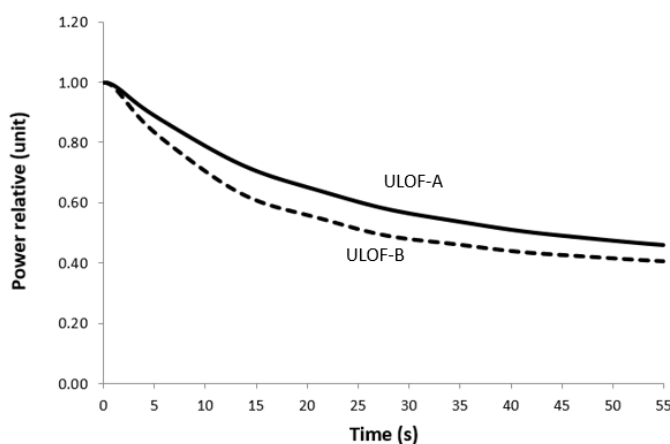
Figure 12. Thermal power relative change of case ULOF-A and ULOF-B

The change in thermal power is shown in Figure 12. Thermal power declines relatively rapidly in the first 30 seconds and then gets more slowly. Thermal power eventually converges at the value of 40.61% for cases ULOF-B and 45.87% for cases ULOF-A. The mass flow rate in the reactor core is lower, i.e. in the range of 32.5% to 37.25%, and the mass flow rate in the steam generatoris only 32.25% up to 37% of the steady state. This proves that the ability of natural circulation has an important role in this ULOF accident case. Overall, the results from this simulation show that the proposed reactor designs survive ULOF accidents. The increase in coolant temperature and outlet temperature does not pass the thermal limitations of a cooling material, structural material and fuel.

Analysis of ULOHS Accident

The reactor accident of ULOHS case is triggered by the loss of cooling capability in the steam generator (SG). After the accident occurs, the cooling capability of theSG can be observed. Figure 13 shows the mass flow rate of coolant at the reactor core and the SG. Shortly after the accident, the flow of coolant at the reactor core is less affected, but the mass flow rate is only slightly decreased. Figure 13 shows the mass flow rate of coolant through the reactor core and the steam generator. Shortly after the accident occurs, the flows of coolant through the reactor core are less affected, but the mass flow rate is only slightly decreased. The mass flow rate of the coolant at the SG looks unstable; proving that there are very large oscillations. In the first 40 seconds, a coolant mass flow rate is unstable but moments later it returns to stability, and achieves a new equilibrium. The oscillation phenomenon can be mitigated by increasing the volume of a cool pool.
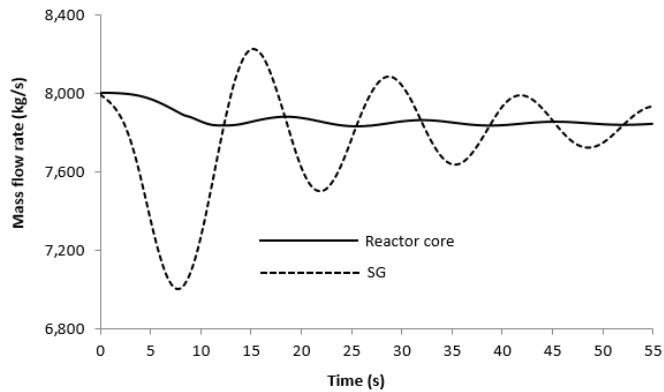
Figure 13. Changes at the coolant mass flow rate and the reactor core SG on ULOHS accident

Figure 14 shows the effect of changes in the volume of cool pool to the oscillation of the coolant flow that passes through the steam generator. Simulations are carried out by increasing the volume of cool pool randomly (the case of ULOH-B and ULOH-C). This is done to prove the increase in the volume of cool pool in the reactor will affect the rate of coolant flow oscillations at the SG. Referring to the previous RVACS design where the heat can be discharged approximately 4.39 MW or 1.95% of thermal power reactors, the results of these calculations can be used as a reference RVACS capability to remove heat from an accident ULOHS.
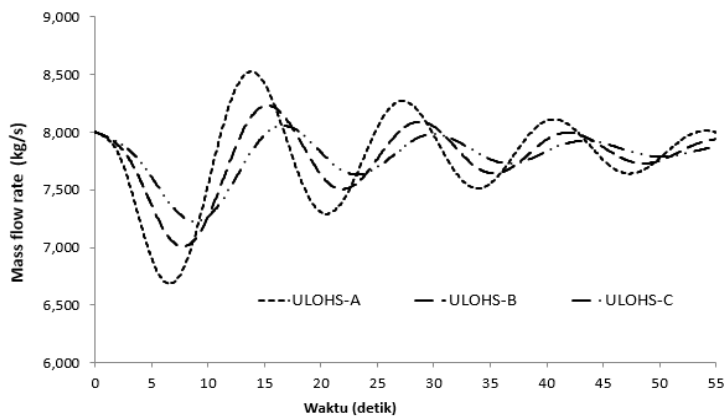
Figure 14. Changes of coolant mass flow rate at SG with various cool pool volume

Figure 15 shows changes of coolant temperature at hot pool and cool pool. Accidents cause loss of cooling capability from SG impacting on the coolant temperature at a cool pool. Shortly after an accident (for 10 seconds), the coolant temperature at a cool pool is still at normal state. But shortly afterwards, cool pool temperature increases, which leads to incapability of the SG to do the cooling. Because the design of SPINNOR reactor with cool pool volume has a large impact on ULOHS, accident can only be sensed in a expanded time, and to obtain thermal equilibrium takes a long time. Increased temperatures of cool pool impacts the rise in temperature at a hot pool. Coolant temperature goes to the reactor core value is higher than when the reactor is in steady state.
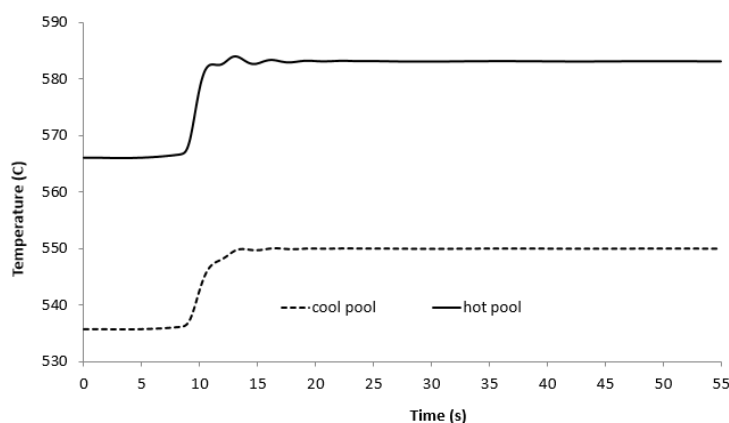
Figure 15. Changes of coolant temperature at hot pool and cool pool, after ULOHS accident

Changing temperatures of coolant, cladding and fuel after ULOHS accident does not affect a temperature hot spot in a short time, so the fuel temperature only increases slightly and temperatures tend to converge at a temperature of 815°C. Likewise, cladding and coolant temperatures do not change drastically. The RVACS is known to be capable of removing heat from the reactor vessel walls. Simulation results prove natural circulation in RVACS is able to function properly. There are differences in temperature change between the cladding and the cooler, but the difference is not so great, as shown in Figure 16. Ten seconds after the accident, temperature of coolant and cladding sudden increases, but then the temperature converge respectively t 583°C and 588°C. The temperature increase of the coolant, cladding or fuel, show they are in a safe margin, meaning that it is still far from the limit to the melting point of the material used.
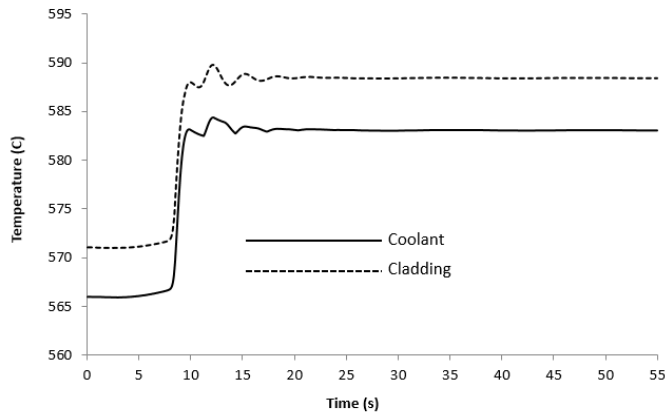
Figure 16. Changes of coolant and cladding temperature after ULOHS accident

Figure 17 show changes of thermal power relative in relation to the ULOHS accident. The thermal power at the reactor decreases quickly at 10s first before it slows down. Thermal power eventually converges towards the value of 57.16%. This proves that capability natural circulation has an important role in ULOHS accident.
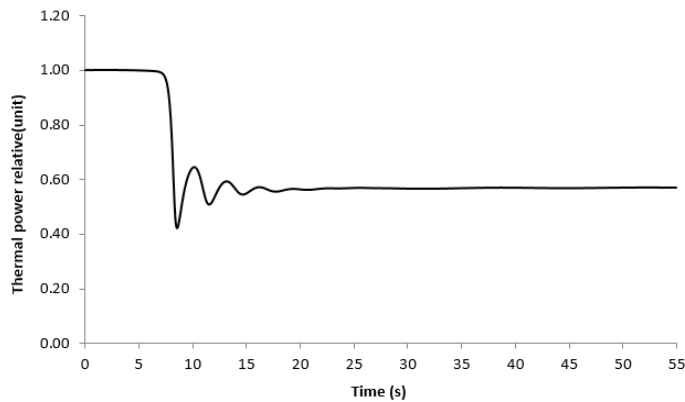
Figure 17. Changes in thermal power after ULOHS accident

Figure 18 shows changes of reactivity after ULOHS accident. This reactivity is dominated by Doppler Effect and reactivity feedback. This is due to the temperature of a cool pool increase, meaning that the temperature at the inlet also experiences substantial increases.
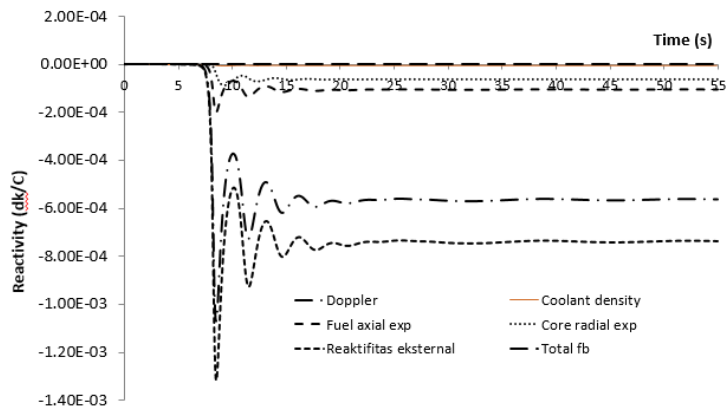
Figure 18. Changes of reactivity for case ULOHS accident

This study is only a simulation whereby ULOF accident simulations are performed to investigate inherent safety performance with the following processes: a reduction of flow rates due to the loss of the pumping power. This then causes the coolant temperature to increase due to an imbalance between power level and the coolant flow rate. Further, analysis for realistic condition will be examined by the reseachers in their future work.

CONCLUSION

In order to study the passive safety system, a system conceptual design of natural heat reduction is done using RVACS. In the reference design, RVACS can reduce 4.39 MW heat or around 1.95% of the total power of thermal reactor. This shows that when there is an accident in the reactor a while after it is shut down, the rest of the heat can be released by RVACS system. The improvement of RVACS works can actually be carried out by optimising several variables. The optimisation is done in two design aspects; first, thermal characteristic aspect that includes emissivity effect of the reactor's vessels and separate plates as well as the air viscosity and conductivity; second, optimisation on geometrical aspects covering the diameter effects, the height of RVACS, and channel gap effects. In general, simulation results show the reactor can easily survive by optimising the roles of natural circulation and passive safety system. Meanwhile, the results of accident analysis simulation due to the loss of reactor coolant show the maximum temperature, cladding, and coolant, have limitation in terms of melting points, meaning that the material reactor does not exceed its thermal limits.

REFERENCES

Abdullah, A. G., & Nandiyanto, A. B. D. (2016). Conceptual design of passive safety system for lead-bismuth cooled fast reactor. *IOP Conference Series: Materials Science and Engineering, 128*(1), 1–6.

Ayyoubzadeh, S. M., Vosoughi, N., & Ayyoubzadeh, S. M. (2012). On an improved direct discrete method and its application in two dimensional multi-group neutron diffusion equation. *Annals of Nuclear Energy, 44*, 1–7.

Bochkarev, A. S., Alekseev, P. N., Korsun, A. S., & Kharitonov, V. S. (2016). Modeling of natural circulation for the inherent safety analysis of sodium cooled fast reactors. *Nuclear Energy and Technology, 2*(4), 294-298.

Butt, H. N., Ilyas, M., Ahmad, M., & Aydogan, F. (2016). Assessment of passive safety system of a Small Modular Reactor (SMR). *Annals of Nuclear Energy, 98*, 191-199.

Ceolin, C., Schramm, M., Vilhena, M. T., & Bodmann, B. E. (2015). On the Neutron multi-group kinetic diffusion equation in a heterogeneous slab: An exact solution on a finite set of discrete points. *Annals of Nuclear Energy, 76*, 271-282.

Chang, S. H., Kim, S. H., & Choi, J. Y. (2013). Design of integrated passive safety system (IPSS) for ultimate passive safety of nuclear power plants. *Nuclear Engineering and Design, 260*, 104–120.

Dai, S., Jin, C., Wang, J., & Chen, Y. (2013). Passive cooldown performance of integral pressurized water reactor. *Energy and Power Engineering, 5*(04), 505-509.

Gaheen, M. A., Elaraby, S., Aly, M. N., & Nagy, M. S. (2007). Simulation and analysis of IAEA benchmark transients. *Progress in Nuclear Energy, 49*(3), 217-229.

Giannetti, F., Di Maio, D. V., Naviglio, A., & Caruso, G. (2016). Thermal-hydraulic analysis of an innovative decay heat removal system for lead-cooled fast reactors. *Nuclear Engineering and Design, 305*, 168-178.

Guo, X., Sun, Z., Wang, J., & Yu, S. (2016). Steady-state performances and scaling analyses for an open flashing-driven natural circulation system. *Progress in Nuclear Energy, 87*, 1-14.

Hossain, K., Buck, M., Said, N. B., Bernnat, W., & Lohnert, G. (2008). Development of a fast 3D thermal-hydraulic tool for design and safety studies for HTRS. *Nuclear Engineering and Design, 238*(11), 2976-2984.

Kazeminejad, H. (2008). Thermal-hydraulic modeling of flow inversion in a research reactor. *Annals of Nuclear Energy, 35*(10), 1813-1819.

Kim, Y. H., & Yang, C. K. (2005). Development of safety analysis methodology for reactivity insertion accidents using modified RETRAN code. *Journal of Nuclear Science and Technology, 42*(11), 1001–1009.

Lee, K., Lee, K. H., Lee, J. I., Jeong, Y. H., & Lee, P. S. (2013). A new design concept for offshore nuclear power plants with enhanced safety features. *Nuclear Engineering and Design, 254*, 129-141.

Lim, J., Choi, S. W., Yang, J., Lee, D. Y., Rassame, S., Hibiki, T., & Ishii, M. (2014). Assessment of passive safety system performance under main steam line break accident. *Annals of Nuclear Energy, 64*, 287-294.

Liu, L., Zhang, D., Lu, Q., Wang, K., & Qiu, S. (2016). Preliminary neutronic and thermal-hydraulic analysis of a 2 MW thorium-based molten salt reactor with solid fuel. *Progress in Nuclear Energy, 86,* 1-10.

Mesquita, A. Z. & do Prado Souza, R. M. G. (2014). Thermal-hydraulic and neutronic experimental research in the TRIGA reactor of Brazil. *Progress in Nuclear Energy, 76*, 183–190.

Nikiforova, A., Hejzlar, P. & Todreas, N. E. (2009). Lead-cooled flexible conversion ratio fast reactor. *Nuclear Engineering and Design, 239*(12), 2596–2611.

Quintero-Leyva, B. (2012). Numerical solution of the multi-group integro-differential equation of the neutron diffusion kinetics in 2D-Cartesian geometry. *Annals of Nuclear Energy, 48*, 130–133.

Saha, P., Aksan, N., Andersen, J., Yan, J., Simoneau, J. P., Leung, L., ... & Kamide, H. (2013). Issues and future direction of thermal-hydraulics research and development in nuclear power reactors. *Nuclear Engineering and Design, 264*, 3-23.

Sun, P., Wang, Z., Zhang, J., & Su, G. (2016). Thermal–hydraulic design and analysis code development for steam generator of CFR600. *Annals of Nuclear Energy, 90*, 256-263.

Su'ud, Z. (2007). Advanced SPINNORs concept and he prospect of their deployment in remote area. In K. Ns, N. Aida, A. Purwaningsih, S. Surwadi, & M. Zen (Eds.), *Proceedings of the International Conference on Advances in Nuclear Science and Engineering in Conjunction with LKSTN* (pp. 199-207). Bandung, Indonesia: Institut Teknologi Bandung dan National Nuclear Energy Agency of Indonesia.

Su'ud, Z. (2008). Safety performance comparation of MOX, nitride and metallic fuel based 25-100 MWe Pb-Bi cooled long life fast reactors without on-site refuelling. *Progress in Nuclear Energy, 50*(2–6), 157–162.

Su'ud, Z., & Sekimoto, H. (1995). Design and safety aspect of lead and lead-bismuth cooled long-life small safe fast reactors for various fore configurations. *Journal of Nuclear Science and Technology, 32*(9), 834-845.

Tyagi, J., Kumar, M., Lele, H. G., & Munshi, P. (2013). Thermal hydraulic analysis of the AHWR—The Indian thorium fuelled innovative nuclear reactor. *Nuclear Engineering and Design, 262*, 21-28.

Vyskocil, L., & Macek, J. (2014). Coupling CFD code with system code and neutron kinetic code. *Nuclear Engineering and Design, 279*, 210-218.

Wang, M., Tian, W., Qiu, S., Su, G., & Zhang, Y. (2013). An evaluation of designed passive Core Makeup Tank (CMT) for China pressurized reactor (CPR1000). *Annals of Nuclear Energy, 56*, 81-86.

Wang, X., Jin, M., Wu, G., Song, Y., Li, Y., & Bai, Y. (2016). Natural circulation characteristics of lead-based reactor under long-term decay heat removal. *Progress in Nuclear Energy, 90*, 11-18.

Wei, C., Sui, Z., Ma, Y., Jing, X., & Shan, W. (2014). Application of improved quasi-static method to high temperature gas cooled reactor simulator. *Nuclear Engineering and Design, 271*, 337-340.

Xia, L., Jiang, J., Javidnia, H., & Luxat, J. C. (2012). Performance evaluation of a 3-D kinetic model for CANDU reactors in a closed-loop environment. *Nuclear Engineering and Design, 243*, 76-86.

Yan, M., & Sekimoto, H. (2008). Safety analysis of small long life CANDLE fast reactor. *Annals of Nuclear Energy, 35*(5), 813–828.

Yulianti, Y., Su'ud, Z., Waris, A., & Khotimah, S. N. (2009). Space-time diffusion equations for three-dimensional control rod withdrawal reactor accident. In *Proceedings of the International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering* (pp. 1–5). Bandung, Indonesia: Institute of Electrical and Electronics Engineers Inc.

Zhao, P., Li, S., Chen, Z., Zheng, J., & Chen, H. (2015). Natural circulation characteristics analysis of a small modular natural circulation lead–bismuth eutectic cooled fast reactor. *Progress in Nuclear Energy, 83*, 220-228.

SCIENCE & TECHNOLOGY

Using Jolly Balance Spring Method to Determine Pure Water Surface Tension Coefficient

Duden Saepuzaman*, Muhamad Gina Nugraha, Regiana Dewi, Fitri Kafiyani and Fanny Herliyana Dewi

Departemen Pendidikan Fisika, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia

ABSTRACT

The surface of a liquid at rest behaves in an interesting way, almost as if it was a stretched membrane under tension. A steel needle can be made to float on the surface of water even though it is denser than water. The surface of a liquid acts like it is under tension, and this tension, acting along the surface, is a product of attractive forces between the molecules. This effect is called surface tension. There are various methods to determine the surface tension coefficient of a liquid. This research focused on determination of pure water surface coefficient using Jolly's Spring Balance. The parameter measured is the height of the liquid bubbles at critical condition that in the right condition will break. Repeated measurement methods that are statistically processed are used to measure uncertainty. The magnitude of pure water surface value using Jolly Balance Spring Apparture's obtained γ (water) = $(3,620 \pm 0.0028) \times 10^{-2}$ N / m at temperature (27 ± 0.25)°C and pressure (68.8 ± 0.005) cmHg. Liquid used in this research is pure water at temperature 27°C.

Keywords: Altitude, jolly's balance spring, surface tension

INTRODUCTION

Surface tension is the surface property of a liquid that behaves like a thin layer of skin that is supple due to stress (Chappuis, 1982). Mathematically, the surface tension is the force of each unit of length parallel to the surface to compensate for the inward pull force of a fluid (Brakke, 1992). This happens because on the surface, the adhesion force (between fluid and air) is smaller than the

cohesive force between the fluid molecules causing the inward force on the fluid surface (Wu, 1973).

A liquid consists of molecules. The forces acting on each molecule have the same magnitude and work in all directions, so that the molecules between them are pulled together (Eijkel & Van Den Berg, 2005). The molecular force resistivity in the liquid is zero. While on the surface, the resultant force points downward and tends to depress the surface layer but only to the extent that the downward force is offset by the upside-down force. Thus, the surface tension is defined as the magnitude of force experienced by the liquid surface per unit length (or equivalent to the energy per unit area) (McConney, Singamaneni, & Tsukruk, 2010). It can be mathematically expressed as (Reyssat, 2014):

$$\gamma = \frac{F_\gamma}{l} \tag{1}$$

with γ is the surface tension (N m^{-1}), $F\gamma$ is the force acting on the fluid surface (N), and l is the length of the object on the fluid surface (m). The determination of the magnitude of the surface tension is characterised by the coefficient of surface tension having unit force per unit length, N m^{-1}, or energy per unit area, J m^{-2}.

There are several methods to measure the surface tension coefficient, for example, pull-off method. As for the pull-off method, a du Noüy ring or Wilhelmy plates (Gaonkar & Neuman, 1984) will be hung under a dynamometer and immersed into the test liquid. Then the dynamometer is used to measure the force used to pull the ring or plate out of the liquid. By analysing the force on the ring or the plate, the surface tension coefficient of the test liquid can be calculated. For the du Noüy ring method, the wetting properties of the surface or interface have little influence. The Wilhelmy plate method is a universal method that is particularly suited to check surface tension over long intervals. Many studies have examined the pull-off method on using a silicon piezoresistive force sensor to improve measurement accuracy, reduce operational complexity and simplify the steps (Tang, Liu, Qian, Sun, & Zhang, 2016).

This study used the pull-off method to measure the surface tension coefficient using Jolly Spring Balance method with circular ring used as force gasifier on fluid surface . This method is simple because we observe the critical condition before the water surface tension break-off.

METHOD

The surface tension of a liquid using Jolly balance uses a thin ring as an indicator of the breaking of the surface tension of the liquid. Tools used are shown in Figure 1.
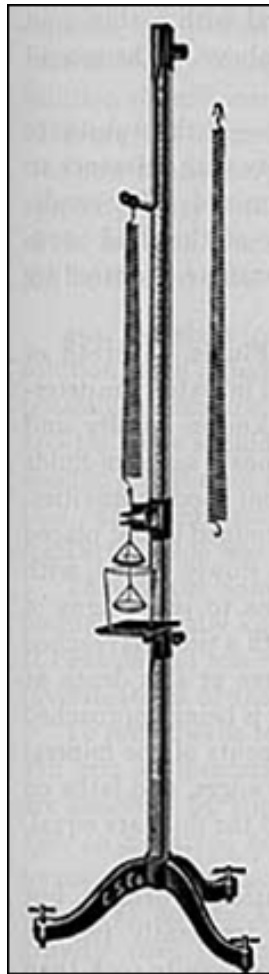
Figure 1. Jolly balance apparatus

At the end of the Jolly's balance apparatus, there is a hanger that serves to hang the spring. There is a stage where the height can be adjusted, the stage is used as a place to store the cup, which is used as a liquid site to determine the value of surface tension coefficient. At the bottom of the stage there is a micrometre screws which are used to lower the cup. Care had to be taken the jolly spring balance stands upright by adjusting the screw on the bottom of the jolly spring balance tripod. This situation can be ensured by using a water pass placed on the stage. In determining the surface tension, a wire ring is used. This wire rings are attached to the springs, which slowly rotate the screw micrometry so as to raise the cup closer to the wire ring until the bottom of the ring reaches the liquid surface directly. The height indicated by the micrometre is the initial height (h_0). After that, the saucer is lowered slowly until the surface tension / liquid membrane breaks precisely. The height at that condition is the final height (h_1). The difference between h_1 and h_0 (h = h_1- h_0) is a parameter calculated repeatedly in the experiment of determining the surface tension by using this Jolly spring balance.

The force acting on the ring is shown in Figure 2.
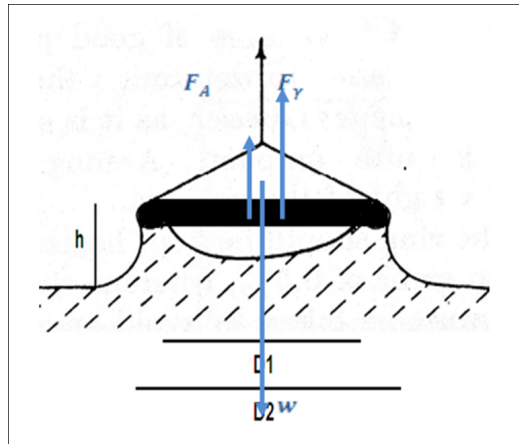
(a)

(b)

Figure 2. The critical condition at the time of the surface membrane / surface tension will break

To keep the object moving on the surface of the fluid (the object is in equilibrium), a total force downwards is required, where the total force is $F_{res} = w - F_\gamma + F_A = 0$ or $w = F_\gamma + F_A$. With the weight of the ring, $w = m.g$; liquid surface tension; $F_\gamma = \gamma \cdot l$; and the magnitude of the buoyant force $F_A = \rho f \cdot g \cdot V$ (Wu, 1973)

Since the object used is a ring, the following satisfies the condition (1) of the ring on the obtained surface tension force:

$$F_\gamma = \gamma \frac{(l_2 + l_1)}{2}$$

$$F_\gamma = \gamma \frac{(2\pi d_2 + 2\pi d_1)}{2}$$

$$F_\gamma = \gamma \pi (d_2 + d_1) \qquad (2)$$

The volume (in the floating force equation) is the difference in the volume of the outer diameter tube (d_2) with the inner diameter (d_1). Therefore, we have the following:

$$F_A = \rho_f . g . (V_2 - V_1)$$

$$F_A = \rho_f . g . \left(\frac{1}{4} \pi d_2{}^2 h - \frac{1}{4} \pi d_1{}^2 h \right)$$

$$F_A = \rho_f . g . \frac{1}{4} \pi h \left(d_2{}^2 - d_1{}^2 \right) \qquad (3)$$

Substituting equations (2) and (3) to w = $F_\gamma + F_A$, the following is obtained:

$$m . g = \gamma \pi (d_2 + d_1) + \rho_f . g . \frac{1}{4} \pi h \left(d_2{}^2 - d_1{}^2 \right)$$

$$\gamma \pi (d_2 + d_1) = m . g - \rho_f . g . \frac{1}{4} \pi h \left(d_2{}^2 - d_1{}^2 \right)$$

$$\gamma = \frac{m . g}{\pi (d_2 + d_1)} - \frac{1}{4} \frac{\rho_f . g . \pi h \left(d_2{}^2 - d_1{}^2 \right)}{\pi (d_2 + d_1)}$$

$$\gamma = \frac{m . g}{\pi (d_2 + d_1)} - \frac{1}{4} \frac{\rho_f . g . \pi h (d_2 - d_1)(d_2 + d_1)}{\pi (d_2 + d_1)}$$

$$\boldsymbol{\gamma = \frac{m.g}{\pi (d_2 + d_1)} - \frac{(d_2 - d_1)}{4} \rho_f . g . h} \qquad (4)$$

with γ surface tension (N /m), m mass of object (kg), acceleration of earth gravity g (m/s^2), d2 outer diameter of ring (m), d_1 inner diameter of ring (m), ρ_f density of fluid (kg / m^3), and h height of fluid when attached to the object (m). Temperature is one of the factors that affect the value of surface tension of the fluid. Generally, when there is an increase in temperature, the surface tension value decreases. This is because when the temperature increases, fluid molecules move faster so that the interaction effect between fluid molecules is reduced. As a result, the surface tension value also decreases. Value of surface tension is influenced by several factors, among others, temperature, pressure, density and concentration of solute (Harrison, Johnston, & Sanchez, 1996).

Measurement of necessary data is done repeatedly to minimise the magnitude of uncertainty. The determination of the surface tension value, $\gamma = (\bar{\gamma} \pm \Delta\gamma)$ N / m, with the uncertainty value

used based on the uncertainty value on the single variable over and over. With $\bar{\gamma}$ is the average surface tension value and $\Delta\gamma$ the number of uncertainties can be expressed as:

$$\Delta\gamma = \left(\frac{(d_2 - d_1)}{4}\rho_f \cdot g\right)(\Delta h)\sqrt{\frac{1}{n-1}}. \tag{5}$$

RESULTS AND DISCUSSION

Data was obtained from the experimental results to determine the surface tension. Water density (ρ_f) value of 1000 kg/m^3 measured using Aerometer instrument, inner (d_1) and outer (d_2) diameter of metal rings, 1.84 cm and 1.98 cm respectively were measured using a Vernier Caliper, and a gravity acceleration value of the research site (g) of 9.8 m/s^2 obtained from literature. The height of water data at the critical condition (Δh) was obtained using jolly balance with repeated measurement 5 times as shown in Table 1.

Table 1
The results of height water surface in critical condition

No	h_1 (± 0.05)x10^{-3} m	h_0 (± 0.05)x10^{-3} m	Δh (± 0.05)x10^{-3} m
1	20.3	14.7	5.6
2	20.4	14.7	5.7
3	20.5	14.9	5.6
4	20.4	14.8	5.6
5	20.3	14.6	5.7
Mean			5.64
Standard deviation			0.0548

Based on Table 1, it could be seen that water level at critical condition has a consistent value with the average of 5.64 cm (5.64 x 10^{-3} m) and standard deviation 0.0548. This value is then processed with other data to obtain the surface tension value using the above formula.

Based on the data processing result, water surface tension value of water γ water is (3.620 $\pm$ 0.0028)$\times$10^{-2} N/m at temperature (27 $\pm$ 0.25)°C and pressure (68.8 $\pm$ 0.005) cmHg. The result is inconsistent with literature findings, i.e. at water temperature 0°C, water surface tension value γ is 7.6$\times$10^{-2} N/m; at water temperature 20°C, water surface tension value γ is 7.2$\times$10^{-2} N/m; and at water temperature 100°C, surface tension value $F\gamma$ is 5.9$\times$10^{-2} N/ m. (Vazquez, Alvarez, & Navaza, 1995)

The inconsistency in findings is due to several factors, namely the presence of convex nature of meniscus in water resulting from the interaction between water molecules and different molecules (in this case with molecules in the ring) (Nandiyanto, Hagura, Iskandar, & Okuyama, 2010). The meniscus symptoms result in an inaccuracy when determining the initial height of the ring just as it touches the surface of the liquid. In addition, the precise bubble conditions where it breaks should be achieved in equilibrium where the resultant force acting on the system is equals to zero, and the resultant force measured in the formula comes only

from the gravity of the ring, the buoyant force of the liquid and the surface tension force (Wu, 1973). However, when experimenting there are external disturbances, such as the presence of vibrations that result in oscillating springs. Thus, the precise of surface tension conditions will break apart by other unmeasured forces in the formulation.

Although the value of surface tension obtained is different from what was reported in the literature, it is not due to the pure water used, but it could also be due to the procedure and the set of tools that still need to be reviewed and further developed.

CONCLUSION

Determination of pure water surface value using Jolly Balance Spring obtained γ (water) = $(3,620 \pm 0.0028) \times 10^{-2}$ N / m at temperature $(27 \pm 0.25)°$C and pressure (68.8 ± 0.005) cmHg. Further studies and research are needed to obtain the accuracy of surface tension values.

REFERENCES

Brakke, K. A. (1992). The surface evolver. *Experimental Mathematics, 1*(2), 141-165.

Chappuis, J. (1982). Contact angles. In G. F. Hewitt, J. M. Delhaye & N. Zuber (Eds.), *Multiphase Science and Technology* (pp. 387-505). Washington: Hemisphere Publishing Corp.

Eijkel, J. C., & Van Den Berg, A. (2005). Nanofluidics: what is it and what can we expect from it?. *Microfluidics and Nanofluidics, 1*(3), 249-267.

Gaonkar, A. G., & Neuman, R. D. (1984). The effect of wettability of wilhelmy plate and du Noüy ring on interfacial tension measurements in solvent extraction systems. *Journal of Colloid and Interface Science, 98*(1), 112-119.

Harrison, K. L., Johnston, K. P., & Sanchez, I. C. (1996). Effect of surfactants on the interfacial tension between supercritical carbon dioxide and polyethylene glycol. *Langmuir, 12*(11), 2637-2644.

McConney, M. E., Singamaneni, S., & Tsukruk, V. V. (2010). Probing soft matter with the atomic force microscopies: imaging and force spectroscopy. *Polymer Reviews, 50*(3), 235-286.

Nandiyanto, A. B. D., Hagura, N., Iskandar, F., & Okuyama, K. (2010). Design of a highly ordered and uniform porous structure with multisized pores in film and particle forms using a template-driven self-assembly technique. *Acta Materialia, 58*(1), 282-289.

Reyssat, E. (2014). Drops and bubbles in wedges. *Journal of Fluid Mechanics, 748*, 641-662.

Tang, L., Liu, G. N., Qian, J., Sun, Q., & Zhang, C. L. (2016). Discussion on the measurement of the surface tension coefficient by the pull-off method. *European Journal of Physics, 37*(2), 025801.

Vazquez, G., Alvarez, E., & Navaza, J. M. (1995). Surface tension of alcohol water+ water from 20 to 50. degree. C. *Journal of Chemical and Engineering Data, 40*(3), 611-614.

Wu, S. (1973). Polar and nonpolar interactions in adhesion. *The Journal of Adhesion, 5*(1), 39-55.

Analysis of Attacks on Mail Disposition Systems Secured by Digital Signatures Equipped with AES and RSA Algorithms

Herbert Siregar, Enjun Junaeti* and Try Hayatno

Departemen Pendidikan Ilmu Komputer, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia

ABSTRACT

Implementation of Information and Communication Technology (ICT) to assist document management in the mail distribution will be useful in improving document management performance effectively and efficiently. Securing information in mail disposition using electronic media should be made in such an order to prevent unwanted things. Cryprography or digital signature are some of the techniques that ensures security. The purpose of this study is to see the effect of digital signatures to maintain data security using Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES) algorithms to the disposition system of letter based on cryptographic purposes, namely Secrecy, Data Integrity, Authentication and Non-repudiation. Several tests of various attacks to the system were conducted. The study concluded the additions of AES algorithm can maintain secrecy and integrity of the letters, while RSA algorithm can maintain authentication and non-repudation of the letters.

Keywords: AES and RSA algorithms, cryptography attacks, digital signature, mail disposition

INTRODUCTION

From time to time the volume of documents - one of which is the mail archives - in an institution or a company will increase, so there is need to have a good document management, so the information contained in the archive can support the performance of the organisation. Implementation of Information and Communication Technology (ICT) to assist document management will be useful in improving document management performance effectively and efficiently, even by utilising internet technology will make it easier for users of information to access information which is loaded on document simultaneously without any obstacle either time or space (Buhalis & Law, 2008). However, it does not have any obstacles,

because information from the documents cannot be guaranteed if not properly maintained. In order to prevent unwanted things such as changes or leaks of the letter contents, the mail security should be safeguarded. This paper discusses the performance of one of the document management systems which is assisted by ICT, a mail disposition system, when facing various cryptographic attacks.

Juju and Cowhand (2015) stated that digital signature can maintain the safety of electronic mails, while Nemavarka and Chakrawarti (2015) used cryptography to do so. Digital signature is a mark in a digital document that offered protection against an attack when a sender or a receiver is trying to have modifications of their messages either when they send them or receive them (Bernstein, Buchmann, & Dahmen, 2009). Schneier (1996) said that cryptography was a science and art to maintain the confidentiality of the message by encoding it into a form that no longer understands its meaning. But cryptography is more than to ensure just privacy; it also ensures data integrity, autentication, and nonrepudiation (Katz, Menezes, Van Oorschot, & Vanstone, 1996).

According to Kumar and Hanok (2015), digital signature was a good solution to maintain document security because it could withstand active attacks, such as forgery attack and choosing cipher text attack (Kumar & Hanok, 2015). They provided timestamped signature scheme which could be verified universally using signer's public parameters in message recovery which is implemented in E-Cash System.

The purpose of this study is to discover the effect of digital signatures to maintain data security using Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES) algorithms to evaluate performance of the mail disposition system based on cryptographic purposes, namely Secrecy (Devi, 2013), Data Integrity (Katagishi, Asami, Ebihara, Sugiyama, & Toraichi, 1999), Authentication (Dhagat & Joshi, 2016) and Non-repudiation (Al-Hammadi & Shahsavari, 1999). The RSA algorithm is an asymmetric cryptographic algorithm that has a key length in bits that can be set. Thus, the longer the bit gets, the more difficult it is to solve because of the difficulty of factoring the very large two numbers and it takes a long time for the decryption process. In the meantime, AES algorithm is a symmetric algorithm that uses a block cipher (Daemen & Rijmen, 2013). During the encryption and decryption process, AES-128 algorithm does 10 cycles of transformation functions, i.e. Add Round Key, Sub Byte, Shift Rows and Mix Column (Schneier, 2007). The addition of asymmetric and symmetric algorithms, i.e. RSA and AES-128, on the digital signature had been carried out in this study for security data maintenance. Several test of various attacks to the system is conducted to evaluate the performance of the system. In conclusion, based on the research outcome, using AES algorithm can maintain secrecy and integrity of the mails, while RSA algorithm can maintain authentication and non-repudation of the mails again those attacks.

RESEARCH METHOD

The application used in this study was a mail disposition system equipped with digital signature using AES and RSA algorithms. This study used files or images with jpg format because they were easily opened using standard platforms and application in the computer operating system. The software used is described in the Figure 1.

Based on Figure 1, an admin can write a mail and view a mail. The AES and RSA algorithms will be including in email communications or viewing an incoming mail. Securing key in the process of digital signature is to sign that the messages sent with ad legality is done by RSA algorithm. The RSA algorithm is used to reach the purposes of cryptographic i.e. authentication and non-repudiation (Nandhakumar, Binu, & Paul, 2013). The AES algorithm is used to ensure security especially within the message encoding process. The AES algorithm is also used in this study to meet the objectives of cryptography i.e. secrecy and data integrity (Mewada, Sharma, & Gautam, 2016), so the content of the messages are protected from actions like tapping data.
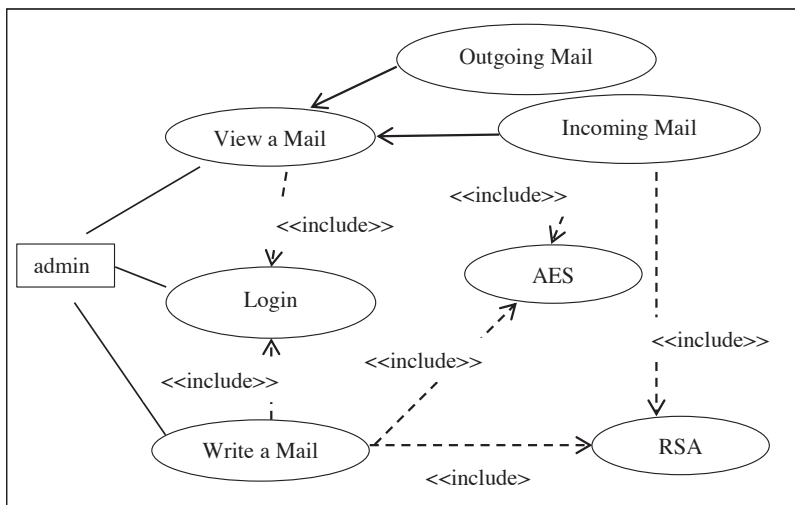
Figure 1. Use case diagram of disposition system of letter

Digital Signature Scheme

A Hash function – modified digital signature scheme (Bao & Deng, 1998) is carried out in this study. Modification is made by adding the encryption process of the mail therein. The AES algorithm serves to encript the message contents that would be delivered to the receiver while RSA algorithm will conceal the signature which is delivered along with the mail. The signature in the system is a key used to encrypt message by AES algorithm. The way RSA algorithm works is slower than that of symmetric cryptography such as DES or AES (Saxena & Kapoor, 2014). Therefore, a symmetric key algorithm, namely AES algorithm would be used to encrypt the message, while an asymmetric algorithm like RSA algorithm, is used to encrypt the key. Figure 2 shows the modification of digital signature scheme in this study.

Figure 2 shows that messages encrypted by AES algorithm would be delivered to the receivers and stored in the database of the system. The AES key used in the system is a combination of an eight-digit number of the sender identity and an eight-digit number of the date of letter sent. Then RSA algorithm would encrypt the key used in the encryption algorithm AES to generate cipher key as well as signature that would be stored in the database.

The RSA algorithms consist of namely public key and private key (Aswathy & Resmi, 2014). In this research, two keys contain prime numbers ranging from 2 to 2000 are randomly selected. The recipients would then decrypt the signature so that the key used to decrypt the messages can be generated.
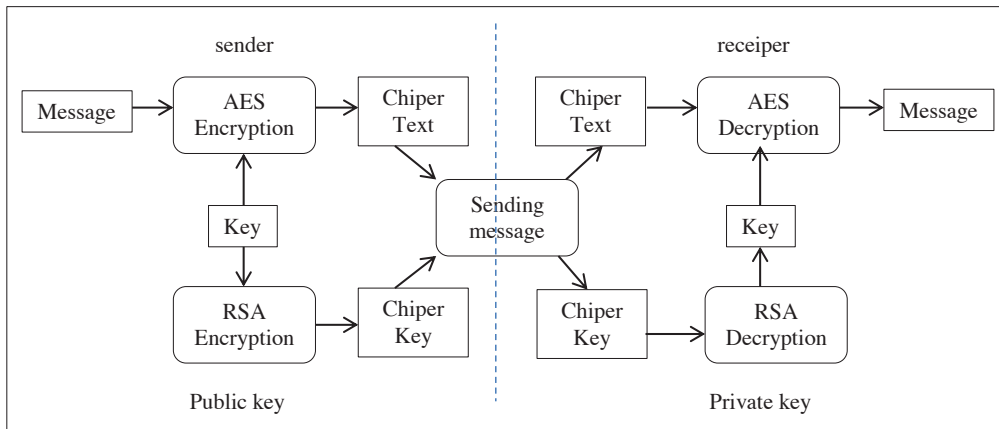
Figure 2. Modification of digital signature scheme

Experiment Testing

Several tests of various attacks had been conducted to achieve the aims of this study. Tests on digital signature schemes had been designed in accordance with the modification of the attacks (Batten, 2013), i.e. addition, reduction, and changing of single character on signature, using different key, and turning aroun time process.

The messege used in this test is shown in the figure 3(a). The keys and signature used in the experiment testing are:

- AES key (plain text) : 1957122620160621
- Public key for RSA   : (7, 1728239)
- Private key for RSA : (739543, 1728239)

The ASCII Code of AES key will be encrypted using RSA algorithm to obtain signature of the message. ASCII Code of plain text (AES key) is:

49 7 53 55 49 50 50 54 50 48 49 54 48 54 49 53

So, the signature (RSA chiper text) used in this experiment testing is:

117649 185193 148877 166375 117649 125000
125000 157464 125000 110592 117649 157464
110592 157464 125000 117649

RESULTS AND DISCUSSION

Figure 3 shows an example of a file (mail) which would be encrypted and encryption as well as decryption results if the receiver uses the correct key to decrypt the message in process of the digital signature of the system.
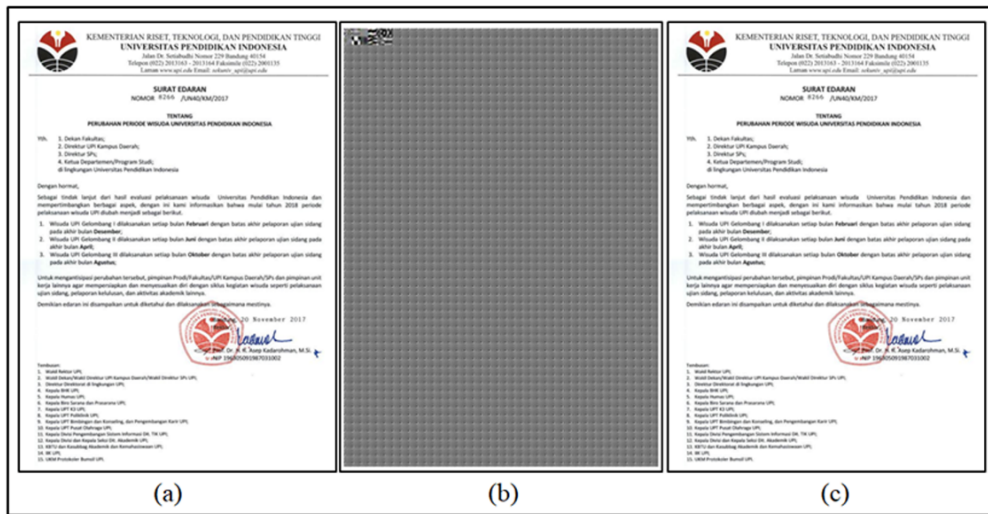
Figure 3. Sample of letter encryption using the AES algorithm

Figure 3(a) shows an example of a mail (image) with jpg format to be sent during the encryption process that used AES algorithm, only the image data will be encoded so the results are also images with jpg format, as shown in Figure 3(b). After the encryption process wis completed, the image is stored on the database system in the form of image cipher text. Figure 3(c) shows the result of decryption of message.

In order to evalutae the performance of the system fofr cryptography purposes, various tests of attack are executed in the system. The addition, reduction, and changing of one character of the signature are run in order to assess the influence of a change in the value of the AES key toward the result of the image decryption. In addition, there is a test to evaluate system resistant toward changing key and turnaround time testing.

Character Changing in the Signature

Various tests of attack character changing in the signature and decrypted results by RSA algorithms is shown in Table 1.

Table 1

Various tests of attack and the results of RSA algorithms decryption

Attack	Signature changes by the Attack	Result of RSA algorithm decryption
Addition of Single Character on Signature	117649 185193 148877 166375 117649 1215000 125000 157464 125000 110592 117649 157464 110592 157464 125000 117649	19571�2620160621
Reduction of One Character	117649 85193 148877 166375 117649 125000 125000 157464 125000 110592 117649 157464 110592 157464 125000 117649	1�57122620160621
Changing of One Character	117649 185193 148877 166375 117649 125000 125000 158464 1215000 110592 117649 157464 110592 175616 110592 148877	1957(22620160621

Based on Table 1, addition of single character on signature attack is done by adding character "1" at the 33rd place of the character sequence on the signature. Then character "1" at the 7th place of the character sequence on the signature is omitted to perform reduction of one character attack. The last at the 45th place of the character sequence on the signature, character "7" replaced by "8" to evaluated changing of one character in the signature towards the result of AES algorithm decryption. Figure 4(a), 4(b), and 4(c) show the results of AES algorithm to decrypt the message using a key which is an outcome using by RSA algorithm decryption process.

(a)　　　　(b)　　　　(c)　　　　(d)
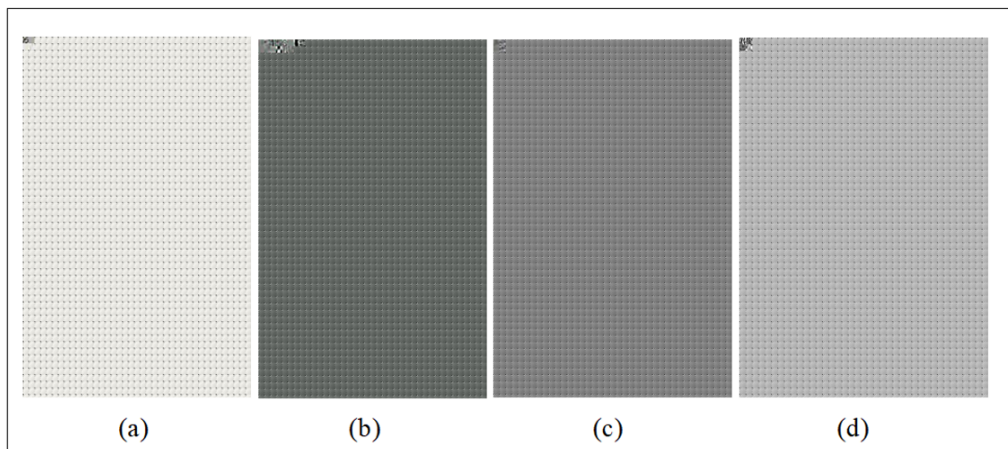
Figure 4. The result of decryption from the various tests

Secrecy of data will secure if no unauthorised parties know the meaning of it (Devi, 2013). Based on the result of AES encryption process, the message has been encrypted without any meaning, whitout AES decryption process, the purpose of the message could not be known, so that confidentiality of the data is maintained. The result of decryption of message in the Figure 3(c) shows that there is no change in the content and purpose massage. But based on Figure 4(a), Figure 4(b), and Figure 4(c), process decryption generates random message, it shows that the accuracy of the message could be guaranteed because the digital signature is very sensitive to a change in data toward the result of decryption of image, so data integrity of the massage can be maintained. It means the AES encryption and decryption process in this digital signature meet the cryptography purposes, i.e. secrecy (Sibert, 2006) and data integrity (Katagish et al., 1999).

Using Different Key

In addition to attacks in the Table, the system is tested to try to deal with "using different key" attacks, i.e. the key is used by receiver to decrypt the signature in the RSA algorithm decryption process is different with right key. This test is administered as an attempt in a case where the receiver uses a different private key which is not the spouse of the public key used to create the signature. Suppose that the key used is:

n = 1685287
public key = 3
private key = 1241787

The results of decryption with the RSA algorithm is:

□□9□□□#□□□#□□9

If this key is used to decrypt the message with the AES algorithm, the decryption result is shown by Figure 4(d). It can be seen that the result does not match the original image in Figure 3(a). This shows that the sensitivity of the digital signature scheme to the keys is good.

Using the right private key to decrypt signature will generate the correct signature, which contains the identity of the sender. This results in the authentication of the sender, and the sender cannot deny that the message is sent by them. The right private key can only be used by the right receiver, this results inauthentication of the receiver. It means that the RSA encryption and decryption process in this digital signature meet the cryptography purposes, i.e. authentication and non-repudiation (Chen, Horng, & Liu, 2013).

Turn Around Time

Turn Around Time test is done to excute the time needed for encryption of plaintext into cipher text and dectyption of cipher text back to plaintext. The image used in the test is 5 .jpg with a size of 1.96 KB to 58 KB. The results of the test is shown in Table 2.

Table 2
Table testing turnaround time

No	File Name	Size (KB)		Time (second)		
		Before	After	Encryption	Decryption	Sum
1	surat_edaran.png	513,00	513,00	36,53	86,51	131,58
2	sample1.jpg	185,00	185,00	13,17	31,20	47,45
3	sample2.jpg	163,00	163,00	11,61	27,49	41,81
4	sample3.jpg	856,00	856,00	60,96	144,35	219,56
5	surat_edaran.jpg	114,00	114,00	8,12	19,22	29,24
6	sample3.jpg	31,00	31,00	3,48	14,06	17,55
7	sample3.jpg	36,00	36,00	4,15	0,72	0,89
8	sample3.jpg	58,00	58,00	0,30	1,14	33.97

Table 2 shows the file size remains the same before encryption and after decryption using the AES algorithm. Hence, AES algorithm does not change the file size \used, so the digital signature maintained the integrity of the message. Therefore, it can be said that the larger the size of the file, the longer the time needed for encryption and decryption processes. It is shown that the time required for process of encryption or decryption is different, depending on the length of plaintext.

Digital signature scheme modification by adding encryption key along with the message and the encryption process when sending messages and adding the decryption process and the decryption key when receiving message is done correctly. The time required for the execution of the system increased because the system would make the process of AES and RSA encryption algorithm when sending messages and make decryption algorithm AES and RSA algorithms when receiving messages.

Implementation of digital signatures using algorithms AES and RSA algorithms on mail disposition system meet the four cryptographic purposes, namely secrecy, data integrity, authentication and non-repudiation. The validity of the sender and receiver are ensured by digital signatures, so the message authentication can be assured (Fu & Wei, 2011). The precise signature could ensure that the sender can not refute that the message was sent by them; it means the signature meets non-repudiation purpose (Belhadj & Akrouf, 2015). The accuracy of the message can be guaranteed by AES algorithm, so it could preserve the secrecy and integrity of data messages in the message delivery process.

CONCLUSION

Based on the research, it can be concluded that the additions of AES and RSA algorithms on the digital signature made the mail disposition system resistant to cryptographic attacks. The protection given to the system meets four objectives of cryptography, namely Secrecy, Data Integrity, Authentication and Non-repudiation.

ACKNOWLEDGEMENT

REFERENCES

Al-Hammadi, B., & Shahsavari, M. (1999). Certified exchange of electronic mail (CEEM). In *Proceedings of the IEEE Southeastcon '99* (pp. 40-43). Lexington, Kentucky: Institute of Electrical and Electronics Engineers Inc.

Aswathy, B. G., & Resmi, R. (2014). Modified RSA public key algorithm. In *Proceedings of the First International Conference on Computational Systems and Communications* (pp. 252-255). Trivandrum, Kerala, India: Institute of Electrical and Electronics Engineers Inc.

Bao, F., & Deng, R. H. (1998). A signcryption scheme with signature directly verifiable by public key. In H. Imai, & Y. Zheng (Eds.), *Proceedings of the International Workshop on Public Key Cryptography* (pp. 55-59). Yokohama, Japan: Springer International Publishing.

Batten, L. M. (2013). *Public key cryptography: applications and attacks.* Hoboken, New Jersey: John Wiley & Sons.

Belhadj, F., & Akrouf, S. (2015). Secure fingerprint-based authentication and non-repudiation services for mobile learning systems. In *Proceedings of the International Conference on Interactive Mobile Communication Technologies and Learning* (pp. 200-204). Thessaloniki, Greece: Institute of Electrical and Electronics Engineers Inc.

Bernstein, D. J., Buchmann, J., & Dahmen, E. (2009). *Post-quantum cryptography.* Boston, MA: Springer Science & Business Media.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management, 29*(4), 609-623.

Chen, Y. C., Horng, G., & Liu, C. L. (2013). Strong non-repudiation based on certificateless short signatures. *IET Information Security, 7*(3), 253-263.

Daemen, J., & Rijmen, V. (2013). *The design of Rijndael: AES-the advanced encryption standard.* Heidelberg, NY: Springer Science & Business Media.

Devi, T. R. (2013). Importance of cryptography in network security. In G. S. Tomar, M. Dixit, & F. Z. Wang (Eds.), *Proceedings of the Communication Systems and Network Technologies* (pp. 462-467). Gwalior, India: Institute of Electrical and Electronics Engineers Inc.

Dhagat, R., & Joshi, P. (2016). New approach of user authentication using digital signature. In *Proceedings of the Symposium on Colossal Data Analysis and Networking* (pp. 1-3). Indore, India: Institute of Electrical and Electronics Engineers Inc.

Fu, D., & Wei, Z. (2011). Research and implementation of a digital signature scheme based on middleware. In *Proceedings of the International Conference on Electrical and Control Engineering* (pp. 2468-2471). Yichang, China: Institute of Electrical and Electronics Engineers Inc.

Jaju, S. A., & Chowhan, S. S. (2015). A modified RSA algorithm to enhance security for digital signature. In *Proceedings of the International Conference and Workshop on Computing and Communication* (pp. 1-5). Vancouver, Canada: Institute of Electrical and Electronics Engineers Inc.

Katagishi, K., Asami, T., Ebihara, Y., Sugiyama, T., & Toraichi, K. (1999). A public key cryptography-based security enhanced mail gateway with the mailing list function. In *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing* (pp. 262-265). Victoria, Canada: Institute of Electrical and Electronics Engineers Inc.

Katz, J., Menezes, A. J., Van Oorschot, P. C., & Vanstone, S. A. (1996). *Handbook of applied cryptography.* Boca Raton, USA: CRC press.

Kumar, P. B. S. P., & Hanok, K. (2015). *Time stamped digital signature scheme with message recovery and its application in E-Cash system* (Unpublished doctoral thesis). National Institute Of Technology Rourkela, India.

Mewada, S., Sharma, P., & Gautam, S. S. (2016). Exploration of efficient symmetric AES algorithm. In *Proceedings of the Symposium on Colossal Data Analysis and Networking* (pp. 1-5). Indore, India: Institute of Electrical and Electronics Engineers Inc.

Nandhakumar, N. K., Binu, A., & Paul, V. (2013). Non repudiation for internet access by using browser based user authentication mechanism. In *Proceedings of the Third International Conference on Advances in Computing and Communications* (pp. 296-299). Kochi, Kerala, India: Institute of Electrical and Electronics Engineers Inc.

Nemavarkar, A., & Chakrawarti, R. K. (2015). A uniform approach for multilevel email security using image authentication, compression, OTP and cryptography. In *Proceedings of the International Conference on Computer, Communication and Control* (pp. 1-5). Indore, India: Institute of Electrical and Electronics Engineers Inc.

Saxena, S., & Kapoor, B. (2014). An efficient parallel algorithm for secured data communications using RSA public key cryptography method. In U. Batra, S. Sujata, & A. Arpita (Eds.), *Proceedings of the IEEE International Advance Computing Conference* (pp. 850-854). Gurgaon, India: Institute of Electrical and Electronics Engineers Inc.

Schneier, B. (1996). *Foundations: Applied cryptography* (2nd Ed.). USA: John Wiley & Sons.

Schneier, B. (2007). *Applied cryptography: Protocols, algorithms, and source code in C.* USA: John Wiley & Sons.

Sibert, W. O. (2006). *U.S. Patent No. 7,058,805.* Washington, DC: U.S. Patent and Trademark Office.

Technical Efficiency Chemical Industry in Indonesia: Stochastic Frontier Analysis (SFA) Approach

Amir Machmud[1]*, Asep Bayu Dani Nandiyanto[2] and Puspo Dewi Dirgantari[1]

[1]*Faculty of Business and Economic Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*
[2]*Departemen Kimia, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia*

ABSTRACT

This study aims to analyse the technical efficiency of Chemical Industry in Indonesia using Stochastic Frontier Analysis (SFA) approach. A survey explanatory method is adopted with panel type data at 4240 inputs and output level of companies during six years of observation (2010-2015). Determinant technical efficiency is estimated by using multiple regression. Both technical efficiency calculations and regression determinants are done using the TE effect approach in Frontier 4.1 software. The result of research indicates that market share variable significantly and negatively affects technical efficiency. Other variables, such as company age, business ownership, ratio concentration, and capacity utilisation affect the technical efficiency are not significant. This finding implies that to improve the efficiency of the chemical industry, market share needs to be increased through various activities such as promotion.

Keywords: Chemical industry, Stochastic frontier analysis, technical efficiency

INTRODUCTION

Synthetic materials have advantages over natural materials on the ease of obtaining properties of the desired material. Today some of humanity's needs can be met through the chemical industry, from textiles, transportation equipment (tires, carbon composites to vehicle bodies), electronic equipment (most electronics components), communication technology (fibre optic cable), drugs, building materials to household appliances. The development of the chemical industry brings about a huge change in life (Plechkova & Seddon, 2008; Tombs & Whyte, 2003).

The chemical industry is small in Indonesia compared with other industrial subsectors, such as food, beverages, tobacco, agriculture and manufacturing industries.

European countries, North America and East Asia (Japan, South Korea, and China). When viewed from the development of the chemical industry from year to year, it appears that the chemical industry in Indonesia is not developing properly, although technological developments in the chemical industry has advanced very rapidly in the last two decades (Nandiyanto, Maulana, Ragadhita, & Abdullah, 2018).

The biggest obstacles that must be faced by the industry in chemistry in Indonesia today are: 1) The absence of linkage of micro and small industries with medium and large industries, even the occurrence of competition for market competition and raw materials among small / micro industries with medium-large industries. This inhibits the development of both parties. 2) Weak structure of the upstream-downstream industry. In developed countries, chemical industries are intertwined with each other, so that output from one factory can be used as inputs at other plants. Thus, a long chain of chemical industries will be created which implies an increase in value added to the goods. If seen from the value of the import of chemicals in Indonesia, the chemical industry is still not able to realise the integration between its sub-sectors of the industry. This has resulted in the weakness of industry in Indonesia because it directly creates the dependence of imported raw materials that impact on efficiency issues (Nandiyanto et al., 2018).

Efficiency is an important indicator in measuring the overall performance of a company's activities (Bunse, Vodicka, Schönsleben, Brülhart, & Ernst, 2011). Measurement of efficiency can provide an assessment of the good operation of a company or organisation. An efficient organisation requires minimal use of input resources to produce output. Many commonly used efficiency measurement tools, for example, use the Cobb-Douglas production function with a solution via an econometric model, the Constant Elasticity Of Substitution (CES) model, and in the latest development of efficiency measurements using Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) (Chiona, Kalinda, & Tembo, 2014; Ebrahimnejad, Tavana, Lotfi, Shahverdi, & Yousefpour, 2014; Aggrey, Eliab, & Joseph, 2010; Andersen & Petersen, 1993; Bessent, Bessent, Kennington, & Reagan, 1982; Charnes, Cooper, & Rhodes, 1978). The SFA method was developed by Aigner, Lovell and Schmidt (1977), Andersen & Petersen (1993), and Chiona et al., (2014). Stochastic frontier analysis is one of the methods used in estimating production limits (frontier) and also measuring the level of production efficiency (Chiona et al., 2014). This analysis uses a parametric approach and uses econometric methods in calculating efficiency. An efficiency measurement consists of two components: 1) technical efficiency that reflects the company's ability to get maximum output from a set of available inputs, and 2) allocative efficiency that reflects the ability of the firm to use input in the optimal proportion in accordance with the cost (Coelli, 1996, 2007).

Measurement efficiency in this case assumes that the firm's optimal production function is known. In practice, the production function is unknown, and it is estimated from the sample data using either a non-parametric piece-wise linear technology or parametric function like Cobb Douglas function (Coelli, 1996, 2007).

Based on the phenomenon, it is necessary to study the implementation of Stochastic Frontier Analysis (SFA) to measure the Efficiency of Chemical industry in Indonesia. This study is intended to collect and process data related to chemical industry efficiency in Indonesia so the following can be known 1) general description of input variables (capital, labour, raw

materials, fuel and auxiliary materials) and output variables (outputs) 2) Compare whether the performance value characteristics of the DEA model have similarities with traditional model performance results, and 3) Identify the efficiency of food production the DEA model and its causal factors.

This study is expected to contribute to theory and praxis (development of knowledge and practical aspects respectively). From the theoretical aspect (science development), this research is expected to increase knowledge insight, especially related to micro economy and efficiency measurement using Stochastic Frontier Analysis approach. For practical aspects, this research provides recommendations for policy makers (government) and the chemical industry to improve efficiency.

LITERATURE REVIEW

The definition of efficiency was first introduced by Debreu (1951) and Koopmans (1951) and has been widely used in the production and production efficiency literature t. Efficiency is input use in producing the most optimal output. It is a condition where people get the most goods / services that can be obtained from scarce resources (Farrell, 1957). In macroeconomics, the general equilibrium approach is a good tool used to calculate efficiency (Mankiw, 2014; Nicholson & Snyder, 2008). The efficiency of a company can be measured from its technical and allocative aspects. Technical efficiency is the company's ability to produce the maximum output from an existing set of inputs. Allocative efficiency is the company's ability to use inputs in optimal proportions, at existing prices. Both types of efficiency are then combined to produce total economic efficiency (Mankiw, 2014; Coelli, 1996).

Technical efficiency is the company's ability to produce maximum output from an existing set of inputs. Allocative efficiency is the company's ability to use inputs in optimal proportions, at existing prices. Both types of efficiency are then combined to produce total economic efficiency (Debreu, 1951; Koopmnas, 1951). Two approaches in the calculation of efficiency are output oriented and input oriented. The output is to maximise the output produced with a fixed set of inputs while the input is to minimise input which is used to produce a fixed output.

Figure 1 shows the differences in technical and allocative efficiency of the input and output approaches. This analysis assumes that the firm produces two inputs (x1 and x2), yielding one y output at full efficient condition at the firm. In the left image, the SS line represents the firm's production isoquant, while the AA line represents the price-input ratio. The 0P line represents the input used in the production process. Then QP represents the number of inputs that can be reduced without reducing output. Thus, what is meant by technical efficiency in the input approach is a comparison between 0Q and 0P.

While the RQ line represents a reduction in production costs that will occur if the company operates efficiently in terms of allocation at the point Q. Therefore, the allocative efficiency can be formulated with 0R versus 0Q. The combination of both types of efficiency is total efficiency. Total efficiency can be formulated with 0R / 0P. The output-based efficiency is shown in the right figure. ZZ line is a possibility of production. The point AB represents technical inefficiency, because with fixed inputs it can produce output at a higher level. Then the measurement of technical inefficiency is 0A / 0B. If the price is known, then the DD line

can represent it. Then it can be written that 0B / 0C is a measure of the allocative efficiency. Thus, the overall economic efficiency can be calculated by multiplication between TE and AE. Some studies have focused on the calculation of technical efficiency and determinant factors.

(a)   (b)
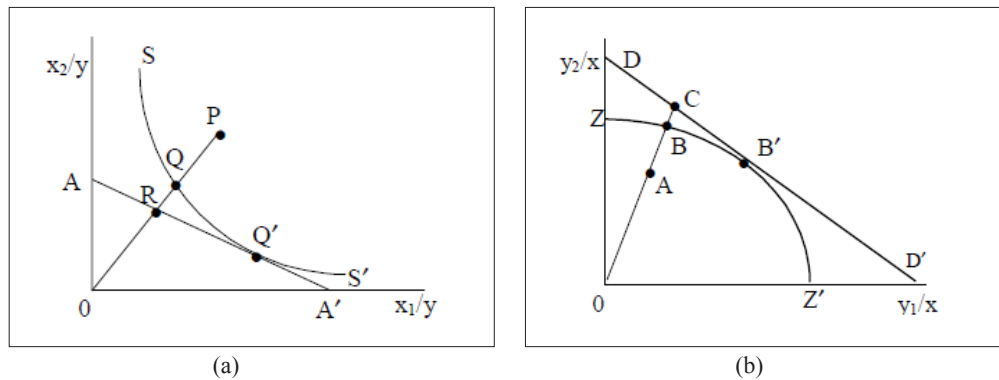
Figure 1. Technical and allocative efficiency, a) input oriented efficiency, b) output oriented efficiency. Adapted from *The measurement of productive efficiency and productivity growth* (pp. 24-25), by H. O. Fried, C. K. Lovell and S. S. Schmidt, 2008, New York, NY: Oxford University Press. Copyright 2008 by Oxford University Press.

An earlier study focused on the food, textile, chemical and metal manufacturing industries in Indonesia (Margono & Sharma, 2006). It uses variable determinants of ownership, company location, size, and age of company. In the food sector, significant ownership factors affect technical inefficiency; in the textile sector, location and size of the company; in the chemical and metals sectors, size of ownership, and age.

A study had looked at the determinants of technical efficiency in the manufacturing sector in Germany. The variables used are industrial effects, ownership, age, legality, ownership structure, size, location, and outsourcing. The industry effect variable is the major determinant of technical efficiency, followed by the size and location variables. Other variables have a very small determining role.

Earlier research also focused on factors affecting technical efficiency in the electronics and optical manufacturing industries in Ireland (Uğur, 2003). This research used data panels with SFA approach. The variables used are investment intensity, worker quality, export intensity. The results of the research show intensity of investment and the quality of the workers affect significantly the overall technical inefficiency. The intensity of exports only signifies.

METHODS

The present study used a survey explanatory method with panel type data of 4240 inputs and output level of companies during six years of observation (2010-2015). Determinants of technical efficiency (TE) are estimated using multiple regression. This research will combine the use of determinant variables that have been used (Aggrey et al., 2010; Uğur, 2003).

Market share variables, company age, capacity utilisation, company ownership, and concentration ratio are variables used in this study.

This study uses the main analysis tool of Stochastic Frontier Analysis (SFA) which calculates technical efficiency (TE) and the factors that affect it. Technical efficiency and the determinant variable are then estimated using Maximum Likelihood Estimation (MLE). The data used are panel data consisting of 4240 inputs and output level companies (DMU) within 6 years of observation (2010-2015). Technical efficiency is calculated using four input variables, namely labour (L), capital (K), raw material resources (M), fuel resources (E). The four variables are then used in the frontier production function by using the trans log variable specification. The model used is:

$$
\begin{aligned}
LnY = \ & \beta_0 + \beta_1 LnK_{ij} + \beta_2 LnL_{ij} + \beta LnM_{ij} + \beta_4 LnE_{ij} + \beta_5 LnT_{ij} \\
& + \beta_6 LnL_{ij}{}^2 + \beta_7 LnK_{ij}{}^2 + \beta_8 LnM_{ij}{}^2 + \beta_9 LnE_{ij}{}^2 \\
& + \beta_{10} LnT_{ij}{}^2 + \beta_{11} LnT^2{}_{ij} + \beta_{12} LnT_{ij} LnL_{ij} \\
& + \beta_{13} LnT_{ij} LnK_{ij} + \beta_{14} LnT_{ij} LnM_{ij} + \beta_{15} LnT_{ij} LnM_{ij} \\
& + \beta_{16} LnT_{ij} LnE_{ij} + \beta_{17} LnL_{ij} LnK_{ij} + \beta_{18} LnL_{ij} LnM_{ij} \\
& + \beta_{19} LnL_{ij} LnE_{ij} + \beta_{20} LnM_{ij} LnE_{ij}
\end{aligned}
$$

where Ln is natural logarithm, T is time (period), is variable parameter. While the inefficiency model used is

$$
TE_{ij} = \delta_0 + \delta_1 MS_{ij} + \delta_2 AGE_{ij} + \delta_3 CU_{ij} + \delta_4 DLOC_{ij} + \delta_5 CR4_{ij}
$$

where TE is technical efficiency, MS is market share, AGE is firm age, DLOC is dummy local ownership (0 = local, 1 = foreign), CR4 is company ratio concentration. Both technical efficiency calculations and regression determinants are done by using the TE effect approach in Frontier 4.1 software.

RESULTS

Calculations using stochastic Frontier Analysis are shown in Table 1, Table 2 and Table 3.

Table 1
Summary of descriptive statistics estimated technical efficiency

Year	Mean	Std. Dev.	Min.	Max.
2010	0.9521	0.0170	0.6596	1
2011	0.9515	0.0167	0.6603	1
2012	0.9500	0.0239	0.5401	1
2013	0.9499	0.0188	0.5414	1
2014	0.9506	0.0183	0.5409	1
2015	0.9497	0.0175	0.5403	1

Table 2
Average technical efficiency based on determinant variables

Variable	2010		2011		2012		2013		2014		2015	
	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total
AGE (Year)												
1-20	0.956	2860	0.956	2752	0.954	2588	0.955	2453	0.956	2331	0.96	2220
21-40	0.946	1167	0.946	1267	0.946	1417	0.946	1541	0.947	1651	0.95	1739
41-60	0.933	162	0.934	170	0.933	180	0.932	188	0.933	196	0.93	212
61-80	0.919	33	0.918	33	0.920	37	0.918	38	0.918	40	0.92	46
>81	0.910	18	0.915	18	0.912	18	0.910	20	0.912	22	0.91	23
LOC												
0 (local)	0.949	3862	0.949	3862	0.947	3862	0.947	3862	0.948	3862	0.95	3862
1 (foreign)	0.982	378	0.982	378	0.981	378	0.981	378	0.982	378	0.98	378
CR4 (%)												
0-10	0	0	0.945	175	0.0	0	0	0	0.0	0	0	0
10.1-25	0.951	3070	0.950	2541	0.949	2806	0.948	3070	0.948	2145	0.95	2287
25.1s-50	0.955	936	0.953	1282	0.951	1093	0.953	928	0.952	1497	0.95	1355
50.1-65	0.978	58	0.981	16	0.957	132	0.958	202	0.957	558	0.96	558
>65.1	0.958	176	0.962	226	0.958	209	0.971	40	0.968	40	0.97	40
CU												
1-20	0.989	41	0.987	21	0.987	36	0.987	31	0.986	36	0.99	34
21-40	0.979	116	0.978	116	0.977	116	0.975	105	0.976	111	0.98	122
41-60	0.965	473	0.965	432	0.963	526	0.962	430	0.964	418	0.96	409
61-80	0.953	2038	0.953	2093	0.952	1951	0.952	2087	0.952	2160	0.95	2127
>81	0.944	1572	0.944	1578	0.940	1611	0.942	1587	0.943	1515	0.94	1548

Table 3
Estimation parameters of stochastic production frontier and technical inefficiency model

Variable	Parameters	Coefficient	Standard Error	t-ratio
Stochastic Production				
Intercept	β0	4.000391	0.995273	4.019391
LnL	β1	0.873324	0.967823	0.902359
LnK	β2	0.056982	0.726318	0.078454
LnM	β3	0.090938	0.780713	0.116480
LnE	β4	0.261519	0.815973	0.320500
T	β5	0.024975	0.969357	0.025765
LnL^2	β6	0.033911	0.449802	0.075390
LnK^2	β7	0.010278	0.022934	0.448170
LnM^2	β8	0.071200	0.129237	0.550924
LnE^2	β9	0.035602	0.162339	0.219306

Table 3 *(continue)*

Variable	Parameters	Coefficient	Standard Error	t-ratio
T^2	β10	-0.005953	0.051172	-0.116334
T*LnL	β11	0.003010	0.030288	0.099369
T*LnK	β12	-0.001187	0.079861	-0.014859
T*LnM	β13	-0.002868	0.102757	-0.027906
T*LnE	β14	0.004884	0.085168	0.057341
LnL*LnK	β15	0.011596	0.390880	0.029667
LnL*LnM	β16	-0.087875	0.561817	-0.156412
LnL*LnE	β17	0.010884	0.215669	0.050467
LnK*LnM	β18	-0.019793	0.428080	-0.046237
LnK*LnE	β19	-0.003460	0.409366	-0.008452
LnM*LnE	β20	-0.067827	0.194989	-0.347849
Technical Inefficiency Model				
Intercept	δ0	-0.002159	0.486978	-0.004433
MS	δ1	-0.001784	0.000309	-5.777606
AGE	δ2	0.000687	0.006107	0.112544
CU	δ3	0.000632	0.002158	0.292972
DLOC	δ4	-0.032479	0.189350	-0.171530
CR4	δ5	-0.000195	0.002645	-0.073604
Other parameters				
Sigma-squared		0.148197	0.058142	2.548885
Gamma		0.000008	0.000109	0.076986
Log likelihood function		-11970.72		
Mean technical efficiency		0.95065		

DISCUSSION

Table 1 is an overview of the technical efficiency of the manufacturing industry in Indonesia. The smallest TE in the observational study is 0.54. This means that for every single input used, it will produce an output of 0.54. The largest number is 1, not least for all years. This means that in every year there are always companies that achieve full technical efficiency. The average indicates that there is no year with a TE of less than 0.94. This figure is supported by a small standard deviation (0.01-0.02). A small standard deviation indicates the resulting slope / fault of the resulting average is relatively small. Of the 100 estimated data that has an average of between 0.94-0.95 only 1-2 data that is not the number. Therefore, it can be said that the average number represents the population, i.e. the average TE of 0.94-0.95 for all years of observation.

Table 2 provides an overview of the association. Judging by its relation to the year variable, technical efficiency for all firms in all years has a downward trend for firms with longer standing ages. Companies aged 1-20 years have an average TE of 0.95-0.96. Companies with more than 81 years of age have an average TE of about 0.91. Therefore, it can be assumed the older the company is, the lower its TE. Additionally, the company with foreign ownership has a larger

TE value than the company with local ownership. In all years of observation, local companies have TE values ranging from 0.94-0.95. Companies with foreign holdings have a TE value of 0.98. Hence, it can be concluded foreign companies are working at a higher level of efficiency than local companies. In terms of concentration ratios, in the first three years of observation all firms have an average TE that varied between 0.95-0.98. The last three years of observation have certain TE value patterns. Companies with higher concentration ratio values have higher TE values. The value varies between 0.94 and 0.97. Thus, the higher the CR4, the higher TE is expected to only be valid for the last three years of observation.

In terms of capacity utilisation, during the observation period, firms showed a higher CU pattern, a lower TE. The average value of TE varies between 0.94-0.99. The higher CU signifies the company is operating at a higher capacity. Therefore, it can be concluded that high production capacity of the company leads to reduced TE. The estimated parameters of TE with selected input sets are shown in Table 3. The approach used is MLE with transcendental logarithmic model. Table 3 provides an overview of the stochastic production estimation coefficients for the samples in this study. The dependent variable used in this estimate is TE. Therefore, a significant positive relationship between dependent and independent variables has a higher independent variable meaning will cause higher TE numbers. This means the company is efficient. This result is in line with that of previous studies (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

MS is suspected to have a positive relationship with TE; the former is calculated based on the ratio between company's sales to the overall sales of the industry. Increased sales means that the company operates on a larger scale. Few competitors means companies with high MS will lead to an oligopoly market. Thus, the company will increasingly compete to improve the efficiency of its production. The estimation results show a significant negative relationship between MS and TE at the 95% probability level. Therefore, the initial hypothesis is rejected. The higher the MS value of the company, the lower its TE value. In other words, it will lead to inefficiency. This result is in line with previous studies (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

Age of the company is suspected to have a positive relationship with TE. The longer the company operates, it will become more specialised. It will improve the technical efficiency of the company. The estimation results show a positive and insignificant relationship between AGE and TE at the 95% probability level. Therefore, it can be said the company's age does not significantly affect its technical efficiency. Production Capacity or CU is suspected to have a positive relationship with TE. This result is in line with previous researches (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

Higher CU means the companies are more productive and achieving full capacity. The debate over the measurement of CU is due to the absence of a definite measure of full capacity in production, the subject matter of full production capacity assessment, and the valuation policy used to measure output. Regardless of the situation, the more companies operate at full capacity, the higher the company's TE. The estimation results show a positive and insignificant relationship between CU and TE at 95% probability level. This means high-low production capacity achieved by the company does not significantly affect technical efficiency of the

company. This result is in line with previous research (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

The DLOC is a dummy variable of company ownership. Domestic-owned companies are assigned a value of 0 and foreign-owned companies are assigned a value of 1. There is no firm hypothesis regarding the relationship between ownership of a company and TE based on citizenship. The results show the coefficient is not significant and therefore, there is no strong evidence to suggest that both local and foreign companies have a higher level of technical efficiency. This result is in line with previous researches (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

The CR4 is the ratio of unbalanced concentration of firms in an industry. It is based on the output share (or sales) of the four largest companies in the industry. It is said not weighted because there is no weighting for large and small companies. The CR4 is suspected to have a positive relationship with TE. A bigger CR4 means the company is producing at a higher level. Thus, the company seems to operate on the oligopoly market. With such a market model the company tries its utmost to improve production efficiency in order to win the competition. This result is in line with previous researches (Aggrey et al., 2010; Margono & Sharma, 2006; Uğur, 2003).

The results show a negative and insignificant relationship between CR4 and TE at 95% probability level. This means that it can be said that CR4 affects TE insignificantly in the sample of this study. Technical inefficiency and other random factors may have influenced the variation in errors. The value of gamma coefficients close to 1 indicates that the variation of the errors is influenced by technical inefficiency. The value of the gamma coefficient away from 1 indicates the variation of error is more influenced by other random factors. The results show the gamma coefficient has a value relatively far from 1. This means random factors other than independent variables and technical inefficiency dominate the variation of error.

CONCLUSION

This research focuses on technical efficiency in chemical industry in Indonesia and its determinant variables. Technical efficiency is calculated using the stochastic frontier analysis (SFA) approach. Determinants of technical efficiency are estimated by using multiple regression. Both technical efficiency calculations and regression determinants are done by one step using the TE effect approach in Frontier 4.1 software. The result of research indicates that market share variable significantly and negatively affects technical efficiency. Other variables such as company age, business ownership, ratio concentration, and capacity utilisation are not significant. This finding implies that to improve the efficiency, the chemical industry needs to increase its market share.

The following are findings of the present study to enhance the growth of chemical industry in Indonesia:

1) Achieve synergy between medium / large industry with small / micro scale industry. The growth of Small / micro industry is vital to boost the growth of medium / large industries. Big industry players and the government can take on this role so that small / micro

industries have direction and production standards that meet the needs of large industries. The medium / large industry can support the development of small / micro industries.

2) Creating a road map for the chemical industry in Indonesia. Thus, will form a chain from upstream to downstream which reduces the amount of imported raw materials. In the implementation, the establishment of new factories can be pursued for industries that process raw materials . Providing incentives to the raw material processing industry can encourage investors to follow the scenario. Additionally, the import duties on raw materials can be increased so that importers will be more keen to source local raw materials.

3) Unnecessary cost cuts, especially licensing fees so investors can reallocate their budget to drive technology development. The latest technology will boost efficiency which will bring down selling prices. In addition, the availability of local raw materials at lower prices will further boost product competitiveness.

ACKNOWLEDGEMENTS

REFERENCES

Aggrey, N., Eliab, L., & Joseph, S. (2010). Firm size and technical efficiency in East African manufacturing firms. *Current Research Journal of Economic Theory, 2*(2), 69-75.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21-37.

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science, 39*(10), 1261-1264.

Bessent, A., Bessent, W., Kennington, J., & Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science, 28*(12), 1355-1367.

Bunse, K., Vodicka, M., Schönsleben, P., Brülhart, M., & Ernst, F. O. (2011). Integrating energy efficiency performance in production management–gap analysis between industrial needs and scientific literature. *Journal of Cleaner Production, 19*(6-7), 667-679.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444.

Chiona, S., Kalinda, T., & Tembo, G. (2014). Stochastic frontier analysis of the technical efficiency of smallholder maize farmers in Central Province, Zambia. *Journal of Agricultural Science, 6*(10), 108.

Coelli, T. J. (1996). *Centre for Efficiency and Productivity Analysis (CEPA).* Working Papers, Armidale: University of New England.

Coelli, T. J. (2007). *A guide to frontier version 4.1: A computer program for stochastic frontier production and cost function estimation.* Working Paper, Armidale: University of New England.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica, 19*, 273– 292.

Ebrahimnejad, A., Tavana, M., Lotfi, F. H., Shahverdi, R., & Yousefpour, M. (2014). A three-stage data envelopment analysis model with application to banking industry. *Measurement, 49*, 308-319.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, 120*(3), 253-290.

Fried, H. O., Lovell, C. K., & Schmidt, S. S. (2008). *The measurement of productive efficiency and productivity growth.* New York, NY: Oxford University Press.

Koopmans, T. C. (1951), *Activity analysis of production and allocation.* New York, NY: Cowles Commission Monographs, John Willey & Sons.

Mankiw, N. G. (2014). *Principles of economics* (7th ed.). Stamford, USA: Cengage Learning.

Margono, H., & Sharma, S. C. (2006). Efficiency and productivity analyses of Indonesian manufacturing industries. *Journal of Asian Economics, 17*(6), 979-995.

Nandiyanto, A. B. D., Maulana, A. C., Ragadhita, R., & Abdullah, A. G. (2018). Economic evaluation of the production ethanol from cassava roots. *IOP Conference Series: Materials Science and Engineering, 288*(1), 012023-1 – 0123023-6.

Nicholson, W., & Snyder, C. (2008). *Microeconomic theory: Basic principles and extensions* (10th ed.). Mason, Ohio: Nelson Education.

Plechkova, N. V., & Seddon, K. R. (2008). Applications of ionic liquids in the chemical industry. *Chemical Society Reviews, 37*(1), 123-150.

Tombs, S., & Whyte, D. (2003). Unmasking the crimes of the powerful. *Critical Criminology, 11*(3), 217-236.

Uğur, A. (2003). *Technical efficiency in Irish manufacturing industry, 1991-1999.* Working Paper, Dublin: Trinity College Dublin.

Validate UML Model and OCL Expressions Using USE Tool

Arifa Bhutto* and Dil Muhammad Akbar Hussain

Department of Energy and Technology, Aalborg University, Esbjerg, Denmark

ABSTRACT

Verification and validation of system models at design level has a huge impact on the quality of the system software engineering process. In general, system modeling and designing Unified Modelling Language (UML) is a standard for the design models of the systems. However, verification and validation of UML models at early design level is not available, but somehow Object Constraint Language (OCL) constraints are defined at the class level to ensure that the model is correctly designed. As for the static–dynamic structure, there is no such mechanism defined in UML/OCL that has a huge impact on the development of the software. Our research is focused on providing verifiable UML/OCL models. Our approach using UML-based Specification Environment (USE) for UML class model is integrated with the OCL constraints to check if the model is correctly designed as well as constraints for verification and validation. In USE, the output is shown as the verifiable UML/OCL models by visual graphical models.

Keywords: OCL constraints, static and dynamic, Unified Modelling Language (UML) and Object Constraints Language (OCL), verification and validation

INTRODUCTION

Unified Modelling Language (UML) is the co-stander for development of the software engineering system. UML is based on different models which are described at different levels of abstraction of the system. However, in UML models for verification, various types of languages are integrated to make sure the model is correctly designed and developed; among them Object Constraint Language (OCL) is used for the verification and validation of static and dynamic properties of the design models defined by Richters and Gogolla(2002), for further model-driven approaches discussed by Dang and Gogolla (2009).

However, there are various tools designed for the verification and validation of the UML/OCL, like OCL constraints solver for UML/OCL that defines the Java API for the class

model using java programming for verification of OCL expression and satisfying the model condition defined by Ali, Iqbal, Arcuri and Briand (2011). UML2Alloy, another project started in 2005, is a scientific approach to use two critical functionalities of the Alloy Analyser to simulate and verification of the UML class diagram contained OCL constraints to ensure that the properties of the model are satisfied by Anastasakis, Bordbar, Georg and Ray (2007), and Cabot, Clarisó and Riera (2014).

UMLtoCSP (Cabot, Clarisó, & Riera, 2007) is a tool for automatic verification of UML models with OCL constraints. The model checks correctness for properties of the models. UMLtoCSP tool checks for satisfiability and contradictory constraints. It supports only UML class diagrams and shows a graphical view of the system (Cabot et al., 2007).

The research is focused on the verification of the UML/OCL expressions by USE tool (Gogolla, Büttner, & Richters, 2007). The tool supports the graphical view of the model and also checks the OCL constraints by applying variants and Pre–Post conditions by logical reasoning illustrated by Doan and Gogolla (2018). The following are the steps as illustrated in Figure 1.
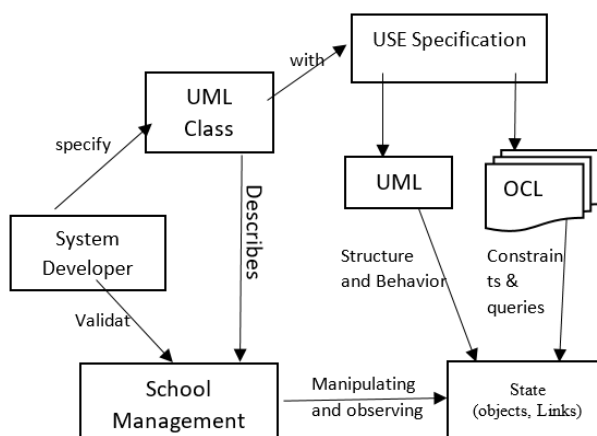
Figure 1. UML/OCL verification and validation using USE tool

Step 1

The design of the application is described by the structure model in UML class, object diagram and behavior of the system including the constraints by the OCL. The USE specification is described in the UML classes, attributes, operations, and association. The class diagrams integrated with OCL constraints as form of invariants and Pre–Post conditions (Gogolla et al., 2007).

Step 2

Open the USE specification model that generates the graphical view of the class model including the relationships among the classes.

Step 3

Check if the model structure is correct, then USE will verify behavioral properties of the system model by analyzing the object diagram, which generates the sequence diagram in connection.

Step 4

The USE model checks the UML class diagram relations, OCL constraints, invariants, and Pre–Post condition to make sure that all OCL constraints are applied correctly and respond to the violation of the constraints as form of the error detected.

In section II, we will further discuss the aforementioned steps in a detailed case study.

CASE STUDY

This section describes the small application model to check and verify the designed approach by using the UML 2.2 and USE tool. The School Management System (SMS) is a complete framework for the management of the school. The main functions of the application are the following models:

- Student admission module

- Class distribution and timetable module

- Employee management module

- Payroll management module

- Accounts management module

- Generating reports

However, the detailed design of all model components of SMS is shown in Figure 2.

VERIFYING METHODOLOGY

Consider the aforementioned study first; we design the component diagram of the case study using UML in Figure 2 and it represents the SMS requirement models of the case study in detail. The internal structure of the components model is based on the object class structure, therefore we design the class structure of the SMS case study using UML, shown in Figure 3. According to the real requirements, class diagrams represent the entities, data types, functionalities, and relationships of various entities. In UML, class structure shows the data types, operations, and relationships. Using the example of our case study, we developed the models describing employees, departments, accounts, with attributes, operations, and relationships of the classes as shown in Figure 3. Furthermore, we integrated the class diagram with OCL constraints, which we described in the natural language. Using OCL language, these constraints are integrated with the UML class diagrams.
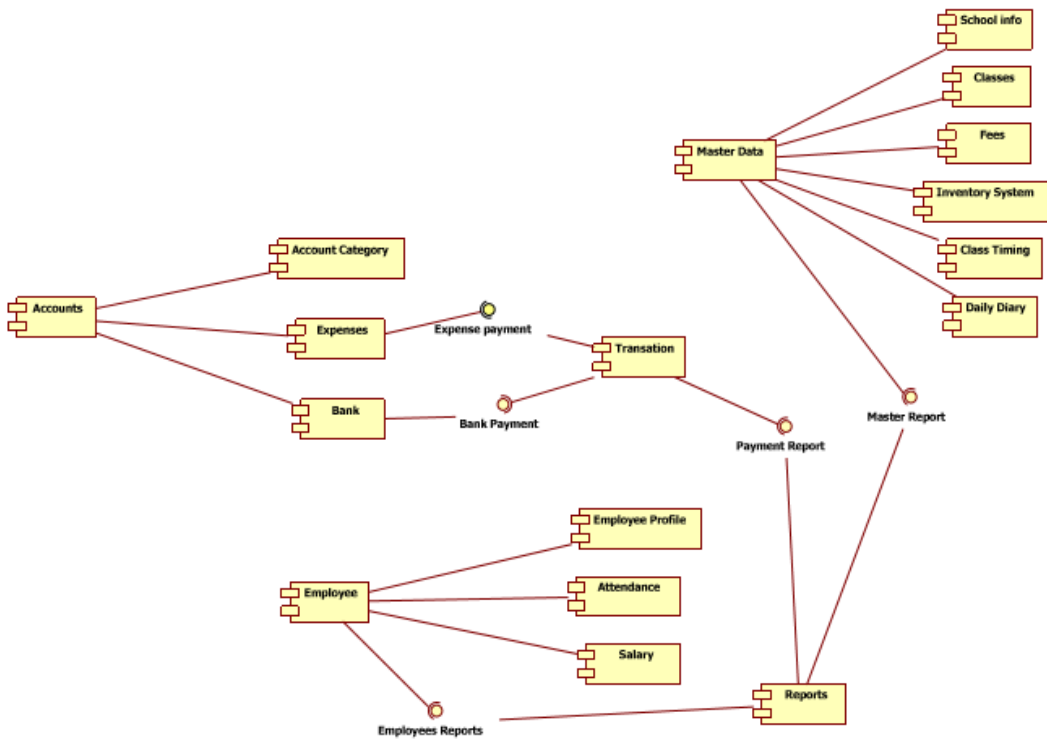
Figure 2. School Management System (SMS) in UML components diagram
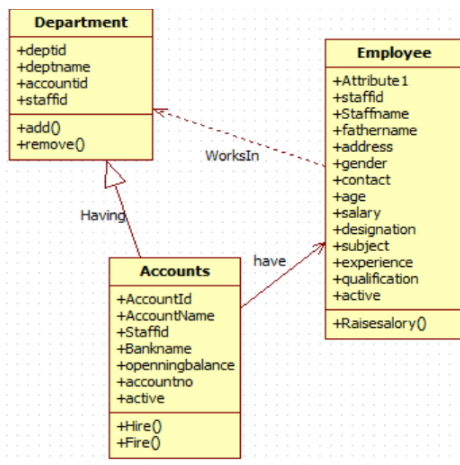
Figure 3. UML class diagram of School Management System (SMS)

Constraints in natural language:

1. *The employer must be working in a department.*

2. *The employer age is not equal to or greater than 20 years.*

3. *Employer accounts must be active.*

4. *If employee's account opening balance is greater than or equal to 5000.*

5. *Employer accounts must be active.*

6. *Employer must be working in department.*

The research methodology is based on selecting the USE tool, which validates the UML class diagram, generating object diagram, sequence diagram from the USE specification. The system also validates the model constraints in the form of OCL, variants, Pre–Post conditions in running the environment automatically.

USE specification of UML class model of SMS Case Study

The USE tool defines the model description in textual format with the extension of use. The UML class diagram is translated into the USE specification that is shown below. Classes, attributes, operations, association, constraints, and Pre–Post conditions in the textual format are listed below:

-- Example illustrating Pre-and Post-conditions

```
model School_Management_System
-- classes
class Employee
attributes
  staffid : Integer
  staffname : String
  fathername : String
  address : String
  gender : String
  contect : Real
  age : Integer
  salary : Real
 designation : String
 subject : String
 experience : Integer
 qualification : String
 active : Integer
 accountid : Integer
 deptid : Integer
```

```
operations
  raiseSalary(rate : Real) : Real
end
class Accounts
attributes
  accountid : Integer
  accountname : String
  staffid : Integer
  bankname : String
  openingbalance : Real
  accountno : Real
  active : Integer
  deptid : Integer
operations
  hire(e : Employee)
  fire(e : Employee)
end
class Department
attributes
    deptid : Integer
    deptname : String
    accountid : Integer
    staffid : Integer
operations
    add (a : Accounts)
    remove (a : Accounts)
end
-- associations
association WorksIn between
  Employee[*]
  Department[1..*]
end
association Controls between
  Department[*]
  Accounts[*]
end
association Having between
  Employee[1]
  Accounts[*]
end
```

```
-- constraints
constraints
context Employee inv:
  age >= 20
 context Department inv:
     staffid = 1
context Accounts inv:
 accountid=active


     context Accounts inv:

     openingbalance >= 5000

     context Employee inv:
     staffid = deptid

     context Accounts inv:
         staffid= accountid
 context Employee::raiseSalary(rate : Real) : Real
  post raiseSalaryPost:

     salary = salary@pre * (1.0 + rate)
   post resultPost:
     result = salary


 context Accounts::hire(e : Employee)
   pre  hirePre1: e.isDefined()
   pre  hirePre2: employee->excludes(e)
   post hirePost: employee->includes(e)


 context Accounts::fire(e : Employee)
   pre  firePre:  employee->includes(e)
   post firePost: employee->excludes(e)
```

After writing the aforementioned specifications in textual form, USE model opens the specifications shown in Figure 3. GUI environment is shown graphically in class diagram, association, OCL constraints, variants, and Pre–Post conditions.

Using GUI environment, models are viewed and the errors are found by messages, structure is correct or not as well as define number of classes, association, invariants, Pre–Post conditions as shown in Figure 4. In the next section, we validate the model using the OCL constraints by applying the object data and runtime we check the results of the UML object, class models
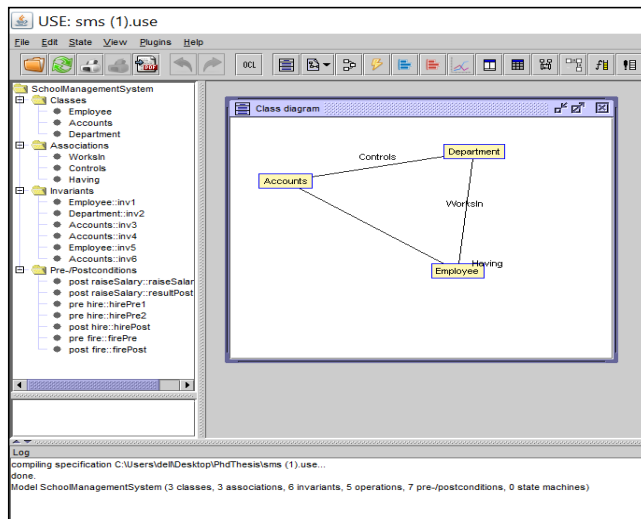
Figure 4. USE tool generating class diagram and association

CREATE OBJECT AND VALIDATE THE DATA

In the USE tool, we can create the object by selecting the option create object shown in Figure 5. We can also directly set the data in the object model by using the option add data as shown in Figure 6 in USE model, we can set the data by command prompt as shown in the listing below:

```
use> !create Accounts : Department
use> !create mehran : Employee
use> !set mehran.staffname :='Ali'
use> !set mehran.salary :=1500000
use> !set mehran.staffname :='Ali'
use> !set mehran.salary :=1500000
use> !set mehran.age :=24
```
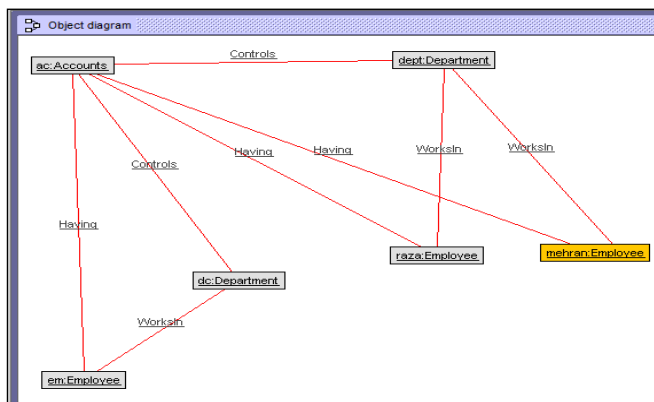

Figure 5. Object diagram of SMS in USE tool
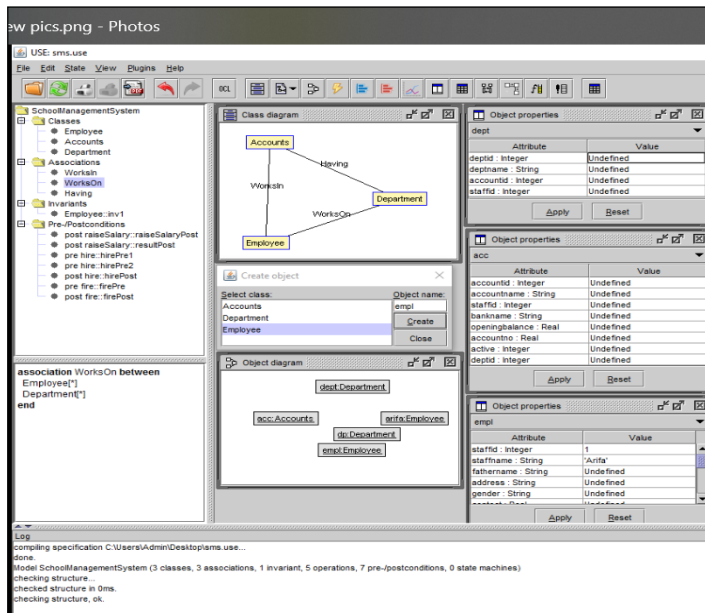
Figure 6. Class, object diagram with data, and properties of the objects

DEFINE MODEL INHERITANCE CONSTRAINTS

Model also checks the constraints by the form of invariants if the constraints are violating the system by giving the message: *Model inherent constraints violated.* In the USE, invariants messages show that there is a problem in the state; we can view the message of the invariants violation in the GUI panel shown at the bottom of the screen. The command prompt also has check command to view the violated comments as shown in the listing below

```
use> check
Multiplicity constraint violation in association
`WorksIn´
  Object `Accounts´ is connected to 0 object(s) of class
`Department´
  but the multiplicity is specified as `1. *'
Multiplicity constraint violation in association
`WorksIn´
  Object `Mehran´ is connected to 0 object(s) of class
`Department´
  but the multiplicity is specified as `1.. *'
```

According to the above class diagram in UML, it is specified that each employee works in the department. USE GUI model shows the object diagram; there is no employee who works in the department in association link. We can fix this in GUI model as well as by command prompt by just typing the insert command for inserting the missing association listed below:

```
use>!insert (mehran, Accounts) into worksIn
```

The USE GUI environment connected the link directly by inserting the object using the object diagram option or inserting the command directly from the command prompt.

The USE diagram, Figure 7, shows the red link in the object diagram between department and employee object, which is the violation of invariants; we can fix it by applying the correct rules for invariants, which is shown in Figure 8 of the model diagrams using USE GUI.

METHOD FOR CHECKING INVARIANTS

Checking class invariants automatically, the USE graphical interface has a capability to view the changes directly. We can change the status and the GUI shows the results. OCL invariants can be analyzed by applying the commands listed below:

```
Department::hire(self:Accounts, e:mehran) [caller:
openter Accounts hire(mehran)@<input>:1:0]
use> info vars
[frame 1]
  e : Employee = mehran
  self : Department = Accounts
[frame 0]
  empty
[object variables]
  Accounts : Department = Accounts
  mehran : Employee = mehran
use> !insert (e,Accounts) into WorksIn
use> !openter mehran raiseSalary(0.1)
use> !set self.salary := self.salary+ self.salary * rate
use> !create Accounts : Department
use> !create mehran : Employee
use> !set mehran.staffname :='Ali'
use> !set mehran.age :=24
use> !openter Accounts hire(mehran)
precondition `hirePre1' is true
precondition `hirePre2' is true
use> info opstack
```

```
  precondition `hirePre1' is true
  precondition `hirePre2' is true
  use> info opstack
  1. Department::hire(self:Accounts, e:mehran) [caller:
  openter Accounts hire(mehran)@<input>:1:0]
  use> info vars
  [frame 1]
    e : Employee = mehran
    self : Department = Accounts
  [frame 0]

  empty
[object variables]
  Accounts : Department = Accounts
  mehran : Employee = mehran
use> !insert (e,Accounts) into WorksIn
use> !set e.salary :=2000
use> !opexit
postcondition `hirePost' is true
use> info opstack
no active operations.
use> info vars
[frame 0]
  empty
[object variables]
  Accounts : Department = Accounts
  mehran : Employee = mehran
use> !openter mehran raseSalary(0.1)
<input>:1:0: No operation `raseSalary'found in class
`Employee'.
use> !openter mehran raiseSalary(0.1)
use> !set self.salary := self.salary + self.salary *
rate
use> !opexit 2200
postcondition `raisesalarypost' is true
postcondition `resultpost' is true
```
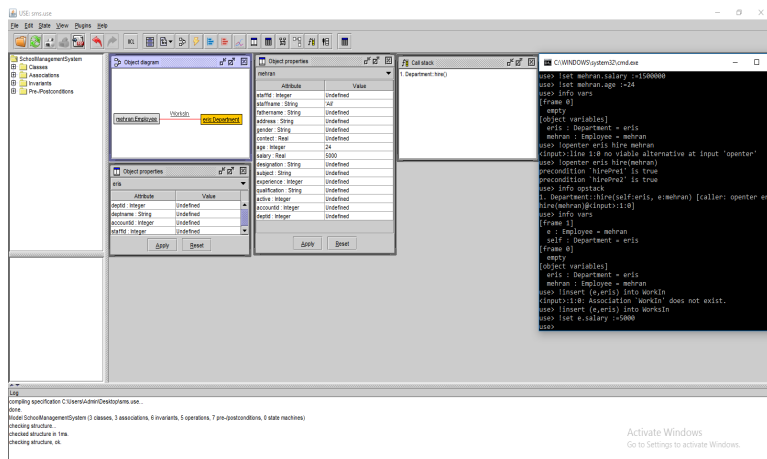
Figure 7. Invariants not satisfied shows in object Link diagram by red line
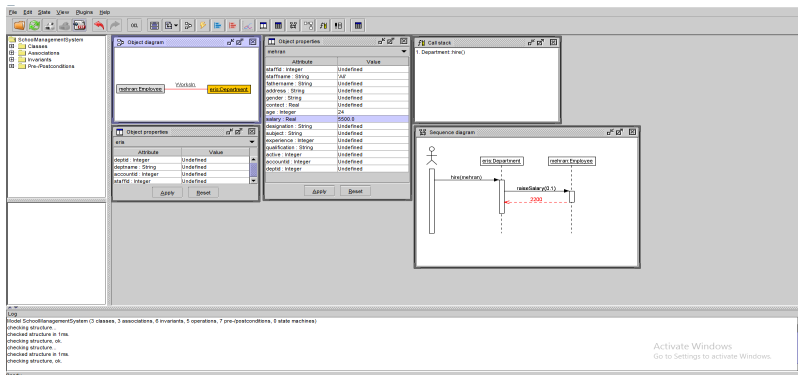
Figure 8. Sequence diagram represent invariants

DISCUSSION

The results of the design model shown by the USE GUI generate the sequence diagram to visualize the variants, Pre–Post conditions by directly applying an object to validate the model according to the data. In Figures 9 and 10, communication model represents the object data including the actual data and relationships of the objects. Figure 11 shows the object, class variants, and counts how many objects are defined as well as class invariants checked- the invariants view of the class by green check that all invariants are correct. Figure 12 shows the complete model behavior and all sub-models like state shows the model status, communication, object, class, class invariants, and sequence. The model output shows that the verification and validation process of the school management system design is done correctly by UML/OCL constraints applied on object class models.

The model shows the red line in Figure 6 when we write the wrong invariants. Figure 7 shows the invariants violation in sequence diagram, which indicates that the model does not provide the right invariants to validate the model.

Figure 9. Sequence diagram represents invariants correct

Figure 10. Communication diagram
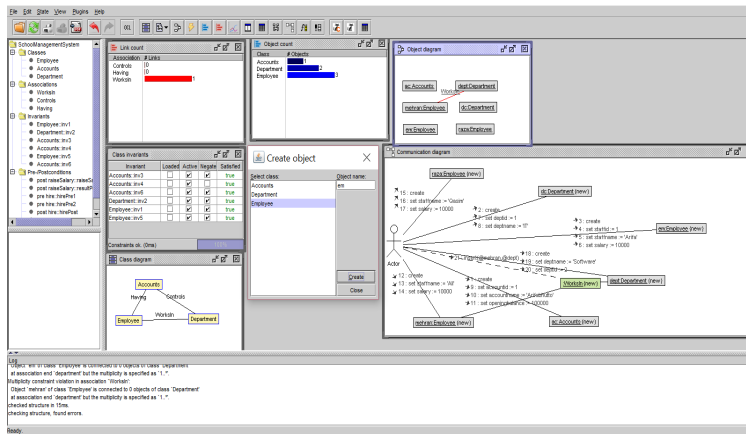
Figure 11. Object, class variants, and counts of the object including sequence diagram (Overall USE Environment)
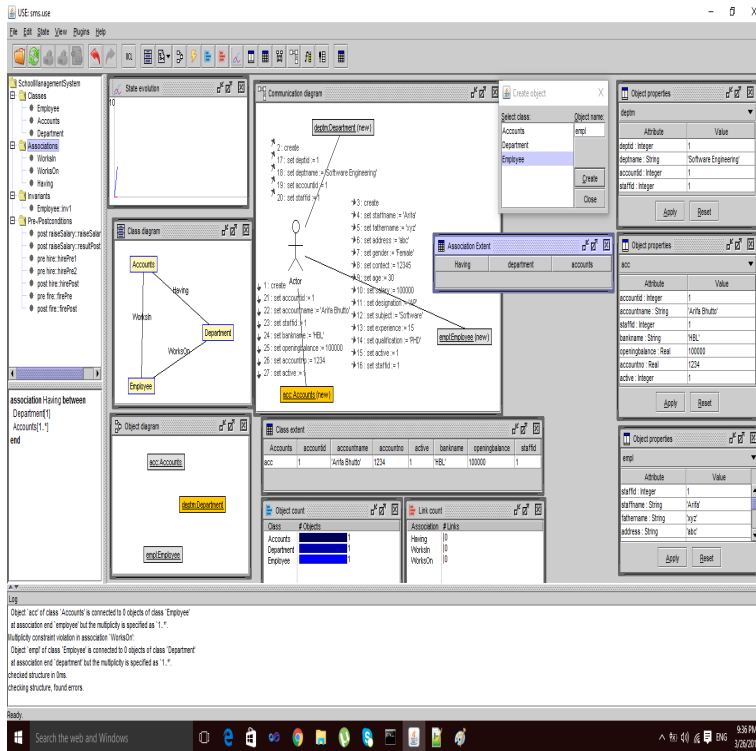
Figure 12. Complete vview of the USE model overview

CONCLUSION

This paper is an approach for verification and validation of structural and behavioral properties of UML model integrated with the OCL constraints. The input we select is USE application that describes the class structure and applies the OCL constraints as a form of invariants. As an output, we found the verifiable UML class, object, and sequence model. We apply this methodology on the small case, which shows very impressive results for the verification and validation of UML models.

Our future work will continue in different directions, like all the models are generated automatically and adding the frame conditions to UML and OCL models.

ACKNOWLEDGEMENT

REFERENCES

Ali, S., Iqbal, M. Z., Arcuri, A., & Briand, L. (2011, July). A search-based OCL constraint solver for model-based test data generation. In *11th International Conference on Quality Software (QSIC), 2011* (pp. 41–50). IEEE.

Anastasakis, K., Bordbar, B., Georg, G., & Ray, I. (2007, September). UML2Alloy: A challenging model transformation. In *International Conference on Model Driven Engineering Languages and Systems* (pp. 436–450). Springer, Berlin, Heidelberg.

Cabot, J., Clarisó, R., & Riera, D. (2007, November). UMLtoCSP: A tool for the formal verification of UML/OCL models using constraint programming. In *Proceedings of the Twenty-Second IEEE/ACM International Conference on Automated Software Engineering* (pp. 547–548). ACM.

Cabot, J., Clarisó, R., & Riera, D. (2014). On the verification of UML/OCL class diagrams using constraint programming. *Journal of Systems and Software, 93*, 1–23.

Dang, D., & Gogolla M., (2009). Precise model-driven transformations based on graphs and met models. In *7th IEEE International Conference on Software Engineering and Formal Methods* (pp. 23–27). IEEE.

Doan, K. H., & Gogolla, M. (2018). Logical reasoning with object diagrams in a UML and OCL tool. In P. Chapman, A. Moktefi, & G. Stapleton, (Eds.), P*roceedings of the 10th International Conference on Diagrams (DIAGRAMS 2018)*. Springer, LNCS, Berlin, Heidelberg.

Gogolla, M., Büttner, F., & Richters, M. (2007). USE: A UML-based specification environment for validating UML and OCL. *Science of Computer Programming, 69*(1–3), 27–34.

Richters, M., & Gogolla, M. (2002). OCL: Syntax, semantics, and tools. In *Object Modelling with the OCL* (pp. 42–68). Springer, Berlin, Heidelberg.

Photonic Crystal Based Micro Mechanical Sensor in SOI Platform

Indira Bahaddur[1]*, Preetha Sharan[2] and P. C. Srikanth[1]

[1]Department of ECE, Malnad College of Engineering, Hassan, India
[2]ECE, Oxford College of Engineering, Bangalore, India

ABSTRACT

Two-dimensional photonic crystals with nano-rod configuration integrated in a silicon-on insulator are analysed in this study. A photonic crystal waveguide suspended over a silicon substrate then weight can be applied on that substrate to change the displacement of substance and to measure sensitivity for pressure in terms of micro units. The overall objective of this work is to detect displacement, which indicates the force applied on the slab with photonic crystals that have line defects. Stress and displacement of the slab reveal the pressure applied. Stress is calculated by the power distribution/excitation in the slab. The displacement of the slab is due to the force, while pressure is determined by the photonic crystal sensor. The quality and sensitivity of the sensor are 1496 and 1200 RIU, respectively. The transmission spectrum is 0.1 micron to 0.5 microns shift, respectively, which are found to be distinct.

Keywords: Photonic crystal (PhC), silicon-on insulator (SOI), micro pressure, stress, optical membrane, nano rods

INTRODUCTION

The silicon-on insulator (SOI) is a fascinating technology introduced some decades ago. It integrates the technique of using a highly sensitive photonic crystal membrane layer and the silicon-on insulation process in a new way and provides a sophisticated design for a photonic crystal-based sensing mechanism. The silicon photonic crystal membrane layer is integrated with circuits on silicon using CMOS technology in a CMOS fabrication process. Better thermal expansion and stability of the insulator in the CMOS (Biallo, Orazio, Sario, Marrocco, & Petruzzelliet, 2006) integration process are the great advantages of silicon photonics. Silicon photonics and photonic crystal sensors have major applications in biochemical sensing,

environmental monitoring, high speed communications and medical diagnostics (Yablonovitch, 1987). The advantages of this technology are low manufacturing cost, a small size, a low weight, low use of power, good reliability and ruggedness.

Photonic technology (Bahaddur, Srikanth, & Sharan, 2006) includes all technology that uses light, creates light, detects light or modifies lights. Light is used to make electricity, for instance in photovoltaic cells for solar panels, while it is created using electricity, as in LED lighting (Joannopoulos, Meade, & Winn, 1995), detected in cameras and modified in lasers. We know that nothing travels faster than light. The high speed of light enables photonic crystal sensing and accounts for its sensitivity (Seitz, 1984) in practical use in products that emit light such as projectors, television screens, light displays in cars and mobile phones and cameras. Sometimes, photonic technology is not discernible, for instance in action (Nieva, Kuo, Chiang, & Syed, 2009). Photonic technology is also used in the oil and gas industries for gathering information on temperature and pressure and providing visual images of drilling operations (John, 1987). High-power lasers are used to cut and weld metal (Tao, Chen, Wang, Qiao, & Duan, 2016).

Structural health monitoring is an important area of application of photonic base strains and displacement sensors. Each structure undergoes deformation for micro pressure or displacements even though such deformations or displacement may not be noticeable to the naked eye as micro pressure and progress deformation of materials depend on the properties of the material such as Young's modulus and density.

The optical membrane structure contributes considerably to fields that use optical communications, photonic crystal sensors and quantum electronics. The study of light, to be more precise, the science of photonics includes the generation of transmission signals. Modulating intensity and, refractive index changes in overall waveguide configuration are key facts involved in photonics (Wu, Jan, & Solgaard, 2013). Sensing the mechanical deformation of membrane structures using optical media is a breakthrough technology of the era, giving sensing mechanism a brighter future. Generation, detection and manipulation of light in nano scale structures (Lu, Xu, & Hu, 2008) and features are primary factors involved in photonic-based micro sensors that measure mechanical pressure or stress detection. The nanometer scale is different from the macroscopic scale. Properties like colour and refractive index depend on geometry. SOI involves placing thin layers of silicon on top of an insulating material in order to speed up the performance of the microprocessor by reducing the capacitance of the transistors, making them operate faster.

The micro-mechanical sensor proposed here was designed using photonic crystal rods in air configuration and simulated using the MEEP tool (MEEP_Introduction). The proposed sensor was integrated with a silicon-on insulator platform and analysed for micro pressure and displacements. Pressure was applied at the centre of the photonic crystal membrane layer due to deviation of the membrane from its original position and structural deformation that could obstruct the light and to allow for wavelength and intensity modulation. The quality factor and sensitivity of the proposed model are discussed.
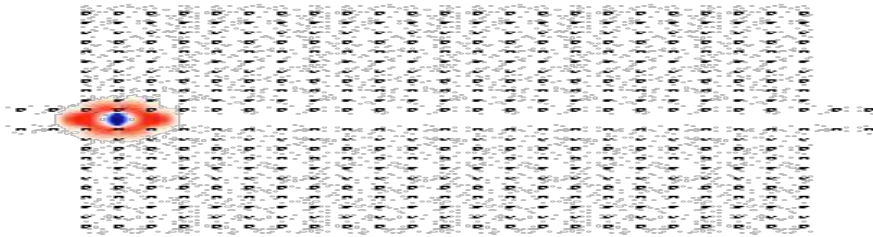
Figure 1. Structure of photonic crystals at zero displacement defect

Design and Operation Principle

The overall sensor sensitivity can be expressed as:

$$St= \varrho P/\varrho T= \varrho T/\varrho h \tag{1}$$

where, T = optical transmittance, P = applied pressure and h = height of photonic crystal.
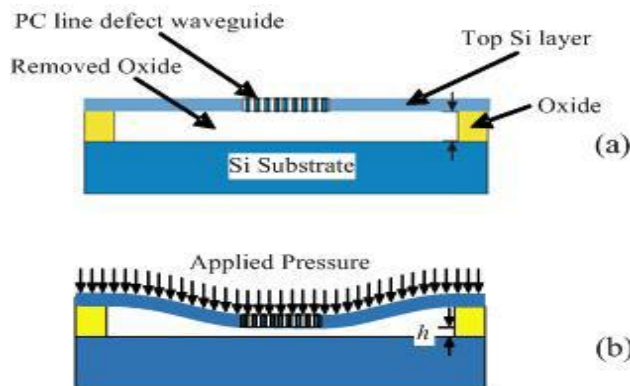
Figure 2. (a) Optomechanical pressure sensor photonic crystal in SOI platform before applying pressure; (b) Sensor structural behaviour during pressure application
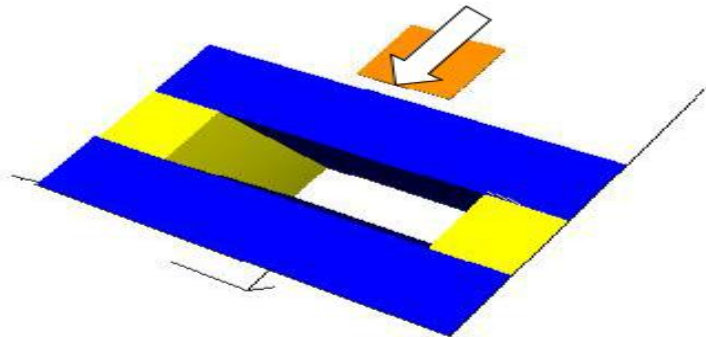
Figure 3. 3D representation of the structure

Figure (3) shows the 3D implementation of the pressure sensor when force is applied on the structure or the Si layer embodied with the photonic crystal sensor (Wu et al., 2013). The photonic crystal sensor tracks the distance between the upper Si waveguide and the lower Si substrate. When force is applied to the structure, stress increases and the sensor experiences a depression in the power absorbed (Bahaddur, 2016).
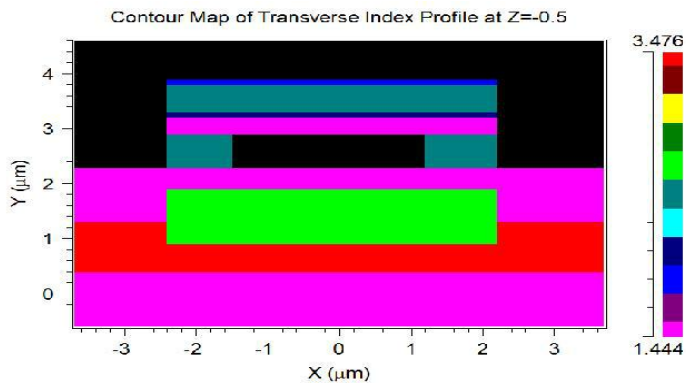
Figure 4. Contour of index profile of the structure

The refractive index profile is depicted in Figure 4. The variation in the colour shows the refractive index distribution in the structure. The stress and restraining movement of the Si slab are justified through propagation, as discussed in the next section.

FDTD SIMULATION RESULTS

The transmission spectrum is from zero micro displacement of the dielectric slab up to 6 micron displacements. In these spectrum results, there is a change in peak amplitude from one displacement to another.
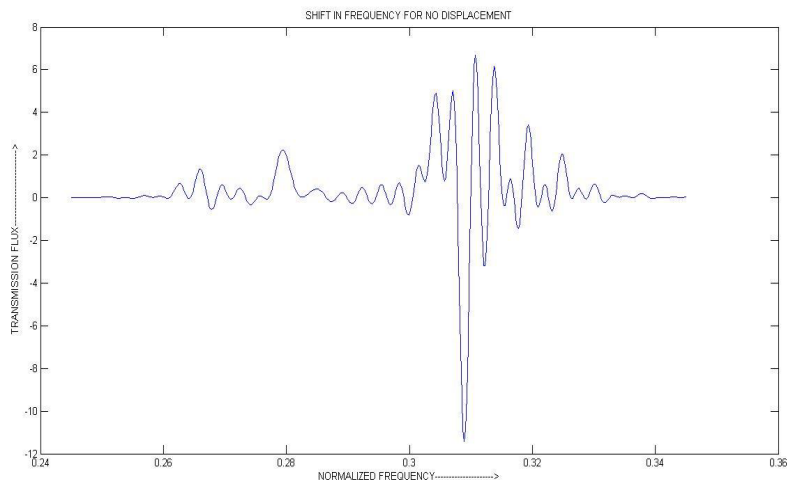
Figure 5. Spectral behaviour with 0-micron displacement
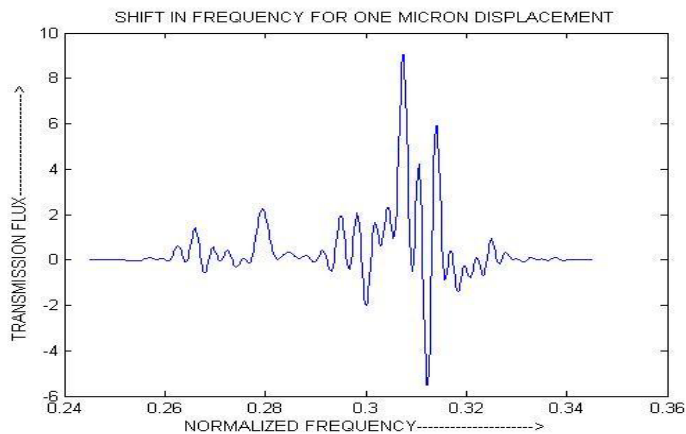
Figure 6. Spectral behaviour with 1-micron displacement
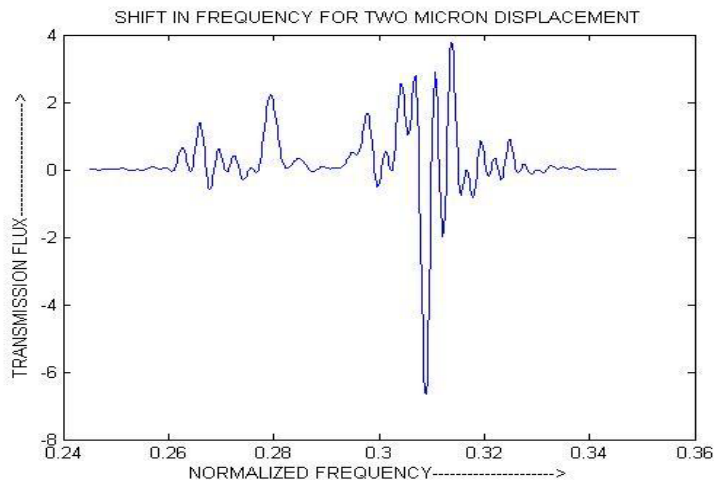
Figure 7. Spectral behaviour with 2-micron displacement

Figure 5 to figure 7 represent the transmission spectrum for 0-micron displacement to 2-micron displacement, respectively. A distinct frequency shift as well as intensity shift is obtained for every displacement, making it an unambiguous solution for various applications.

STRESS ANALYSIS

For the 0-micron displacement slab, stress was lower and a peak was observed at 1.3 μm wavelength. Displacement ranging from 1 to 6 microns produced the shortest wavelengths of 1.2-1.4 μm, as shown in the table below.

Table 1
Micro displacement and corresponding wavelength values

Displacement	Amplitude/ Wavelength	Intensity
No displacement (Zero displacement)	0.4	1.2
1 micron	0.23	1.2
2 micron	0.2	1
3 micron	0.18	1.254
4 micron	0.15	1.3
5 micron	0.1	1.36

From the above readings it was observed that for displacement, the sensor not only exhibited the wavelength shift, but also exhibited relative intensity shifts. This characteristic of the device can be implemented in barometers and other pressure/stress-orientated applications such as water level controllers.

CONCLUSION

Highly sensitive photonic crystals integrated in a silicon-on insulator (SOI) was analysed for micro displacement and stress. A novel approach in the design using an SOI platform was used for the photonic crystal-based pressure sensor, while the oxide layer was partially etched, thus providing a window for stress analysis by forming a cavity. For each structural deformation of the photonic crystal membrane in the silicon-on insulator, there was a shift in wavelength for range of micro displacements in the proposed sensor design up to a quality factor of 1495 and a high sensitivity of 1200 RIU. This type of sensor has remarkable applications in blood pressure monitoring, damage monitoring and structural health monitoring.

REFERENCES

Bahaddur, I., Srikanth, P. C., & Sharan, P. (2016, March). A photonic crystal based pressure sensor. *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT),* (pp. 3902–3905). IEEE.

Biallo, D., Orazio, A. D., Sario, M. D., Marrocco, V., & Petruzzelliet, V. (2007). Photonic crystal sensors. *Journal of the European optical society - Rapid Publications, 2*(2007). doi: 10.2971/jeos.2007.07017.

Joannopoulos, J. D., Meade, R. D., & Winn, J. N. (1995). *Photonic crystals: Modeling the flow of light.* New York, NY: Princeton University Press. Retrieved from http://ab-initio.mit.edu/book/

John, S. (1987, June). Strong localization of photons in certain disordered dielectric super lattices. *Physical Review Letters, 58*(23), 2486–2489. https://doi.org/10.1103/PhysRevLett.58.2486

Lu, Z., Xu, X., & Hu, X. (2008, January). Extracellular superoxide dismutase deficiency exacerbates pressure overload-induced left ventricular hypertrophy and dysfunction. *Hypertension, 51*(1), 19–25. doi: 10.1161/HYPERTENSIONAHA.107.098186

Nieva, P. M., Kuo, J., Chiang, S. H., & Syed, A. (August 2009). A novel moems pressure sensor: Modelling and experimental evaluation. *Sadhana (Indian Academy of Science), 34*(4), 615–623. https://doi.org/10.1007/s12046-009-0039-4

Seitz, W. R. (1984, May). Chemical sensors based on fiber optics. *Analytical Chemistry, 56*(1), 16A–34A. doi: 10.1021/ac00265a711.

Tao, S., Chen, D., Wang, J., Qiao, J., & Duan, Y. (2016). A high sensitivity pressure sensor based on two-dimensional photonic crystal. *Photonic Sensors, 6*(2), 137–142. https://doi.org/10.1007/s1332.

Wu, X., Jan, C., & Solgaard, O. (2013, August). Monolithic photonic crystal-based fiber-tip Fabry-Pérot static pressure sensor. In *International Conference on Optical MEMS and Nanophotonics (OMN), 2013* (pp. 49–50). IEEE.

Yablonovith, E. (May, 1987). Inhibited spontaneous emission in solid-state physics and electronics. *Physical Review Letters, 58*(20), 2059-2062. doi: https://doi.org/10.1103/PhysRevLett.58.2059

SCIENCE & TECHNOLOGY

Adaptive MOEMS Based Micro Pressure Sensor Using Photonic Crystal

Johnson, O. V.[1]* and Preeta Sharan[2]

[1]Research and Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India
[2]Department of E&C Engineering, Oxford College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The Micro Opto Electro Mechanical System (MOEMS) is a most promising cutting-edge technology development that uses recent trends and has huge potential for use in sensing applications. This simulation study describes the Photonic Crystal- (PhC) based micro pressure sensor, which is highly position-sensitive and free from external electromagnetic interference. Functionality of MOEMS in the photonic crystal-based micro pressure sensor is achieved through the movement of the two piston-shaped slab structures analysed in Rods in Air (RIA) and Holes in Slab (HIS) that belong to photonic crystal configurations. Displacement of the micro cavity due to applied pressure gives rise to shifts in wavelength. It is found that for each submicron displacement starting from 0 to 0.25 mm of the piston-shaped slab that is embedded, an approximate range of wavelength shift of 0.0001 for rods in air and for holes in slab configurations occurs. The simulation design shows a remarkable response in terms of intensity shift for desirable range of wavelengths 1.36 μm to 1.44 μm for RIA and 1.377 to 1.382 for HIS, thus confirming that wavelength is adaptable. The performance parameters such as Q factor and deflection range for wavelength and intensity are observed for both RIA and HIS configurations and it is found that the sensor with the HIS configuration shows better performance with Q factor of 15897 compared with the sensor with RIA configuration, which remained at a Q factor of 2482. Deformation of structure for applied pressure exhibits a linear relationship with a resonant wavelength shift. Structural variation in relation to wavelength shift exhibits a pressure sensitivity of 58.4 μm/Pa and 0.98 μm/μPa for each configuration.

Keywords: Photonic crystal, rod in air, holes in micro cavity, Perfectly Matching layer (PML), pressure, total direct strain, Finite Difference Time Domain (FDTD)

INTRODUCTION

Photonic crystal (Sheikhaleh, Abedi, & Jafari, 2016) has a periodic optical structure that controls the flow of light by affecting the motion of photons much in the same way

as solid ionic lattices affect electrons. Photonic crystal is available in one-dimensional, two-dimensional and three-dimensional configurations. Photonic crystal used as a sensor (Kovacs, Ivanov, & Mescheder, 2015) is used widely in most applications because of their advantageous physical characteristics such as being highly sensitive and having transmittance and reflectance. By using 1D, 2D, or 3D photonic crystal structure, it is possible to control the flow of light; this is not possible using conventional optical materials (Trigona, Ando, & Baglio, 2016). A combination of the optical and mechanical system to actuate the electrical structure developed is known as the Micro Opto Electro Mechanical System (MOEMS). A MOEMS micro sensor uses an input signal sensed by a micro sensor element that is transduced by a transduction unit to produce an output signal. Using the optical system along with MEMS technology (Yang, Tian, Wu, Yang, & Ji, 2013) allows for a lot of benefits such as low weight, micro size, elimination of cross talk or electromagnetic interference and immunity to variation in temperature and moisture content. MOEMS uses many advanced fabrication technologies in micromachining (Biallo, Sario, Orazio, & Marrocco, 2007), encapsulations and die bonding, among others. A more extensively developed technology for the MOEMS pressure sensor is micromachining, which mostly uses capacitive piezo resistive or resonant frequency technology. These techniques mould pressure into displacement or strain it through a membrane (Tung, Dao, & Sugiyama, 2010).

In one work, the force and strain sensing capability of the noval PhCWG- (Yablonovith, 1994) based bridge structure was studied. A pair of air holes in a silicon linear waveguide acted as a reflector. A free standing bridge beam structure was proposed comprising a photonic crystal waveguide for strain and force sensing, as the beam structure is a common mechanical structure used in MEMS-based physical sensors (Bahaddur, Srikanth, & Sharan, 2016). It was concluded that a change in defect length coupled with an elongation and or deformation in x direction would be the most sensitive parameter to resonant wavelength shift (Motamedi, 2010). A novel nano-scale optical torsion-free photonic crystal pressure sensor was analysed. It consisted of a PhC waveguide and side-coupled piston-type micro cavity. During the analysis, optical properties of resonant mode on the applied strain were systematically studied (Imada Noda, Chutinan, Mochizuki, & Tanaka, 2002). A linear relationship between applied strain and shift in wavelength with pressure sensitivity of 0.50 nm/nN was seen. Work was accomplished relevant to a tunable PhC sensor that was highly sensitive to position where the polysilicon micro cavity with silicon pillars were incorporated. MEMS flexures (Levy, Steinberg, Boag, Krylov, & Farb, 2007), using two-dimensional photonic crystal cavities where photonic crystal mounted on flexures, were used; the deformation was transferred to the photonic crystal. Defects were introduced in the photonic crystal and it was analysed with a mechanical flexure using the silicon as a material with consistent characterisation.

The photonic crystal sensor (Choudhury, 2009) has been intensively studied for change of refractive index as variation in pressure can shift the resonant wavelength, in turn inducing change in the refractive index of structure. Investigation of the optical sensor related to MEMS technology (Radhakrishnan & Chen, 2008) has proven to be useful in a wide area of applications like microfluidics and blood pressure measurement, among others. In this paper, the function of a novel micro-scale MOEMS-based micro pressure sensor (Zamora, 2011) using photonic crystal technology (Boutami et al., 2007) was demonstrated. Then design consisted of two

inverted piston-type slab structures (Subramanian, Upadhyaya, & Sharan, 2017) coupled with photonic crystal configurations. The performance parameters such as Q factor, sensitivity and deflection range for wavelength and intensity for applied pressure were scrutinised for two different configurations of the photonic crystal. Characterisation and simulation of the MOEMS device (Zouache, Hocini, Harhouz, & Mokhtari, 2016) used photonic crystal technology to observe change in the peak resonance mode of transmission spectrum due to deformation of the micro cavity in the PhC after applying pressure. Structural variations due to the applied pressure (Shakhnov & Zinchenko, 2014) for the piston-type slab structure were studied and strain due to mechanical deformation evaluated by finite element analysis. The proposed piston-type micro cavity coupled with photonic crystal configuration in an array of slabs was observed.

DESIGN APPROACH AND WORKING PRINCIPLE

The architecture of the proposed micro pressure sensor using photonic crystal is shown in Figure 1. The design approach involved the design of photonic crystal for rods in air configuration and holes in slab configuration and Gaussian pulse source and was kept at the centre of two piston-type micro cavity structures. Finite Difference Time Domain (FDTD) solves mathematical equations representing electric and magnetic field effects. This interaction of current and magnetic fields with one another and with matter and source uses the Maxwell equation and was analysed in MIT electromagnetic wave propagation (MEEP). Using the FDTD code form behind the tool MEEP provided the output of transmission spectrum and computed fluxes for the specified equations. The PML Boundary Layer or nonphysical boundary condition controlled the confinement of the light propagation within the waveguide. Two piston-shaped micro cavities made of silicon were embedded in the photonic crystal rods in air and holes in micro cavity configurations.

Table 1
Parameters of PhC configurations

Type lattice structure	Square Lattice
Lattice constant 'a'	1 µm
Radius of rods 'r'	0.17 µm (For rods in air configuration)
Radius of holes	0.17 µm (For holes in micro cavity configuration)
Refractive index of air	1.000293
Dielectric constant of silicon rods	11.74
Wavelength of light 'λ'	1550 nm
Total frequency to which flux is calculated during MEEP simulation	500
Light source used	Gaussian pulse of frequency 0.4 and width of pulse 0.3

When pressure was applied on the two ends of the piston-shaped micro cavity, both micro cavities were considered to be moving towards each other. Movement of piston type structure was made for micron range. In the present work, the micro cavity was displaced by 0 to 0.25 µm value in an increment of 0.05 µm. Pulses of Gaussian light source passed through the slab,

as shown in Figures 3(a), 3(b), 3(f) and Figures 5(a) and 5(c). The propagation of light can be seen throughout the micro cavity, due to the presence of the PML Layer, which absorbed the most reflected light waves and confined the light on the slab that had less dispersion. As the micro cavities were displaced, electromagnetic waves passing through the piston-type slab structure were transmuted, shifting the wavelength transmuting the transmission spectrum. The corresponding Maxwell equations governing the photonic crystal light propagation were:

$$\nabla . \, B = 0 \quad \nabla \times E + \frac{\partial B}{\partial t} = 0 \tag{1}$$

where, B = magnetic flux density; J is current density; E = electric field; H = magnetic field

$$\nabla . \, B = 0 \quad \rho . \, \nabla \times H - \frac{\partial D}{\partial t} = J \tag{2}$$

where, D = displacement field; ρ = density

$$S_{\lambda\text{-}n} = \frac{\lambda 2 - \lambda 1}{n2 - n1} \tag{3}$$

where, $S_{\lambda\text{-}n}$ = sensitivity ; λ2-λ1 = difference in wavelength ; n2-n1 = refractive index difference.
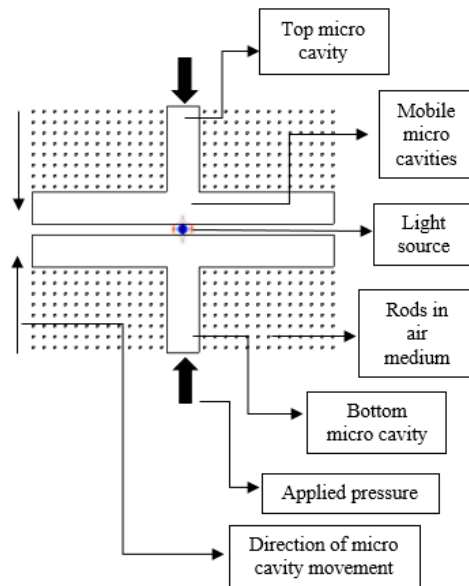
Figure 1. Photonic crystal configuration

RESULTS AND DISCUSSION

Referring to Figure 2 for the rods in air configuration, the slab was moved in the direction shown in Figure 1. Pressure was applied as shown in Figure 1 and Figure 2, and the slab was displaced to an increment of 0.05 μm from the initial position. Light was propagated from the

source centre to throughout the micro cavity, giving remarkable confinement for a particular frequency as both the micro cavity moved as shown in Figures 3(a) and 3(b). Figure 3(c) shows index distribution for the rods in air configuration. The monitors shown in Figure 2 retrieved data pertaining to wavelength and intensity variation as the slab moved from 0.05 µm to 0.25 µm. The effective index of the rods in air configuration for the proposed structure is shown Figure 3(e). Figure 3(d) depicts the 3D visualisation of the RIA configuration. The effective index of 0.991645 was identified for the RIA configuration. The index profile distribution showed the refractive index of material used during the analysis.
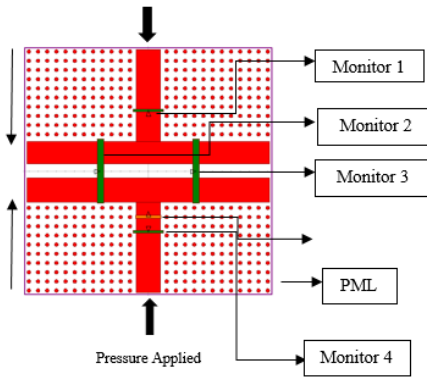
Figure 2. Photonic crystal RIA configuration with monitors

(a) (b) (c)

(d) (e) (f)
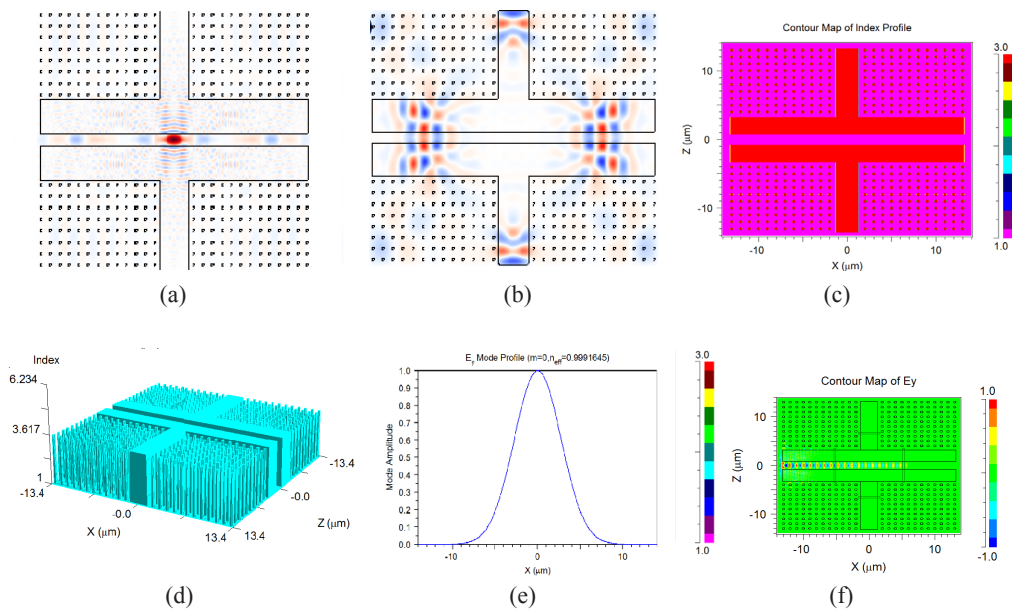
Figure 3. (a) Sample light propagation for 0.1µm displacement for RIA configuration; (b) Sample light propagation for 0.15 µm displacement of piston-shaped slab in RIA configuration; (c) Index profile distribution for RIA configuration; (d) 3D model visualisation for RIA configuration; (e) n_{eff} calculation for RIA configuration; (f) Light propagation between gaps of two piston-shaped slab

The MOEMS piston structure for holes in micro cavity configuration is shown in Figure 4. Light can be seen propagating through the proposed sensor design through the holes in the slab configuration with less dispersion, showing better confinement throughout the mobile slab. Varying flux values were obtained for each submicron displacement of the slab. Index distribution of holes in the slab configuration is indicated in Figure 5(f). n_{eff} for holes in micro cavity configuration is shown in Figure 5(e). The effective index found during the analysis was 1.888345. According to the parameters specified in Table 1, simulation was carried out allowing Gaussian pulse on the piston-shaped slab that was embedded as shown in Figures 5(a) and 5(b). The rods in air and holes in slab configurations with embedded structures used the photolithography technique to build in practical. Initially, two piston-type slabs with no displacement were used, that is no pressure was applied on the slabs. At the second level, both the inverted slabs were moved towards each other for a displacement of 0.05 µm and this was incremented at the step size of 0.05 µm. The propagation of light throughout the slab is shown in Figure 5(c).
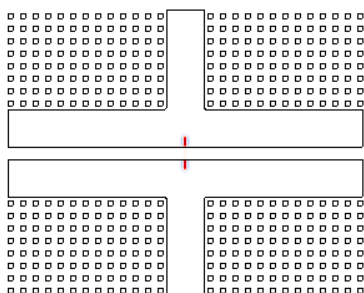
Figure 4. Proposed MOEMS-based photonic crystal micro pressure sensor in holes in micro cavity configuration
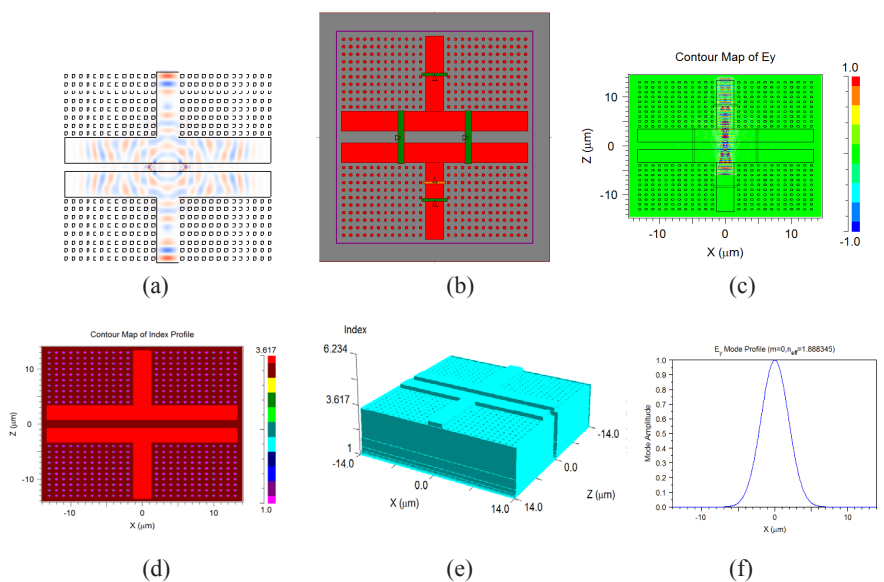
Figure 5. (a) Light propagation for 0.1µm displacement HIS configuration; (b) Photonic crystal HIS configuration with monitors; (c) Light propagation in HIS configuration; (d) Index profile distribution; (e) 3D model visualisation of HIS configuration; (f) n_{eff} calculation for HIS configuration

Figure 6(a) to Figure 6(h) show the shift in wavelength for each two displacement range 0 and 0.05 µm, 0.1 µm and 0.15 µm and 0.2 µm and 0.25 µm for rods in air and holes in micro cavity configuration. The results of the displacements and wavelength shift are shown in Table 2 and Table 3 below. The maximum Q factor approximation of 15892 was observed for the HIS configuration as indicated in Figures 7(a) and 7(b). The Q factor was analysed for changing the aspect ratio of configuration for the radius of the holes and rods, as depicted in Figure 7(b). Absorbing boundaries of structure due to pulse excitation adopted the period of time taken for light to travel from the piston-type slab structure to the surroundings and vice versa. The Finite Difference Time Domain (FDTD) method evaluated during the simulation of light travel in HIS and RIA configuration was used to analyse the decay rates to quantify the Q factor.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
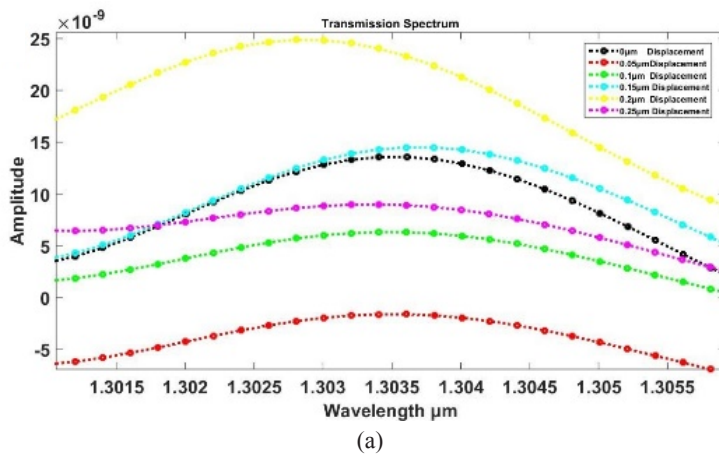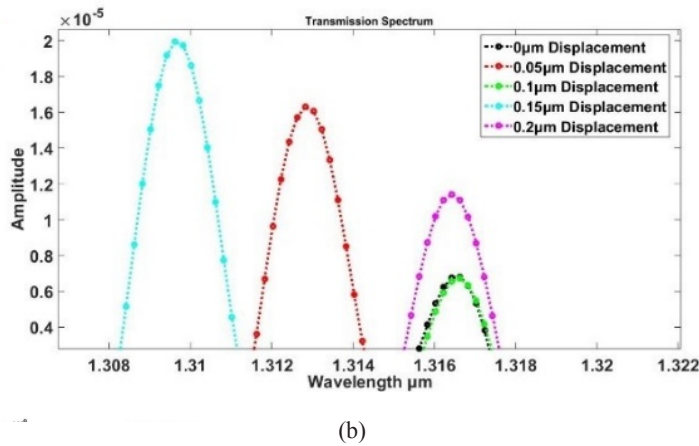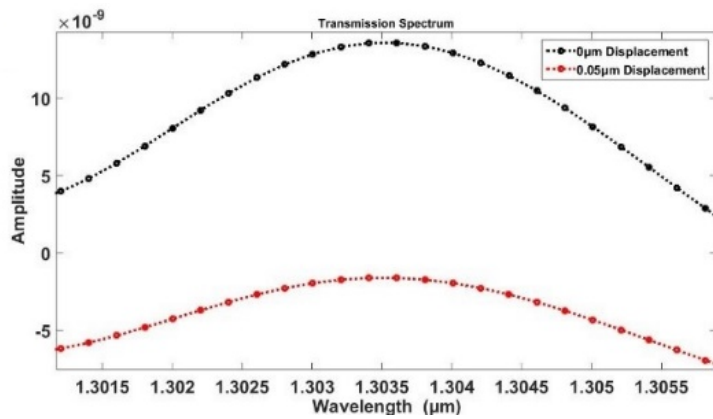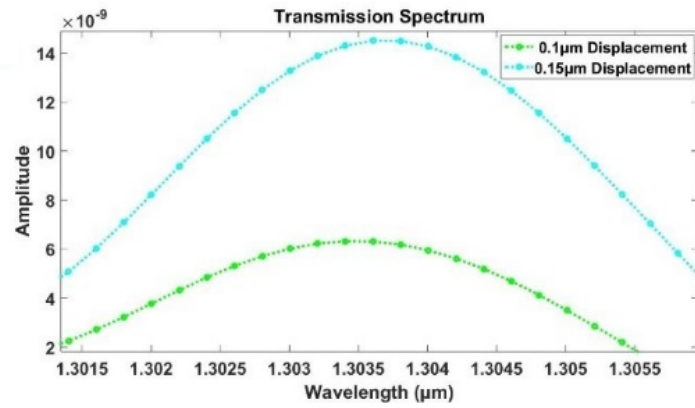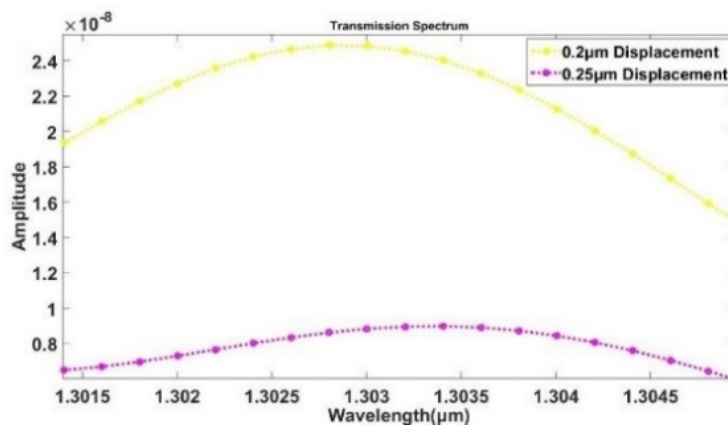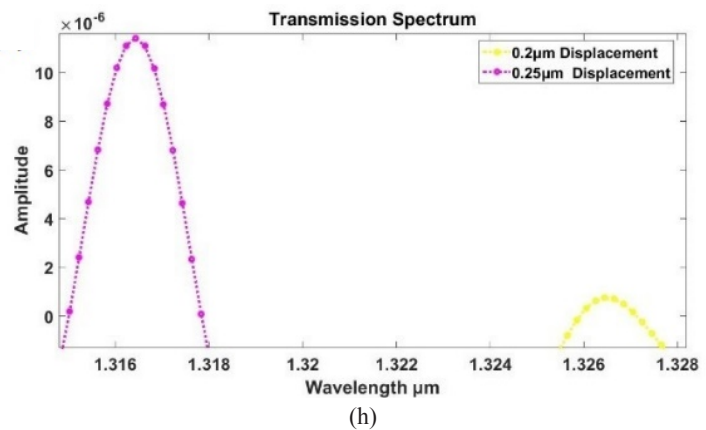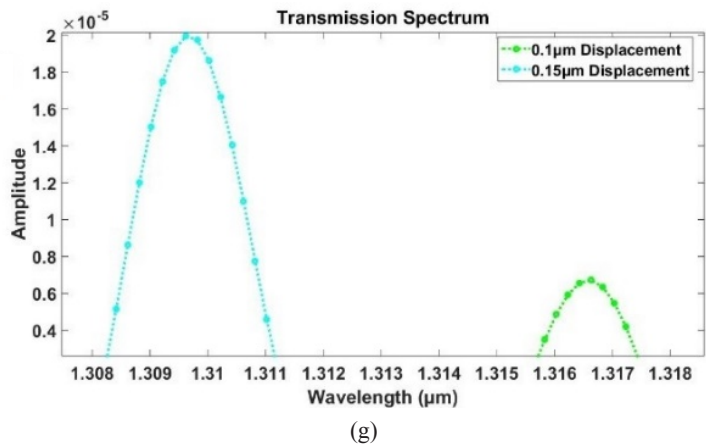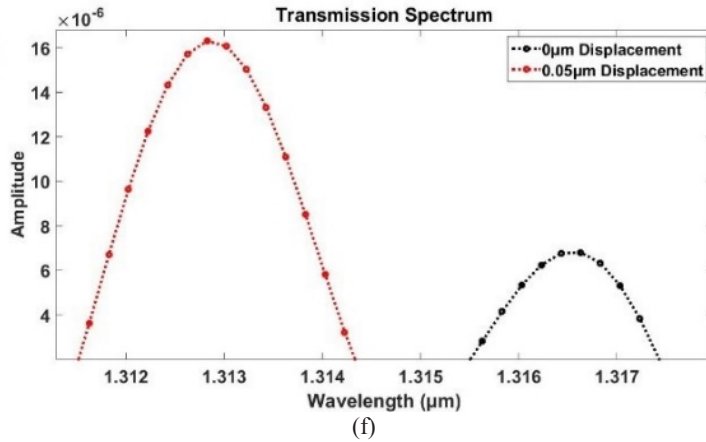
Figure 6. (a) Combined spectral behaviour of wavelength shift for RIA configuration; (b) Combined spectral behaviour of wavelength shift for HIS configuration; (c) Distinct shift in wavelength for the 0 and 0.05 µm displacement RIA configuration; (d) Distinct shift in wavelength for the 0.1 µm and 0.15 µm displacement RIA configuration; (e) 0.2 µm and 0.25 µm displacement in the RIA configuration; (f) Distinct shift in wavelength for the 0 and 0.05 µm displacement HIS configuration; (g) Distinct shift in wavelength for the 0.1 µm and 0.15 µm displacement HIS configuration; (h) Distinct shift in wavelength for the 0.2 µm and 0.25 µm displacement HIS configuration

Table 2
Result for holes in slab configuration

Displacement	Wavelength	Wavelength shift
0	1.3033	0
0.05	1.3034	0.0001
0.1	1.3036	0.0002
0.15	1.3038	0.0002
0.2	1.3037	0.0007
0.25	1.3034	0.0003

Table 3
Result for rods in air configuration

Displacement	Wavelength	Wavelength shift
0	1.3165	0
0.05	1.3129	0.0036
0.1	1.3166	0.0037
0.15	1.3167	0.0001
0.2	1.3260	0.0093
0.25	1.3164	0.0096

(a)

(b)

(c)

(d)

(e)

(f)
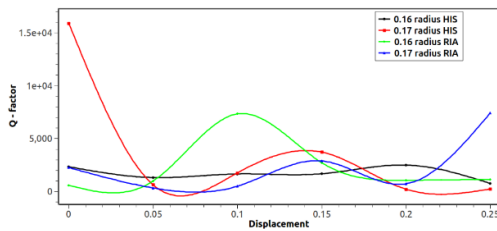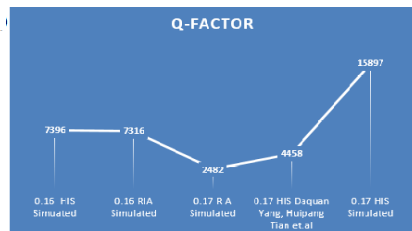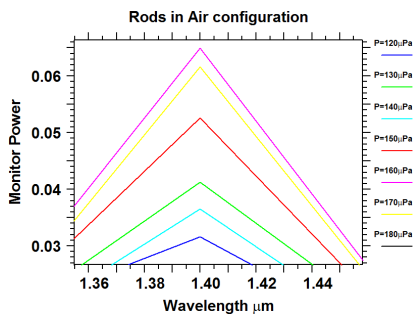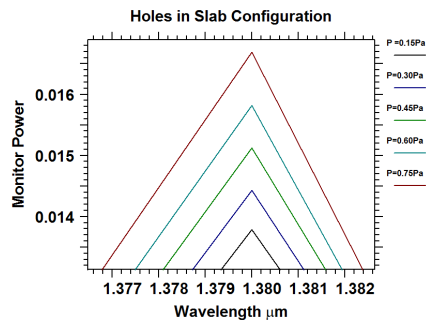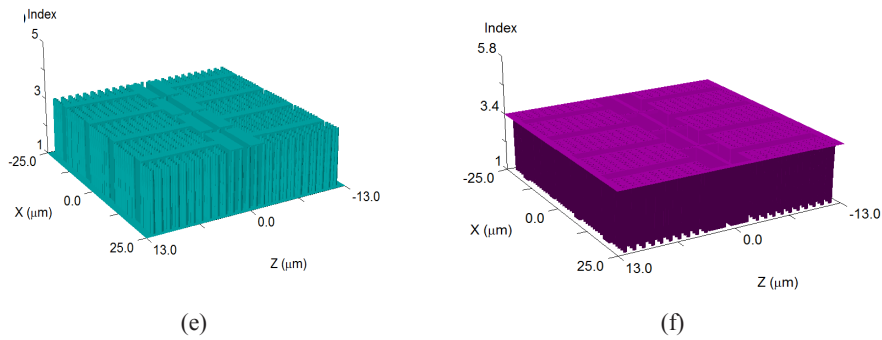
Figure 7. (a) Quality factor obtained for 0.16 RIA, 0.16 HIS, 0.17 RIA and 0.17 HIS for displacement of 0 to 0.25 μm; (b) Maximum quality factor achieved by changing the aspect ratio for HIS and RIA configurations; (c) and (d) Various applied pressure and wavelength shift for RIA and HIS configurations; (e) and (f) Array of piston-type structure slabs for bulk pressure with higher accuracy

Figure 7(c) and Figure 7(d) show wavelength peak in relation to power and applied pressure. There was a shift in amplitude for each increase in pressure from 120 micro Pascal to 180 micro Pascal in the case of the rods in air configuration. Pressure varied from 0.15 Pascal to 0.75 Pascal, while power and wavelength peak were monitored by optimising the output in the holes in slab configuration. A distinct shift was observed in intensity with respect to the applied pressure in the holes in the micro cavity configuration compared with the rods in air configuration.

To study the structural response of the sensor due to variation in mechanical parameters, the proposed micro structure was modelled in Abaqus CAE as shown in Figure 8. Due to applied pressure on both ends of the piston type of slab, structural variation was exhibited, resulting in mechanical distortion and the linear perturbation method was adopted to analyse the designed structure. A mapped quad mesh was generated using the model as shown in Figure 9(a), and suitable boundary conditions and constraints were assigned as per the movement of the sensor structure. Force was applied at both ends of the piston structure, such that back and forth movement of the structure was noticed. For each increase in the force, increase in the overall displacement of structure was identified in Figure 9(c).

$$\varepsilon = \frac{\sigma}{E} \text{ or } \mu = \frac{|\in lateral|}{|\in axial|} \tag{4}$$

$$\epsilon_x = \sigma_x / E \tag{5}$$

where, ε = strain, σ = stress, E = Young's modulus, ϵ_x = strain in x direction, σ_x = stress in x direction

$$\text{Strain} = \Delta L / L \tag{6}$$

where, ΔL = change in dimension, L= initial dimension

Young's modulus of silicon (E) = 150Gpa, Poisson's ratio (μ) = 0.14, Density of silicon (ρ) = 2328 kg/m^{-3}, Stress acting on the structure σ.

By correlating the mechanical parameters of the HIS configuration, we analysed the VonMises stress and total direct strain distribution of the micro structure for each increase in force. Figure 9(c) and Figure 9(d) shows stress and strain distribution in the proposed micro structure. Structural variation due to applied pressure with respect to wavelength shift decided the pressure sensitivity of the piston-type structure in both the configurations of the photonic
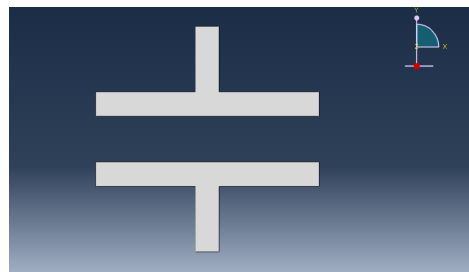
Figure 8. CAD model of proposed micro structure

(a)

(b)

(c)

(d)
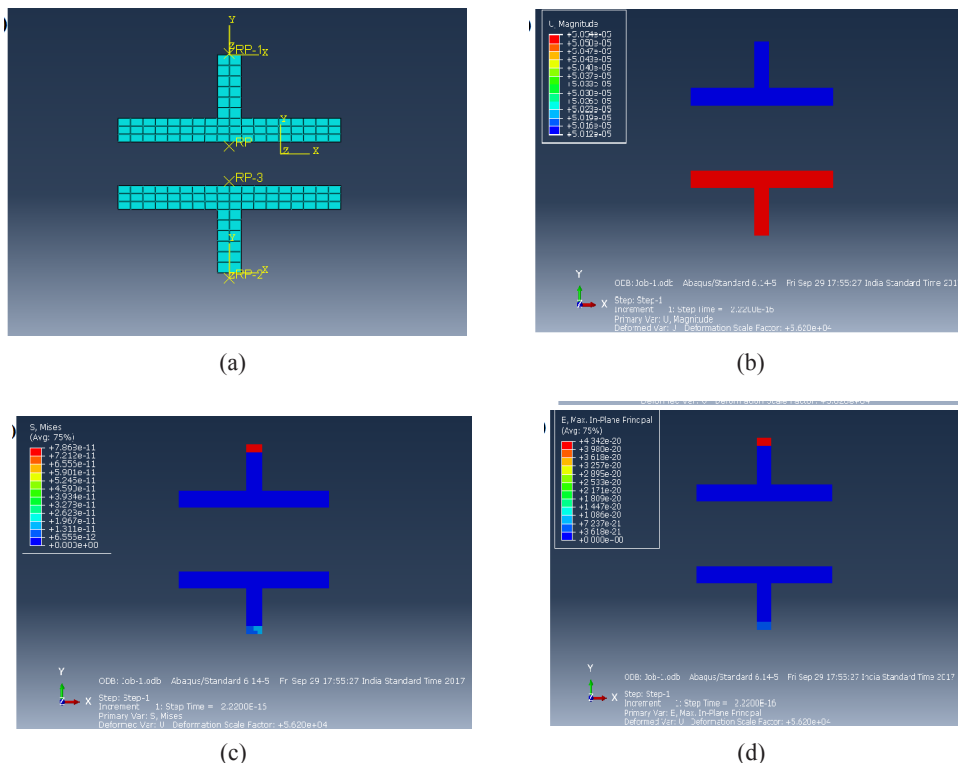
Figure 9. (a) FEA discretisation of the proposed structure; (b) Displacement behaviour of the proposed structure for applied pressure; (c) Sample stress behaviour of proposed structure; (d) Strain simulation for piston-type structure

crystal structure. Due to the stability feature of the HIS and RIA configuration for micron level movements of the piston-type slab, pressure sensitivity of 58.4 µm/Pa was recorded for the rods in air configuration, while for the holes in slab configuration, it was 0.98 µm/Pa.

The CAD model was developed by considering the same design parameters as those used in the photonic-based MOEMS structure shown in Figure 8. The designed CAD model is capable of showing the same behaviour as that in photonics. The basic fact behind the FEA analysis was to obtain the mechanical displacement and strain caused due to applying micron-level pressure. Linear relationship between wavelength shift and strain was observed during the analysis as shown in Figure 10(c). Further simulation with Abaqus FEA was performed to observe the stress behaviour for the piston-type structure. This analysis for mechanical behaviour was based purely on Hooke's law. Graphical representation shown in Figure 10(a) shows a linear relationship between stress and strain for the varying pressure units. Figure 10(b) captures the behaviour of strain against wavelength. The graph shows that there was a gradual increase of wavelength shift for strain ranging from 0.5e-19 to 1.4e-19. An abrupt increase in wavelength shift was observed after the direct strain value of 1.4e-19.
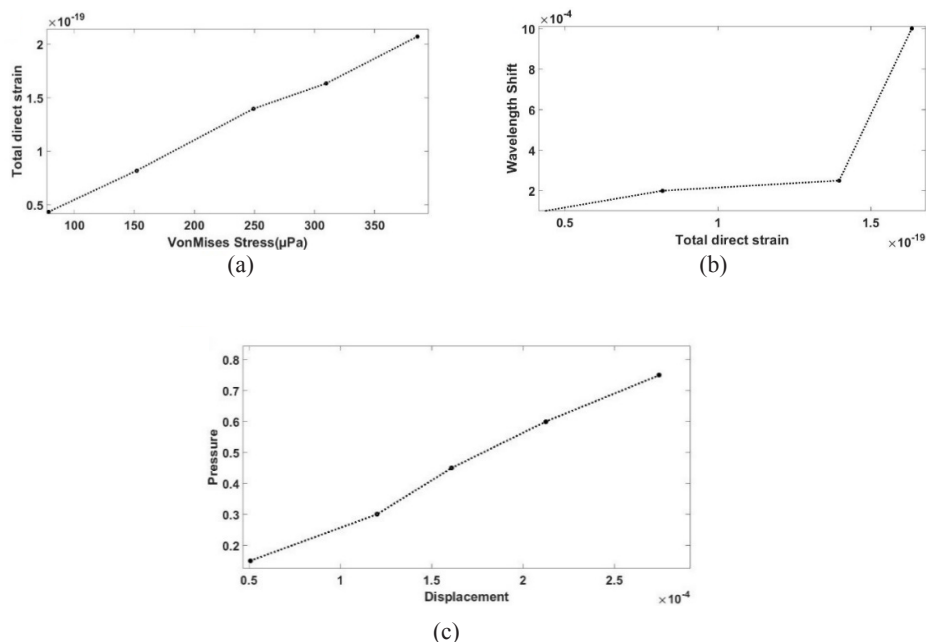
Figure 10. (a) Stress vs strain; (b) Linear relationship between wavelength shift and strain; (c) Applied pressure and displacement behaviour

CONCLUSION

This paper documents the use of an adaptive MOEMS micro pressure sensor using photonic crystal technology. The proposed sensor depends on the movement of two piston-type slabs embedded in RIA and HIS configurations. Simulation wavelength and intensity modulation were achieved for change in the mechanical parameters such as displacement and pressure. Using FDTD, the behaviour of photonic crystal-based pressure sensing technology was investigated. The simulation results showed that shifts in wavelength and intensity for varied pressure units in relation to wavelength. A maximum quality factor of 15897 was obtained for the holes in slab configuration compared with a quality factor of 2482 obtained for the rods in air configuration for a radius of 0.17 µm, while a Q factor of 7396 was obtained for the holes in slab configuration and 7316 for the rods in air configuration for a radius of 0.16 µm. A distinct intensity shift for a desirable range of wavelength from 1.3 µm to 1.38 µm was obtained in the holes in micro cavity configuration, and this was found to be remarkable compared with the shift obtained for the rods in air configuration. Linear behaviour of strain with respect to resonant wavelength shift was observed during analysis of the mechanical parameters. A sensitivity of 54.8 µm/Pa and 0.98 µm/Pa was achieved for the rods in air and holes in slab configurations, respectively.

The proposed sensor provides various benefits such as reliability and protection from electromagnetic interference. The results of this study showed that the design parameters and specifications greatly influenced sensor performance, including sensitivity, selectivity and sensing range. It was also shown that the sensing range could be adjusted and also that sensor sensitivity can be enhanced using the optimised design parameters. This type of sensor has tremendous application potential in biomedical instrumentation, especially for determining the blood pressure of neonatal babies.

ACKNOWLEDGEMENT

REFERENCES

Bahaddur, I., Srikanth, P. C., & Sharan, P. (2016). *Photonic crystal nano cavity pressure sensor*. India: OSA Publishing. https://doi.org/10.1364/PHOTONICS.2016.Tu4A.69

Biallo, D., Sario, M., Orazio, D., & Marrocco, V. (2007, May). High sensitivity photonic crystal pressure sensor. *Journal of the European Optical Society Rapid Publications, 2*(2007), 1-5. doi: 10.2971/jeos.2007.07017

Boutami, S., Bakir, B. B., Leclercq, J. L., Letartre, X., Seassal, C., Rojo-Romeo, P., ... & Viktorovitch, P. (2007). Photonic crystal-based MOEMS devices. *IEEE Journal of Selected Topics in Quantum Electronics, 13*(2), 244–252.

Choudhury, P. R. (2009). *MEMS and MOEMS Technology and Applications.* India: PHI Publication Edition.

Imada, M., Noda, S., Chutinan, A., Mochizuki, M., & Tanaka, T. (2002, May). Channel drop filter using single defect in a 2D photonic crystal micro cavity waveguide. *Journal of Light Wave Technology, 20*(5), 873–878. https://zapdf.com/channel-drop-filter-using-a-single-defect-in-a-2-d-photonic.html

Kovacs, A., Ivanov, A., & Mescheder. (2015). Tunable narrow band photonic crystal for MOEMS based scanning systems. *Euro Sensor, 120*, 811–815. https://doi.org/10.1016/j.proeng.2015.08.669

Levy, O., Steinberg, B. Z., Boag, A., Krylov, I., & Farb, G. (2007). Mechanical tuning of two dimensional photonic crystal cavity by micro electro mechanical flexures. *Sensors and Actuators A: Physical, 139*(1-2), 47–52. https://doi.org/10.1016/j.sna.2006.11.018.

Motamedi, M. E. (2010). *MOEMS*. India: New Age International (P) Ltd.

Radhakrishnan, R., & Chen, C. C. (2008). Design and modelling of nano mechanical sensor using silicon photonic crystal. *Journal of Light Wave Technology, 26*(27), 839–846. doi: 10.1109/JLT.2007.915273

Shakhnov, V. A., Zinchenko, L. A., & Kosolapov, I. A. (2014, October). Simulation of distributed MOEMS for smart environments. In *10th International Conference on Advanced Semiconductor Devices & Microsystems (ASDAM), 2014* (pp. 1–4). IEEE.

Sheikhaleh, A., Abedi, K., & Jafari, K. (2016). A proposal for an optical MEMS accelerometer relied on wavelength modulation with one dimension photonic crystal. *Journal of Light wave Technology, 34*(22), 5244–5249. doi: 10.1109/JLT.2016.2597539

Subramanian, S., Upadhyaya, A. M., & Sharan, P. (2017, August). MOEMS-based accelerometer sensor for supplementary restrain system in automobile passenger safety. *Indian Journal of Science and Technology, 10*(29), 1-6. doi: 10.17485/ijst/2017/v10i29/117327

Trigona, C., Ando, B., & Baglio, S. (2016). Fabrication and characterization of an MOEMS Gyroscope based on photonic band gap materials. *IEEE Transactions on Instrumentation and Measurement, 65*(12), 2840–2852. doi: 10.1109/TIM.2016.2608078

Tung, B. T., Dao, D. V., & Sugiyama, S. (2010, August). Investigation of strain sensitivity of photonic crystal nanocavity for mechanical sensing. In *International Conference on Optical MEMS and Nanophotonics (OPT MEMS), 2010* (pp. 183–184). IEEE.

Yablonovith, E. (1994). Photonic crystals. *Journal of Modern Optics, 41*(2), 173–194. https://doi.org/10.1080/09500349414550261

Yang, D., Tian, H., Wu, N., Yang, Y., & Ji, Y. (2013). Nanoscale torsion-free photonic crystal pressure sensor with ultra-high sensitivity based on side-coupled piston-type microcavity. *Sensors and Actuators A: Physical, 199*, 30–36. https://doi.org/10.1016/j.sna.2013.04.022

Zamora, R. (2011). Optical characterization of photonic crystals as polarizing structure for tunable MEMS devices. In M. Kusserow & H. Hilmer (Eds.). *16th International Conference on optical MEMS and Nanophotonics* (pp. 83-84). IEEE. doi: 10.1109/OMEMS.2011.6031069

Zouache, T., Hocini, A., Harhouz, A., & Mokhtari, R. (2017). Design of pressure sensor based on two-dimensional photonic crystal. *Acta Physica Polonica A, 131*(1), 68–70.

Photonic Crystal-Based Micro Interferometer Biochip (PC-IMRR) for Early Stage Detection of Melanoma

Nandhini, V. L.[1]*, K. Suresh Babu[2], Sandip Kumar Roy[3] and Ketan Pandit[4]

[1]Department of Electronics and Communications Engineering,
Government Sri Krishnarajendra Silver Jubilee Technological Institute, Bangalore, India
[2]Department of Electronics and Communications Engineering,
University Visvesvaraya College of Engineering, KR Circle, Dr. Ambedkar Veedhi, Bangalore, India
[3]Department of Electronics and Electrical Engineering, AMET University, Kanathur, 135,
East Coast Rd, Chennai, Tamil Nadu, India
[4]Department of Mechanical Engineering, PESIT South Campus, Hosur Road, Konappana Agrahara,
Electronic City Bangalore, India

ABSTRACT

The paper describes a unique approach for a label-free biosensor designed for early-stage detection of malignant and invasive melanoma. Refractive Index variation analysis plays a vital role in the photonic crystal-based sensor design. The photonic crystal-based micro interferometer biosensor has been designed in rods in air configuration. The design comprises a hexagonal ring structure between two bus waveguides forming sensing and reference arms. The early-stage diagnosis of melanoma is dependent on the comparison of normal cells against affected cells. The diagnosis is conducted using the arms of the interferometer. The sensitivity of the biosensor is determined by the phase shift/wavelength difference between the arms of the interferometer sensor. The sensitivity of the designed sensor is 4000nm/RIU. The simulations are done using the FDTD technique. The relative shifts in frequency and wavelength are due to the refractive index deviation in the sensing arm. Quality factor is obtained at 10654.3 for the wavelength 1550 nm.

Keywords: Bus waveguides, early stage diagnosis, FDTD, interferometer, melanoma, photonic crystal, quality factor, refractive index deviation

INTRODUCTION

Cancer is caused by uncontrolled cell division in the body and comprises about more than 200 different forms, depending on the origin of the uncontrolled growth in the body. Cancer is diagnosed in terms of stages depending on the rate at which cancer cells spread in the

body. The stages of cancer are suspected from the occurrence of tumours and their impact on nearby lymph nodes. Skin cancer is the most commonly occurring cancer in Western countries like the USA, Canada etc. Squamous and basal cancers are the most commonly occurring skin cancers. Melanoma is the most dangerous skin cancer, with an occurrence of 1% among the population. The stages of melanoma depend on variations in skin texture.

Photonic crystal- (PC) based optical sensors have emerged as a unique solution in early detection of cancer. Although sensors based on conventional techniques are well established, photonic crystal-based sensors have gained the attention of researchers across the world (Girijamba, Srikanth, & Sharan, 2016). PC is characterised by periodic arrangement of rods (for rods in air) or holes (for holes in slab) (Pierre, Villeneuve, & Joannopoulos, 1996). The benefit of using PC is that it modifies the size and location of the holes/rods in the lattice structure, allowing the output spectrum to be modulated to sense even minute values; this cannot be achieved using traditional optical sensor-based devices (Joannopoulos, Johnson, Winn, & Meade, 2008). PC provides a unique solution for practical applications, where the monitoring of refractive index (RI) variations are important, such as monitoring of changes in complex structures bio analytes.

Sharan and Sharma (2015a) presented that for a small RI change, a moderate shift in the frequency is observed and hence, concluded that PC can be used as a sensor. Sharan and Sharma (2015b) proposed an optical-sensor design for urine analysis for diabetic applications. The sensor design consisted of a two-dimensional PC ring resonator structure (Mallika, Bahaddur, Srikanth, & Sharan, 2015) to achieve a Q factor of 210. Mondal, Sharan and Hussain (2017) presented the modelling of the Fiber Bragg Grating (FBG) for structural health monitoring application; it was geometrically designed in the wavelength window of 1.568-1.580 μm.

Despite the vast research conducted by Sharan and Sharma (2015c) in the field of PC-based sensors, commercialisation of such sensors is not yet achieved. So, new techniques must be developed to make PC-based sensing more cost effective. Also, sensitivity needs to be increased to measure comparatively lower magnitudes of RI variations. This can be done by manipulating the lattice parameters and structure of the PC, for example, by varying the rod diameter or by changing the structure of PCRR.

In this paper, we present a PC-based Interferometric Micro Ring Resonator (PC-IMRR) structure. The simulation of the operation of the sensor was carried out using the Finite Difference Time Domain (FDTD) method (Taflove, Oskooi, & Steven, 2013). A comparison of the results of the sensitivity tests was also done.

MATERIALS AND METHOD

A unique hexagonal-shaped resonant cavity was modelled using bus. The proposed model of the resonant structure reduced the scattering losses and increased light confinement. Therefore, the model was proposed to improve various features such as sensitivity and Q factor, among others. The Plane Wave Expansion (PWE) and FDTD methods were employed for obtaining the required PBG and normalised transmission spectra for the proposed structure. This paper

covers the modelling of the designed structure and also presents a simulation of the device's use. The results were later analysed and discussed. The theoretical analysis of 2D PC was carried out using the PWE method by Pendary (1996) and the FDTD method. The PBG of periodic and non-periodic structures and propagation modes were calculated using the PWE method primarily.

$$\nabla \times \left(\frac{1}{\varepsilon(r)} \nabla \times E(r) \right) = \frac{\omega^2}{c^2} E(r) \qquad (1)$$

The PWE method was realised using Maxwell's equations where, $\varepsilon(r)$ is the dielectric function and ω is the angular frequency, $E(r)$ is the electric field of the periodic structure, and 'c' is the speed of light in free space.

The solution of Equation (1) was represented in the form of a band structure. The spatial detention of the photon in the PC was achieved through the introduction of defects. The propagation of electromagnetic modes inside the PC structures was studied using the FDTD method that was introduced by Yee (1966). The FDTD was considered the most significant solution to Maxwell's equation, given its simplicity. In this simulation, the FDTD mesh size and time step were $\Delta x = a/20$ and $\Delta t = \Delta x/16$, respectively. The 2D FDTD method was used to obtain the transmission spectrum of the proposed sensor. The performance measurement parameter for the proposed sensor was the Q factor. The Q factor measured the losses in the cavity.

$$U(t) = U(O) \exp \left[\frac{-(\omega_o t)}{Q} \right] \qquad (2)$$

The Q-factor was calculated using [11], where λ_r is the central wavelength and $\Delta\lambda$ is the channel bandwidth.

$$Q = \frac{\lambda_r}{\Delta\lambda} \qquad (3)$$

PROPOSED DESIGN

The proposed design is a PC-based Interferometric Micro Ring Resonator (PCIMRR) consisting of two bus waveguides and a hexagonal bent waveguide using the rods in air configuration. The design parameters used were: lattice size, 21x21nm; lattice constant, 200 nm; rod diameter, 900 nm; the lattice constant a, 520 nm; radius of non-defected rod, 115 nm. The proposed sensor is the design using the square lattice PC. The total amount of rods in 'X' and 'Z' directions is 22 and 19, respectively, as shown in Figure 1.
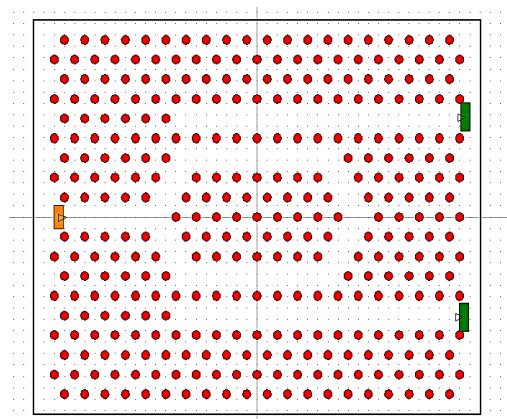
Figure 1. PC-based micro interferometer biosensor

The hexagonal lattice structures designed with silica as the base material comprised circular rods based on slab. As can be seen in Figure 1, two bus structures were created to observe the shift in resonance wavelength. The operating wavelength was chosen as 1550 nm in order to achieve higher sensitivity. Different levels of bio samples were placed in the air and the corresponding shift in resonance wavelength was observed. The effect of changes in RI on the sensor was also analysed.
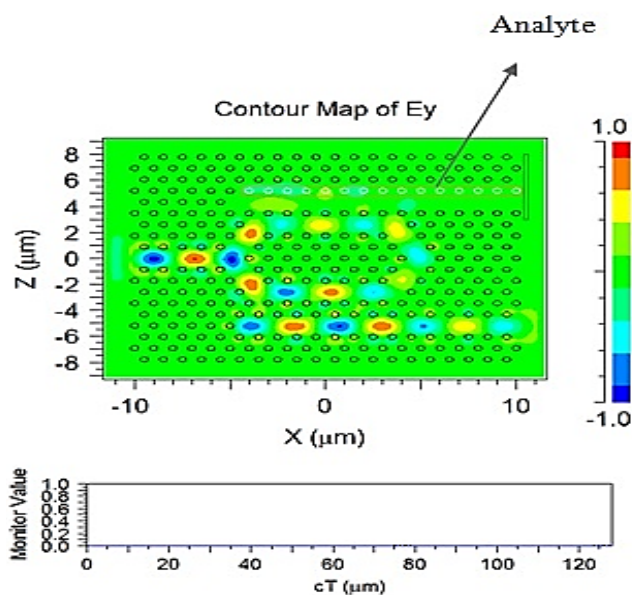
Figure 2. Excitation inside the interferometric cavity

The above contour depicts the excitation inside the interferometer. The light travelling from the micro hexagonal ring is guided to one of the two independent bus waveguides. Two excitation monitors were kept at the ends of the bus waveguides. Of the waveguides was considered the sensing waveguide, while the other was considered the reference waveguide. The sample was placed on the sensing arm so as to compare the phase variation of the sensing arm and reference arm.
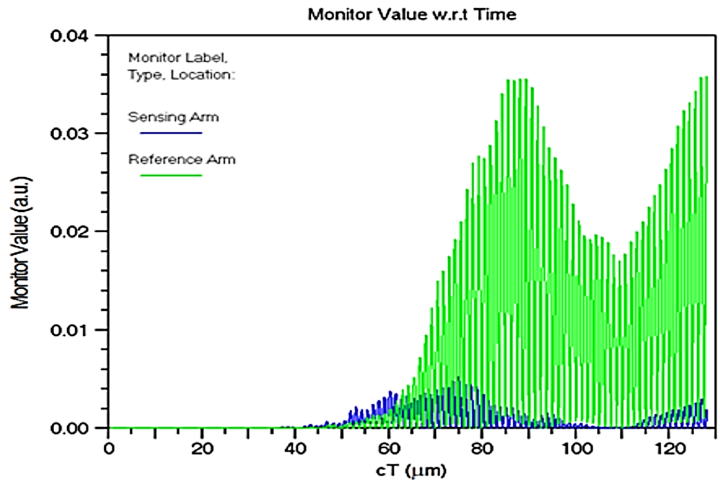
Figure 3. Light intensity comparison of the two bus waveguides

Figure 3 depicts the intensity comparison for the two bus waveguides, one with the normal melanin sample in the sensing arm. The intensity value of the reference arm was 0.037 a.u. and 0.005 a.u. Intensity variation and deviation by wavelength can be easily compared; hence, sensitivity can be calculated.
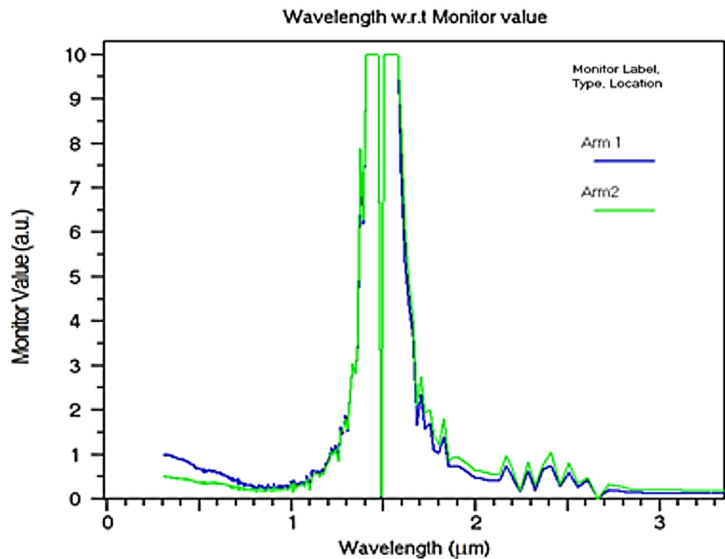
Figure 4. Light intensity (monitor value) w.r.t. wavelength for the comparison of bus waveguides

Figure 4 illustrates the typical behaviour of wavelength with respect to monitor power, where the responses from the reference and sensing arms were recorded and compared. They were found to be identical, with a negligible variation of a peak intensity value of 9.87 a.u. at wavelength 1.405 μm, where the source was 1.55 μm.

RESULTS AND DISCUSSION

Figure 5 demonstrates the light intensity with respect to the wavelength variation of the two arms, with each holding different samples. The intensity monitor placed at one arm depicted the wavelength response of the epidermis and the intensity monitor located at the other arm showed the response of the normal melanocyte. Due to high absorption of light, the melanin response lagged by 0.02 μm.
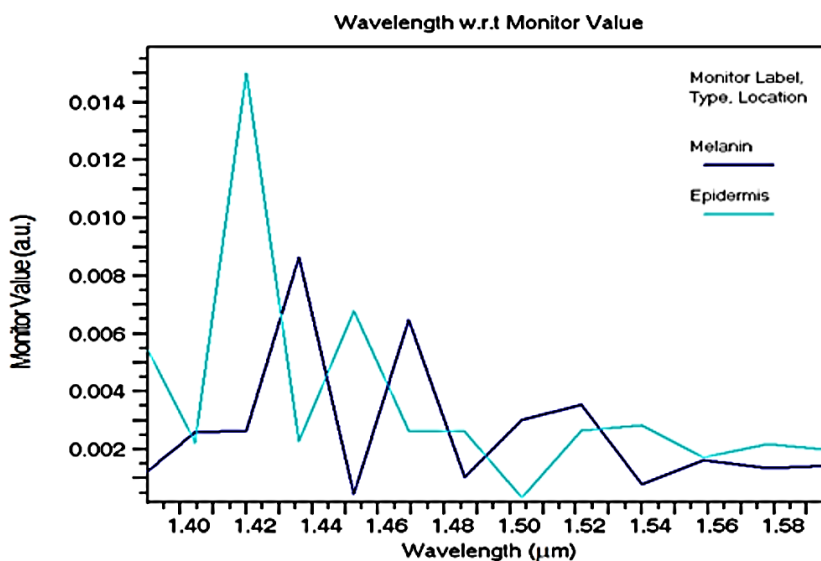
Figure 5. Light intensity (monitor value) w.r.t. wavelength for the comparison of bus
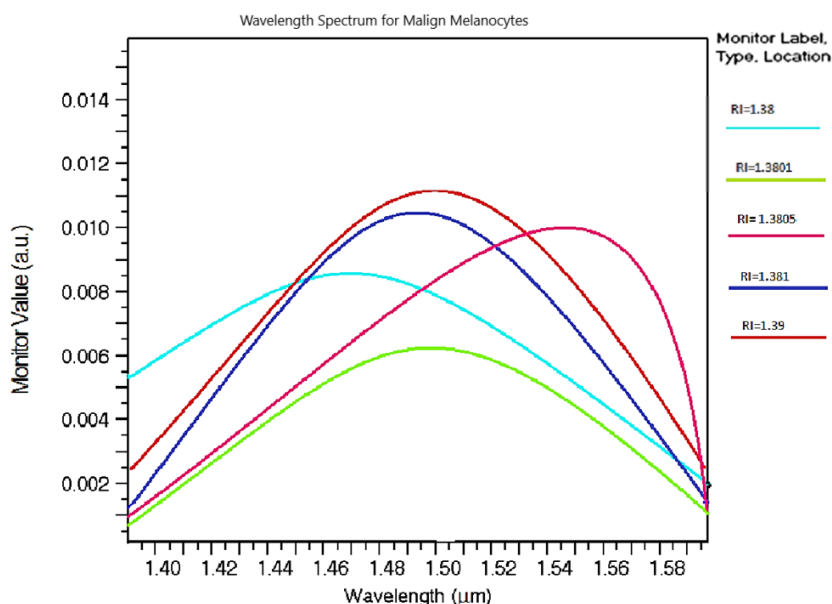
Figure 6. Wavelength response for malign melanocyte values ranging from 1.38+0.01

The figure shows the wavelength spectrum for malign melanocytes, where the peak values of malign cells occurred at the wavelength range of 1.45 to 1.55 µm.

CONCLUSION

Our work was aimed at minimising the size of devices at nano-scale for integration with integrated optical chips. The designed device can be used extensively by the layman with minimum training and if produced in bulk, will be a low-cost device. We designed and developed the sensor for skin cancer detection using bio analytes. Each element of the sample has a unique RI and our design approach involved the use of a unique RI for each element as the signature of the element. The presence of the disease can be established by matching the signature of a cancer-infected cell in the sample. The accuracy and reliability of the results were demonstrated using state-of-the-art simulation tools. The result obtained shows sensitivity of 4000 nm/RIU and a Q factor of 10654.3. The figures were much higher than the previously published results of 250 nm/RIU for Mach-Zehnder-type optical sensor implementation.

REFERENCES

Girijamba, D. L., Srikanth, P. C., & Sharan, P. (2016). Lab-on-chip based optical biosensors for the application of dental fluorosis. *Optik-International Journal for Light and Electron Optics, 127*(6), 3480–3483.

Joannopoulos, J. D., Johnson, S. G., Winn, J. N., & Meade, R. D. (2008). *Photonic crystals molding the flow of light* (2nd ed.). New Jersey: Princeton University Press.

Mallika, C. S., Bahaddur, I., Srikanth, P. C., & Sharan, P. (2015). PC ring resonator structure for temperature measurement. *Optik-International Journal for Light and Electron, 126*(20), 2252–2255.

Mondal, B., Sharan, P., & Hussain, Z. (2017). Modelling and simulation of pressure sensitivity of Bragg Grating sensor for structural health monitoring application. *International Journal of Advanced Computer Research, 1*, 73–77.

Pendary, J. B. (1996). Calculating photonic band structure. *Journal of Physics: Condensed Matter, 8*(9), 1085.

Sharan, P., & Sharma, P. (2015a). Design of PC based biosensor for detection of glucose concentration in urine. *IEEE Sensors Journal, 15*(2), 1035–1042.

Sharan, P., & Sharma, P. (2015b). PC based ring resonator sensor for detection of glucose concentration for biomedical applications. *IJETAE, 4*, 2250–2459.

Sharan, P., & Sharma, P. (2015c). An analysis and design of pc-based biochip for detection of glycosuria. *IEEE Sensors Journal, 15*(10), 5569–5575.

Taflove, A., Oskooi, A., & Steven, G. (2013). *Advances in FDTD computational electrodynamics: Photonics and nanotechnology.* United States of America, USA: Artech House Antennas and Propagation Library.

Villeneuve, P. R., Fan, S., & Joannopoulos, J. D. (1996). Microcavities in photonic crystals: Mode symmetry, tunability, and coupling efficiency. *Physical Review B, 54*(11), 7837-7842.

Cheat-Proof Communication through Cluster Head (C3H) in Mobile Ad Hoc Network

Abu Sufian[1]*, Anuradha Banerjee[2] and Paramartha Dutta[3]

[1]*Department of Computer Science, University of Gour Banga, West Bengal, India*
[2]*Department of Computer Application, Kalyani Government Engineering College, West Bengal, India*
[3]*Department of Computer and System Sciences, Visva-Bharati University, West Bengal, India*

ABSTRACT

The mobile ad hoc network (MANET) is a wireless network based on a group of mobile nodes without any centralised infrastructure. In civilian data communication, all nodes cannot be homogeneous-type and not do a specific data communication. Therefore, node co-operation and cheat-proof are essential characteristics for successfully running MANETs in civilian data communication. Denial of service and malicious behaviour of the node are the main concerns in securing successful communication in MANETs. This scheme proposed a generic solution to preventing malicious behaviour of the node by the cluster head through the single hop node clustering strategy.

Keywords: Ad hoc network, black-hole attack, cheat-proof, malicious node, MANET

INTRODUCTION

Mobile ad-hoc networks (MANETs) is an infrastructure-less, self-organising network in which a set of mobile nodes (capable of receiving and transmitting radio signals) can quickly set up a temporary network (Toh, 2002). This type of network is very useful in emergency situations such as a battle field or a rescue operation after a natural disaster, in commercial applications like vehicular ad hoc networks and in communication in conference halls, among other instances (Helen & Arivazhagan, 2014). There are many underlined protocols that are available to establish this type of network (Abolhasan, Wysocki, & Dutkiewicz, 2004). We can classify all these protocols into three broad categories: (a) proactive or table driven such as DSDV (Perkins & Bhagwat, 1994), WRP (Murthy & Aceves, 1996) and FSR (Pei, Gerla, & Chen, 2000); (b) reactive, such as AODV (Perkins & Royer, 1998), AOMDV

(Marina & Das, 2001) and DSR (Johnson, Hu, & Maltz, 2007); and (c) hybrid, where some part of the network is proactive and another is reactive, such as EMR-PL (Banerjee, Sufian, & Duta, 2018) and TORA (Park & Corson, 2001).

In most routing protocol network security and lack of co-operation among nodes are yet to be solved. It is true that MANETs is not effective in civilian data communication because of lack of co-operation and the malicious behaviour of some nodes, while in specific communication such as military data communication, it is effective. Attacks by some malicious nodes from inside the network are a main security issue in MANETs. This malicious behaviour is not an issue of military communication or any other specific communication because all nodes of such type of communication are of the same type and they work specifically in that communication. But in general data communication and networking, mobile nodes are open and the types are different such as communication made through mobile phones, laptops, palmtops, PDAs, among others. Therefore, network security is very much important.

In our scheme of cheat-proof communication through the cluster head (C3H), we used single hop-clustering strategies as a generic mechanism to increase network security through the cluster head (CH) so that MANETs can be successful in civilian data communication. We know that the clustering scheme is more scalable and we can see it in FESC (Banerjee, Dutta, & Sufian, 2018). In single hop-clustering, all the nodes are attached to CH directly and turns the entire network into different partitions called clusters. CHs are connected to each other through gateway nodes and establish a MANETs. The CHs take responsibility for network security and increase co-operation within networks. Network security is a big challenge from the start for MANETs (Deng, Li, & Agrawal, 2002). All the network security and co-operation attacks in MANETs can be classified into two broad categories: selfish attack and malicious attack. A malicious attack is more harmful than a selfish attack in this type of network. This scheme explains step by step the idea behind prevention, such as malicious attacks.

The rest of the paper is organised as follows: In the next section similar works are studied, while the third section explains the clustering strategy. Our proposed solution to the problem of malicious attacks is outlined in the fourth section, followed by a discussion on the simulation results, and in the final section, the conclusion and future scope are given.

LITERATURE REVIEW

Many node co-operation and cheat-proof schemes have been proposed. Marti, Giuli, Lai and Baker (2000) suggested a routing scheme based on DSR to detect misbehaving nodes using a 'watchdog' and providing labels using a 'pathrater'. Nodes are classified using this 'pathrater' so that misbehaving or malicious nodes can be avoided. Levente Buttyan and Jean-Pierre Hubaux proposed a virtual-currency-based scheme called 'Nuglet' (Buttyan & Hubaux, 2001) to increase node co-operation in MANETs. The researchers of this scheme used two purse models; one was the Packet Purse Model (PPM), where Nuglet is debited from the source of the packet and the other was the Packet Trade Model (PTM), where Nuglet is debited from the source or destination of the packet. This scheme also described the purpose of increasing Nuglet for a node and discussed the security of these Nuglet.

Michiardi and Molva (2002) proposed a reputation-based scheme called CORE. This scheme stimulates the selfish node to avoid selfish behaviour such as denial of service attack. Zhong, Chen and Yang (2003) proposed a simple, cheat-proof, credit-based system for mobile ad-hoc networks with selfish nodes called SPRITE. This is an incentive credit- or debit-based system without any tamper-proof hardware. Here nodes could get inceptive by showing receipt of forwarded messages from Credit Clearance Service (CCS).

Kargl et al. proposed the Advanced Detection of Selfish or Malicious Nodes in Ad Hoc Networks (Kargl, Klenk, Schlott, & Weber, 2014). This scheme explained activity-based overhearing, iterative probing and unambiguous probing to detect malicious and selfish nodes in the network. Nasser and Chen (2007) described an intrusion-detection scheme. Here, malicious nodes detected by overhearing the network then responding. This is an enhanced version of 'watchdog' and 'pathrater'.

Kang, Shakshuki and Sheltami (2010) presented a misbehaving node detection scheme at IIWAS in 2010. They used a different Intrusion Detection System (IDS) as watchdog to detect malicious nodes. The IDS, called Enhanced Adaptive ACKnowledgement (EAACK), attempted to overcome difficulties faced by the watchdog. On their part, Enrique Hernandez-Orallo et al. proposed cocoa as a collaborative-contact-based 'watchdog' (Orallo, Olmos, Cano, Calafate, & Manzoni, 2015) to effectively detect selfish nodes speedily. This scheme was said to depend on a 'watchdog', whereas CoCoWa used collaborative work based on the diffusion of local awareness of selfish nodes.

Chang et al. proposed the Cooperative Bait Detection Approach (CBDS) (Chang, Tsou, Woungang, Chao, & Lai, 2015) based on DSR. This scheme exploits both proactive and reactive defence architectures. The authors used the reverse tracing approach to defend a collaborative attack by malicious nodes. Berri et al. (2017) presented a reputation-based node cooperation model at an international conference. According to their scheme, co-operation between nodes can be increased by adding or deducting the reputation of nodes in the network. If a node denied service to a reputed node, the node with the greater reputation would 'lose'. A node could gain more reputation by serving a reputed node and less for serving a non-reputed one.

CLUSTERING SCHEME DETAILS

The main problem of MANETs is the mobility of nodes and for that reason, topology is highly dynamic. Due to this high mobility traditional protocols and the security scheme of fixed networks do not work in ad hoc networks. A clustering strategy can mimic the topology of a traditional network and reduce scalability; not only is this essential for MANETs, it can also reduce other problematic issues in MANETs. Here, we have adopted FESC (Banerjee, Dutta, & Sufian, 2018), a single hop-clustering scheme with some modifications for cheat-proof and co-operation among nodes in MANETs.

There are three types of nodes in this scheme namely, Cluster Head (CH), gateway node and ordinary member. The single hop-clustering scheme is a strategy where all the mobile nodes are attached to some elected CHs directly, making the entire network of many groups

of nodes headed by each CH. CHs are elected temporarily according to high residual energy, bandwidth and low mobility of node compared with other nodes of a cluster. A portion of an instance of our clustering scheme is shown in Figure 1.
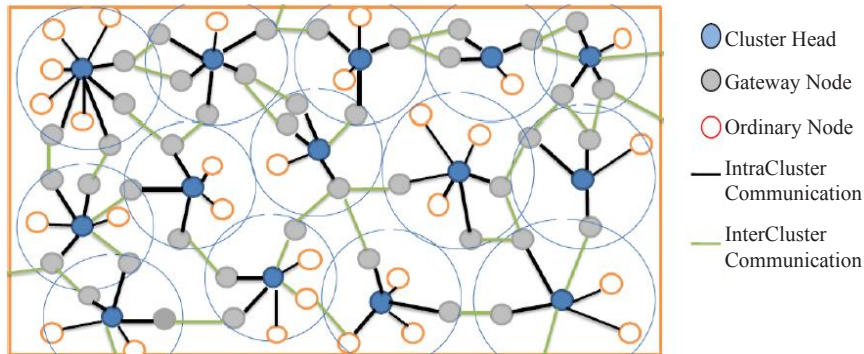
Figure 1. An instance of our clustering

Cluster Formation Strategy

The most important strategy in cluster schemes is the election of a Cluster Head (CH). The stability of the routing directly depends on stability of CHs. In this scheme, we assume CHs are fully supportive and trusted nodes; therefore, it is very important to choose a good candidate for CH. This is a single hop clustering scheme and another important strategy involving adding a node to a cluster, deleting a node from the cluster and merging two clusters to form a new cluster as done in FESC.

Electing Cluster Head (CH)

Four important factors of node were measured and combined with the help of fuzzy logic to get the final metric. This final metric was used to elect the CH and gateway node. The four important factors or metrics were: residual energy, trust value, mobility of a node and connectivity to downlink neighbours.

Residual energy of node. According to the functionality of node, at least 40% of residual energy is required to remain operational.

Let Ei be the total residual energy of node ni; Ei the expanded energy till current time; and *rem_eng(i)* the current residual energy as a fuzzy variable with values between 0 and 1. Therefore, the current residual energy will be represented by Equation 1:

$$res\_eng(i) = 1 - \frac{e_i}{E_i} \tag{1}$$

A value of '*res_eng*' below 0.4 means 'worst result' and one that is close to 1 means 'best result'.

Trust value of node. The initial stage is when a node enters the network; here, 0.5 is assigned as the default trust value, where '*trust_value*' is a fuzzy variable and the values 0 to 1 indicate the trusted level of a node. There are two more supporting variables: '*earn_trust*', which can take any natural number starting from 2, and '*loose_trust*', which is also a natural number ranging from 1 up to '*earn_trust*'. Initial values assigned to '*earn_trust*' and '*loose_trust*' are 2 and 1, respectively. When a node successfully transfers a packet its '*earn_trust*' value increases by 1. If any kind of selfish behaviour is shown its '*loose_trust*' value decreases by 1 and if any malicious behaviour is shown, it loses all its '*trust_value*', and its overall trust values decreases to 0. Trust value will also decrease after successful completion of the data transfer request by one unit. A node can ask the CH to transfer its data packet only if it has a positive trust value. The entire activity is carried out by CH. Equation 2 is used to calculate the current '*trust_value*' of a node.

$$trust\_value = 1 - \frac{loose\_trust}{earn\_trust} \tag{2}$$

More trust means more chances to become a CH as well as more credit to send a data packet.

Mobility of node. In order to stabilise the clustering scheme in MANETs, the CH of each cluster should be less mobile with its downlink neighbours compared with other nodes of the same cluster.

Let transmission power of a signal of node na be *trans_power(a)* and power of this signal when it being received at node n_i be *recv_power$_b$(a)*, while the current distance between node n_a and n_b at the time of the *i*-th HELLO message is *dist$_i$(a,b)*. Therefore, as per Frii's transmission, Equation 3 is:

$$recv\_power_b(a) = K.\ trans\_power(a)/\ dist_i^q(a,b) \tag{3}$$

where, K is constant and q is a factor with the values 2, 3 or 4, depending upon the environment. Re-writing Equation 3, we get:

$$dist_i(a,b) = \sqrt[q]{\frac{K.trans\_power(a)}{recv\_power(a)}} \tag{4}$$

Suppose *t* is the time interval between two consecutive HELLO messages and *n* is the number of HELLO messages observed. Therefore, the effective mobility of node n_a compared to its downlink neighbours is calculated using Equation 5. The average mobility called '*avg_mobility*' of a node n_a with respect to all its downlink neighbours is given in Equation 6.

$$mobility_b(a) = \frac{\sum_{i=2}^{n}(dist_i(a,b)\ -\ dist_{i-1}(a,b))}{(n\times t)} \tag{5}$$

$$avg\_mobility(a) = \frac{\sum_{k=1}^{n_d} mobility(a)}{k} \tag{6}$$

Downlink neighbours connectivity. The CH should have more downlink neighbours compared with other member nodes. Here, we assume that the current CH has the standard number of downlink neighbours. This number of downlink neighbours of CH calculated by the fuzzy membership value of that CH and initial standard membership value of this parameter is 0.5. Any node that has more downlink neighbours has a greater chance of becoming a CH.

Let the number of downlink neighbours of the current CH be '*ndnb_CH*' and of the node ni be '*ndnb_n_i*'. At first, the range of number of downlink neighbours needs to be fixed according to the current standard, as given in Equation 7.

$$\cdots\cdots dnb\_n_i = \begin{cases} ndnb\_n_{i,} & ndnb\_n_i & ndnb\_ \\ \\ ndnb\_ & , & \text{otherwise} \end{cases}$$

$$\text{Therefore, } dn = \frac{ndnb\_n_i}{2 \times ndnb\_CH} \tag{7}$$

where, '*ednb_n_i*' is the effective downlink neighbour and '*dnc*' is the downlink neighbour connectivity; clearly the range of '*dnc*' is from 0 to 1. It is also assumed that '*ndnb_CH*' will never be zero as before it becomes zero, the CH will be changed.

SEVERAL SECURITY ISSUES IN MANETS AND OUR PROPOSED SOLUTION

Besides other challenges such as dynamic topology, energy constrained and lack of bandwidth, MANETs faces two more serious challenges, which are selfish and malicious behaviour of the node. These challenges may come from some node(s) within the network. According to behaviour of the nodes, all nodes of the network can be classified into three categories: normal node, selfish node and malicious node. The normal node works in the expected way and is therefore not a concern. The selfish node works in unexpected ways, whereas the malicious node does more harm to the network. This scheme proposes a solution to security threats that come from malicious node(s) within the network. A malicious node raises the main security challenges in MANETs. Different types of malicious attacks and our proposed solution are discussed below.

Black-Hole Attack

Malicious nodes can participate in communication of other nodes by making a false route reply (RREP) packet mentioning the shortest path to the intended destination. The source node could fall into this trap of the malicious node and start sending data packets through this malicious node. Therefore, malicious nodes will be able to drop those packets or perform other more harmful work such as tampering with the data packets etc.

Figure 2 shows a portion of MANETs, where the malicious node is *m*, the source node, *s* and the destination node is *d*. The source node, *s*, wants to communicate with the destination node, *d*. Therefore, node *s*, will broadcast the route request (RREQ) packet to its neighbours,

including p, and p will also broadcast this RREQ to its neighbour nodes in the same way. In this way, the malicious node, *m*, gets the RREQ packet meant for the destination node, *d*, and node *m* provides a route reply (RREP) packet containing false information saying it has the shortest path to the destination *d*. The source node, *s*, believes the malicious node, *m*, and starts sending data packets to *m*. Now *m* can drop those packets or do more harmful work to them. Similarly, reverse-direction packets can also be captured by malicious nodes.
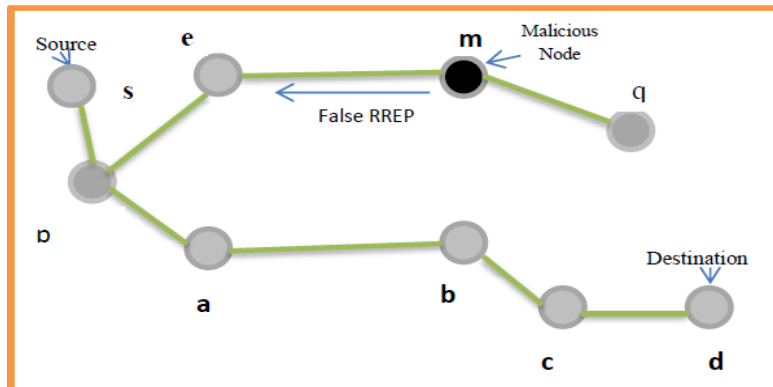
Figure 2. Black-hole attack by node *m*
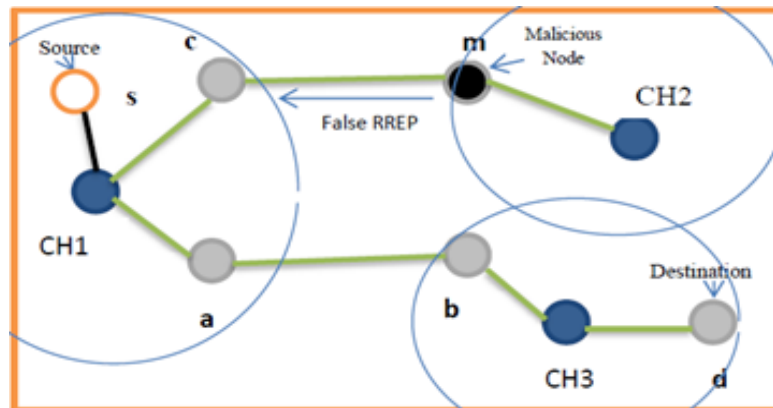
Figure 3. Solution to the black-hole attack

This black-hole attack can be prevented through the use of the CH. The CH can exert 'punishment' on malicious nodes and separate them from the network. As discussed earlier, in the clustering scheme, communication is performed through the CH, gateway node, source node and destination node. In this scheme, the black-hole attack can arise from the gateway node only and between two neighbouring CHs; at the most, two gateway nodes can participate in a communication path. The gateway node is one hop away from the CH. Therefore, activity can be monitored by the CH easily. If any gateway nodes carry out a black-hole attack, that is, when its first false-information reply enters the communication path but the node starts dropping packets, it can be easily caught red-handed from consultation with the neighbouring CH. The malicious node can then be separated and taken out of the cluster as well as the network itself.

In the example mentioned in the Figure 2, malicious node *m* can be bound by two successive CHs, CH1 and CH2, as shown in Figure 3. Between these two CHs gateway nodes, *c* and *m*, can only assist communication between these two cluster heads, CH1 and CH2. Here, node *m* is directly monitored by the cluster head, CH1; therefore, node *m* has no chance to carry out the black-hole attack as all communication is controlled only by CHs in our scheme. Therefore, here, this particular required communication can be done by CH1 and CH3.

Wormhole Attack

Here, two successive malicious nodes collude and make a wormhole between them. Whenever a route request arrives, colluding nodes hide their node information, so the source node does not perceive the presence of the two colluding nodes. The source node estimates the path length, which is less than two hops away from the actual path length. Therefore, the probability of selection of this path is very high. If this path is selected, then the source node will start sending the data packet through this path, and the two malicious nodes can drop the packets or do more harmful work such as tampering.
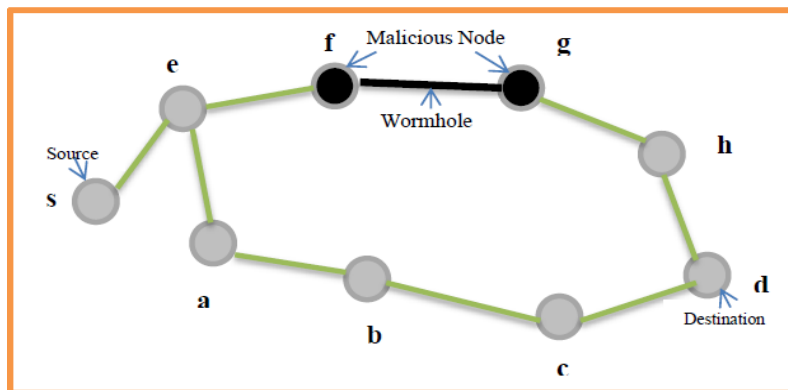
Figure 4. Wormhole attack by nodes, *f* and *g*
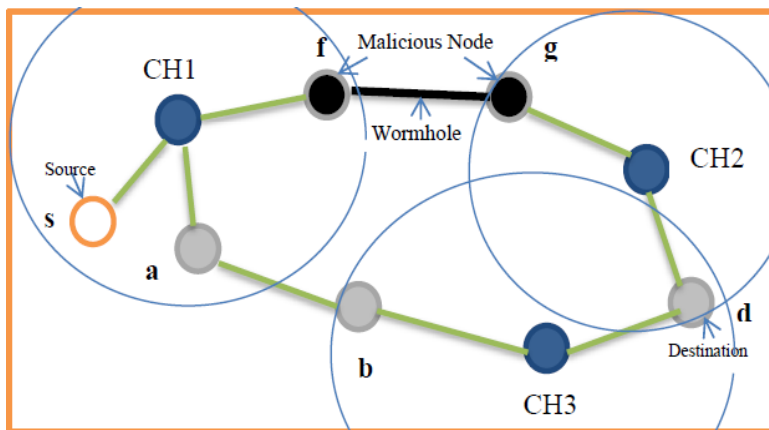
Figure 5. Solution to the wormhole attack

In Figure 4, in the portion of the network shown, source node, *s*, wants to communicate with destination node, *d*. Source nodes broadcast their RREQ packet to their neighbours, including *e*. When the RREQ is broadcast, node *f* will receive it, and it passes through the wormhole to node *g*. Node *g* sends it to node *h* without laying any marks on it. Therefore, node *h* understands that this route request packet comes from node *e* directly, but this was not the case; in this way, a virtual shortest path *s-e-h-d* is established but the actual path is *s-e-f-g-h-d*. Now, source node *s* selects this virtual path *s-e-h-d* instead of the real shortest path, *s-a-b-c-d*. When source node *s* starts sending data packets through this shortest path, the malicious node is able to capture those data packets. Reverse-direction data packets also can be captured in the same way.

As mentioned earlier, at the most, two gateway nodes could be between two successive CHs. In a wormhole attack, two successive nodes collude to make a wormhole between them; therefore, a wormhole attack is not feasible in a single hop clustering scheme because two gateway nodes are directly monitored by the CH. This is the advantage of the single hop clustering scheme. Even activity initiated by gateway nodes is easily caught by two successive CHs. In the example illustrated in Figure 4, malicious nodes *f* and *g* collude to make a wormhole between the two nodes. However, this portion of the networks seen through our scheme would look like what is seen in Figure 5. Here, node *f* is directly controlled by the cluster head, CH1, and *g* is directly controlled by CH2. The nodes, *f* and *g*, cannot collude to make a wormhole without knowing their respective cluster heads. Therefore, wormhole attacks are not feasible in our single hop clustering scheme.

Spoofing (Impersonation Attacks)

Some malicious nodes hide their addresses and use the address of another node during communication and does harm to the network. In such instances, a normal node gets falsely blamed, creating a loose trust level for such type of malicious nodes in the network. This type of attack is called spoofing.

In this scheme, every node is just a single hop away from its respective CH and sends a reply to HELLO messages from time to time to the respective CH. If any node tries to carry out spoofing attacks by hiding its address, the CH will check the address or identity through the link the node uses to connect to its CH, and if CH finds that the node is hiding its address, the CH will catch the malicious node in its cluster red-handed.

Slander Attack

This is quite similar to spoofing. Here, the malicious node attempts to reduce the overall trust of another node, but the malicious node does not hide its address. Instead, it colludes with other malicious nodes to send false information about a normal node to reduce their trust level within the networks.

A cluster member node might collude with other cluster member nodes and give false information to the CH to reduce the trust value of the target node. However, this is not possible for this scheme as explained earlier as every node is directly attached to its respective CH; therefore, increase or decrease of trust value of a node is directly done by the respective CH

without any certification of other member nodes. So, a slander attack can be resisted through the CH.

Routing Table Overflow Attack

This type of attack occurs basically on proactive routing, where the routing table is maintained by each node and even routes are not required. A malicious node sends false information by claiming it has many routes to many nodes, but those nodes do not actually exist. In this way, malicious nodes try to overflow the routing tables of other nodes so that other nodes cannot add more real route information into the routing table.

Some member nodes may unnecessarily send false information to its CH to overflow the routing table of that CH; therefore, the required routing information can be dropped by the CH. However, here the CH is the only node in a cluster that maintains the route and stores the routing table and, as already mentioned, this scheme assumes that the CH node is trust worthy. Therefore, the question of this kind of attack does not arise in our routing scheme.

Grey-Hole Attack

Here, malicious nodes flow along the same principle of behind a black-hole attack, but the malicious node drops selective packets such as data packets, and allows passage to control packets such as the RREQ packet. Therefore, another node falsely perceives the malicious node as a normal node.

A grey-hole attack is very difficult to catch because the malicious node passes the controlling packet, triggering the same resistance to the grey-hole attack as to a black-hole attack. This can come only from a gateway node. This scheme assures data packet delivery only after getting the acknowledgment message of the respective data packet from neighbouring CH. Therefore, this scheme can raise adequate resistance to grey-hole attacks.

SIMULATION

The simulation environment is given in Table 1. Performance analysis of the algorithms was done using the network simulation (NS-2) version 2.33. C3H was compared with CCS and EAACK, which are two state-of-the art approaches of detection of selfish and malicious nodes. Simulation metrics are a percentage of correct detection, malicious nodes, network throughput (percentage of data packets that can reach their respective destinations) and end-to-end delay per session.

In C3H, each CH computes the trust value of its members based on their previous activities and this trust value is considered along with residual energy and relative velocity. Therefore, if a node ceases to forward one particular message, the CH can easily investigate the chances of its complete exhaustion and breakage of links.

Table 1
Simulation parameters

Requirements	Specification
Topology area	500 m × 500 m
Traffic type	Constant bit rate(CBR)
Packet size	512 bytes
HELLO packet interval for original versions of protocols	10 ms
Node mobility	10-30 m/s
Signal frequency	2.4 GHz
Channel capacity	2 Mbps
Transmission power	300-600 mW
Receiving power	50-300 mW
Mobility model	Random waypoint
Radio range	50-100 m
Initial energy of nodes	5-10 j
Pause time	1 s
Number of nodes	20, 40, 60, 80, 100

Unlike CCS and EAACK, C3H considers residual energy of nodes and relative velocity between a CH and its members. If residual energy is very high and relativity is low, but the node does not respond to the message forwarding request of its CH, it is accused of malicious activity and its trust value reduces. If this trust value reduces below a pre-defined limit, the node is blacklisted network-wide. As seen in Figure 6, correct detection of malicious activity is higher in case of C3H.
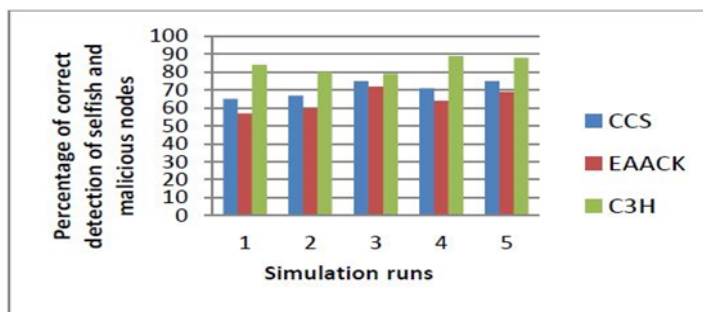
Figure 6. Percentage of correct detection of malicious activities in different simulation runs

The reason for this is that the competitors mentioned do not consider factors like energy and velocity and therefore, sometimes punish non-malicious nodes; this is not the right behaviour. In this way, we lose links to certain good nodes and also, packets generated by them are not forwarded to their respective destinations; no nodes cooperate with them. So, network throughput in C3H is much higher than CCS and EAACk, as seen in Figure 7.
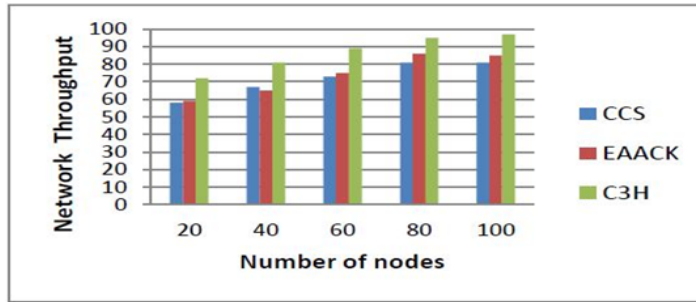
Figure 7. Network throughput vs number of nodes

Figure 8 shows end-to-end delay, which is much less in C3H due to availability of a number of good links.
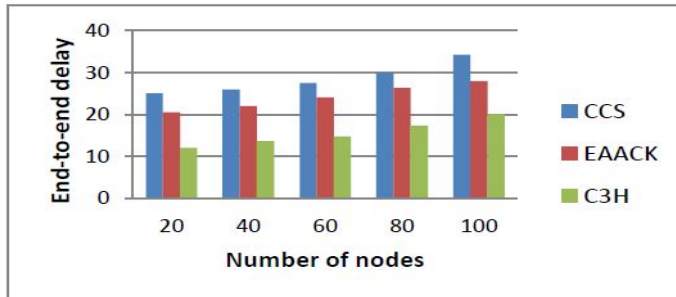
Figure 8. End-to-end delay per session vs number of nodes

CONCLUSION AND FUTURE SCOPE

In this scheme, cluster heads (CHs) are the most vital node as this scheme assumes the CH to be the most trustworthy node. Therefore, choosing the best candidate for cluster head is extremely important, and this scheme does the same using four parameters described earlier under the subsection, 'Electing Cluster Head (CH)'. Through these CHs, malicious attacks can be avoided and prevented. The CH takes packet transfer requests and gives processing priority based on trust values of that node. This clustering strategy can also be used to increase node co-operation, which is essential to successful data transmission in civilian data communication using MANETs.

REFERENCES

Abolhasan, M., Wysocki, T., & Dutkiewicz, E. (2004). A review of routing protocols for mobile ad hoc networks. *Ad Hoc Networks, 2*(1), 1–22.

Banerjee, A., Dutta, P., & Sufian, A. (2018). Fuzzy-controlled energy-efficient single hop clustering scheme with (FESC) in ad hoc networks. *International Journal of Information Technology, 10*(3), 213–327.

Banerjee, A., Sufian, A., & Duta, P. (2018). EMR-PL: Energy-efficient multipath routing based on link life prediction in ad hoc networks. *Journal of Information and Optimization Sciences, 39*(1), 285–301. doi: 10.1080/02522667.2017.1374733.

Berri, S., Varma, V., Lasaulce, S., Radjef, M. S., & Daafouz, J. (2017, May). Studying node cooperation in reputation-based packet forwarding within mobile ad hoc networks. In *International Symposium on Ubiquitous Networking* (pp. 3–13). Springer, Cham.

Buttyan, L., & Hubaux, J. P. (2001). *Nuglets: A virtual currency to stimulate cooperation in self-organized mobile ad hoc networks.* Technical Report No. DSC/2001/001, Swiss Federal Institute of Technology, Lausanne.

Chang, J. M., Tsou, P. C., Woungang, I., Chao, H. C., & Lai, C. F. (2015). Defending against collaborative attacks by malicious nodes in MANETs: A cooperative bait detection approach. *IEEE Systems Journal, 9*(1), 65-75.

Deng, H., Li, W., & Agrawal, D. P. (2002). Routing security in wireless ad hoc networks. *IEEE Communications Magazine, 40*(10), 70-75.

Helen, D., & Arivazhagan, D. (2014). Applications, advantages and challenges of ad hoc networks. *Journal of Academia and Industrial Research (JAIR), 2*(8), 453–457.

Hernandez-Orallo, E., Olmos, M. D. S., Cano, J. C., Calafate, C. T., & Manzoni, P. (2015). CoCoWa: A collaborative contact-based watchdog for detecting selfish nodes. *IEEE Transactions on Mobile Computing, 14*(6), 1162–1176. doi:10.1109/TMC.2014.2343627.

Johnson, D., Hu, Y. C., & Maltz, D. (2007). The dynamic source routing protocol (DSR) for mobile ad hoc networks for IPv4. *IETF RFC 4728.*

Kang, N., Shakshuki, E. M., & Sheltami, T. R. (2010). Detecting misbehaving nodes in MANET. In *Proceeding iiWAS '10 Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services* (pp. 216–222). ACM. doi: 10.1145/1967486.1967522

Kargl, F., Klenk, A., Schlott, S., & Weber, M. (2014). Advanced detection of selfish or malicious nodes in ad hoc networks. In *European Workshop on Security in Ad-hoc and Sensor Networks* (pp. 152-165). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-30496-8_13

Marina, M. K., & Das, S. R. (2001). On-demand multi path distance vector routing in ad hoc networks. In *Proceedings of the Ninth International Conference on Network Protocols, IEEE Computer Society* (pp. 14–23). Washington, DC, USA.

Marti, S., Giuli, T. J., Lai, K., & Baker, M. (2000). Mitigating routing misbehavior in mobile Ad hoc networks. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking* (pp. 255-265). ACM.

Michiardi, P., & Molva, R. (2002). Core: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In *Advanced Communications and Multimedia Security* (pp. 107-121). Springer, Boston, MA. doi:10.1007/978-0-387-35612-9_23

Murthy, S., & Aceves, J. J. G. L. (1996). An efficient routing protocol for wireless networks. *Mobile Networks and Applications, 1*(2), 183–197.

Nasser, N., & Chen, Y. (2007). Enhanced intrusion detection system for discovering malicious nodes in mobile ad hoc networks. In *Communications, 2007. ICC'07. IEEE International Conference on* (pp. 1154-1159). IEEE. doi: 10.1109/ICC.2007.196.

Park, V., & Corson, S. (2001). *Temporary-ordered routing algorithm (TORA)*. Internet Draft, draft-ietf-manettora-spec-04.txt.

Pei, G., Gerla, M., & Chen, T. W. (2000). Fisheye state routing: A routing scheme for ad hoc wireless networks. In *Communications, 2000. ICC 2000. 2000 IEEE International Conference on* (Vol. 1, pp. 70-74). IEEE. doi: 10.1109/ICC.2000.853066.

Perkins, C. E., & Bhagwat, P. (1994). Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. In *ACM SIGCOMM Computer Communication Review* (Vol. 24, No. 4, pp. 234-244). ACM.

Perkins, C. E., & Royer, E. M. (1998). Ad-hoc n-demand distance vector routing. *draft-ietf-manet-aodv-02.txt.*

Toh, C. K. (2002). *Ad hoc mobile wireless networks: Protocols and systems.* New Jersey: Prentice Hall PTR.

Zhong, S., Chen, J., & Yang, Y. R. (2003). Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies* (Vol. 3, pp. 1987-1997). IEEE. doi: 0-7803-7753-2/03

Wavelength Selectivity Using Adaptive Shortest Path Algorithm for Optical Network

Piruthiviraj P*, Preeta Sharan and Nagaraj Ramrao

Department of ECE, The Oxford College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The problem of routing and wavelength assignment is apparent in the dynamic all-optical network that plays an important role in the optical transport layer network. It is solved by minimising the connection blocking since the grooming adaptive shortest path algorithm shows comparably better results in terms of the calculation to find blocking probability. The shortest path algorithm used in this paper contains the present network state information, and each node creates a shortest path tree towards all the other nodes, which form node pairs by connecting each branch in the tree. The adaptive shortest path algorithm will find the shortest path throughout the network path and it chooses the best path from the available source-destination. Considering the number of nodes as 14 and comparing for different topology, it has been observed that the wavelength usage in each node varies with respect to different topology. Additionally, a comparative study of wavelength usage has been achieved for topologies like random, ring and tree.

Keywords: Adaptive shortest path algorithm, blocking probability, topology, wavelength assignment

INTRODUCTION

The problem of routing and wavelength assignment is a crucial issue that has received broad consideration among researchers (Karmi & Chlamtec, 1989). The demand for

bandwidth by network users is increasing rapidly. Optical fibre network supports WDM network as its backbone for all solutions related to network issues. Over the decades, Internet traffic has been consistently increasing and has continued to see rapid growth with respect to number of users (Zang, 2000). Internet traffic is due to the huge amount of network capacity that is utilised on bandwidth-demanding networking applications such as IP telephony, video streaming, video conferencing and sharing of the network in peer-to-peer node configuration. The wavelength routed network uses all-optical channel fibre cables for users to communicate with one; this is referred to as

the lightpath (Hurai, 1997). A lightpath utilised in a WDM optical network occupies multiple spans of fibre links. WDM networks without wavelength conversion is assigned as a common wavelength throughout the network and this property is said to be a wavelength continuity constraint.

ROUTING AND WAVELENGTH ASSIGNMENT

Routing Methods

Figure 1 shows lightpaths assigned to all combination pairs of access nodes with different wavelengths, λ1 and λ2. Using a set of connection links, the issues related to lightpaths are carried out through a routing path that is created between the nodes and a wavelength based on wavelength continuity constraint that is assigned, leading to a routing and wavelength assignment (RWA) problem (Kershenbaum, 1995). In general, a connection request in path selection is classified into three types: static routing, incremental and dynamic routing. In static traffic, the network connections (lightpaths) are pre-defined and the issue relates to lightpath connections using this structure (physical topology), resulting in reduction of network resources like wavelength usage in the network.
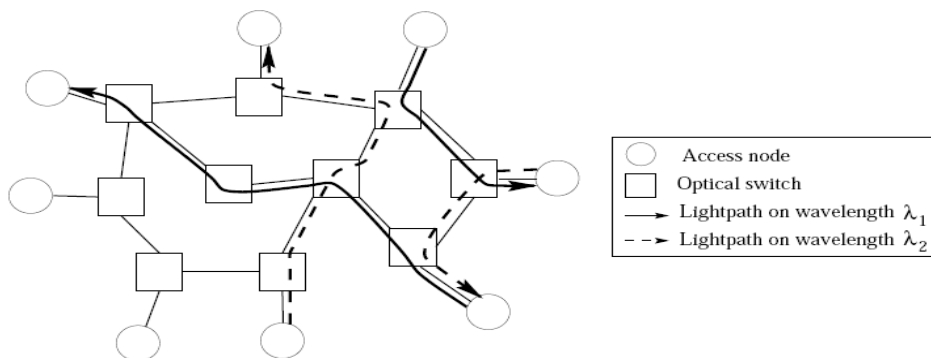
Figure 1. Optical WDM network

Multiple connections can be attempted using static traffic with the given number of wavelengths to minimise blocking ratio. Issues stemming from the RWA problem that are related to static traffic are called static light path establishment (SLE) problems. In a request involving incremental traffic, connection requests arrive in an orderly manner, and the lightpath is assigned for each connection indefinitely. In dynamic traffic, the lightpath assigned to each connection is released after the request is performed within a small interval of time. In both cases, the lightpath setup and wavelength assignment are done to accommodate a multiple number of connections and to reduce blocking that may be arise at any time. This approach is defined as the dynamic lightpath establishment (DLE) problem (Karmi, 1992).

Routing Sub-Problem

The SLE problem is formulated as two main sub-problems: i) Routing problem; and ii) Wavelength assignment problem. The routing sub-problem arises as three basic approaches: fixed routing, fixed alternate routing and adaptive routing.

Routing Types

Fixed routing is performed by creating a routing path fixed from the given pair of sources to destination connection. This simplest and most straightforward approach is performed as a fixed shortest path routing. The fixed routing uses either the standard shortest routing path algorithm such as Dijkstra's or the Bellman Ford algorithm and the shortest path route is established between the given pair of nodes in the pre-defined routing path. The fixed routing using the shortest path routing is established between Node 0 and Node 2.

Fixed alternate routing is considered to have multiple routes. Every node is assigned with a route according to the routing table, which contains a fixed order of routes to the entire destination node. The routing table is ordered with different shortest path routes as the primary route and secondary route. The primary route is linked between the source node and destination node, while the secondary route acts as a link between the pair of source-destination links, acting as an alternate route without any link to the first route.

Adaptive routing shows the better performance. According to the network state of information, the route is selected dynamically as a route between the source and the destination node. The adaptive shortest path algorithm is well suited for the WDM networks as it calculates the unused links and the used links. When there is a connection request, it calculates the shortest path from the source to the destination node.

Wavelength Assignment Heuristics

In the static wavelength assignment algorithm, based on the given set of lighpaths and routes, a wavelength is assigned to every lightpath; lightpaths do not share a similar wavelength on an optical fibre. The number of lightpaths assigned to any physical fibre link estimates the congestion based on the path selected for each connection. Each lightpath is chosen with the same wavelengths, minimising the number of wavelengths by utilising the wavelength continuity constraint feature.

Lightpaths arriving in incremental or dynamic traffic in an orderly manner should follow the heuristics methods for assigning wavelengths. With static wavelengths, the number of wavelengths assigned is fixed to avoid the dynamic problem and to minimise connection blocking. Although there are 10 wavelength assignment heuristics, the first-fit wavelength assignment technique is commonly used among all the methods as it shows the best performance in terms of blocking probability. In the first-fit scheme, wavelengths are numbered and when there is availability of wavelength, the lowest number of wavelengths is selected rather than the highest number of wavelengths, that is, the initial wavelength available is selected first. Compared with other wavelength assignment techniques, the computation cost is lower as it is not required to search the complete wavelength space for all the routes.

Shortest Path (SP) Algorithm

In Dijkstra's shortest path algorithm (Ribeiro, 2011), the shortest path route is created from every node to its respective destination node in the network. Node pairs are created by connecting the branches of the tree emerging from the individual node. Based on the SP algorithm, only the shortest path route forms a tree between the different node pairs. Topology changes that occur in a dynamic manner give rise to the adaptive shortest path (ASP) algorithm.

Generally, the ASP algorithm, according to the network state information shortest path, is calculated, after which it utilises the wavelengths for the shortest path route. As mentioned earlier, all wavelengths are not available in dynamic systems and they are represented by the traffic grooming method. In turn, they assign the wavelength to the Dijkstra's SP algorithm using the wavelength continuity method so that the connection link and the node pairs of a tree for the new route to be selected are updated.

Three criteria need to be met for the new route:

1) It should have a path that is shorter than available in the previous route.

2) A free wavelength should be available with a grooming wavelength.

3) It should not overrule the traffic constraints.

The first method is the same as the traditional SP algorithm, and a lightpath is assigned to the existing traffic, while the second method keeps a record of current information pertaining to the wavelength assigned, and when a new route is created, it makes sure that there is an available wavelength. The third method involves looking into the TR distance during the last regeneration. When a route is selected using the ASP algorithm, a wavelength is assigned for all the routes using the first-fit algorithm. In the static SP algorithm, the shortest path route is fixed in which a wavelength is assigned to the same, whereas the fixed alternate path algorithm first uses a shortest path and then changes to a second path as an alternative choice (link-disjoint method) if the other is fully occupied due to existing traffic in the network.

ADAPTIVE SHORTEST PATH ALGORITHM

The routing algorithm used here selects a route based on the network state information and it is computed after the arrival of a call request. Every router occasionally broadcasts link information to all other routers. This information is utilised to build a perspective view of the topology based on the related connection cost functions. Each router then checks the shortest path from the source to the destination. Topology is represented in the means of a graph G (V, E), in which V indicates the set of nodes and E the set of edges. Every link in the router is connected with a weight, wij, which denotes the cost of that particular link. The cost of the link altered is calculated using the number of wavelengths used.

DESIGN AND SIMULATION

The connection cost function is described as follows.

Given that $P = \{ e_1, e_2, \ldots, e_L \}$ $\forall$ e_i ε is a route path composed by L (2) and i= 1, 2, 3,…, M, M therefore indicates the maximum number of links active in network. Then, the total cost of the path, P, is calculated as the sum of all the link costs.

$$C_{T,P} = \sum_{i=1}^{L} Cej \; P \qquad (1)$$

where, $C_{T,P}$ represents as total cost of the path C_ej as the individual cost of the link e_i ε P. Therefore, cost function is calculated as cost value added when a connection is established and cost value is subtracted when the connection is ended.

$$C^n_{ij} = \{ \; C^{n-1}_{ij} + 1, \backslash C^{n-1}_{ij} \; - 1 \qquad (2)$$

The initial cost of all the links is given as:

$$C^o_{ij} = 1 \;, \; \forall \; (i,j) \; \varepsilon \; E \qquad (3)$$

The connection setup increases the cost value in each connection link and the liberation decreases the cost. It occurs up to a maximum of value $C_{ij} = \infty$, which represents the occupation of all the wavelengths. The cost of a function will increase by excluding the next request if a connection is established in a route that results in an equal load distribution for the network (Humblet, 1996).

Traffic Model

On the assumption of a network topology and traffic load, blocking performance for the SP algorithm is proposed in this paper. Blocking probability is calculated using the adaptive shortest path algorithm and first-fit algorithm (Ribeiro, 2011). Independence models described (Subramaniam, 2014) in the paper use recursive algorithms that are applicable for the dense networks and increase the blocking probability. Some models are assumed to be used for small networks that have an extensive algorithm. The model used in this paper was modified from previous works based on an algorithm that has less computational complexity, which is not suitable for a large number of wavelengths (Wason, 2011). In some models (Sun, 2003), it is proposed that the first-fit algorithm is based on overflow traffic and used as a recursive algorithm, which is suitable for a small number of wavelengths.

In this paper, performance analysis of blocking probability was calculated and analysed using the adaptive shortest path algorithm based on the first-fit wavelength assignment algorithm. The framework used was the adaptive SP algorithm with wavelength assignment algorithm. The network for different topology without a wavelength conversion analytical model is described and simulated using the above algorithm to calculate blocking probability.

Mathematical Model of Blocking Probability

Blocking probability was calculated without wavelength conversion. The wavelength continuity constraints were implemented in assigning the wavelengths. If there is insufficient free wavelength available in the network, blocking probability will be increased. Call blocking is generally calculated as:

$$P_B = \text{(Number of calls blocked)/(Total number of calls generated)} \tag{4}$$

Additionally, blocking probability is calculated using the standard Erlang formula:

$$P_B = L^W / W! \; / \sum_{i=0}^{W}(L^i/i!) \tag{5}$$

where, P_B is the blocking probability with respect to load, L, and wavelength, W. The algorithms used in this paper were the first-fit wavelength assignment algorithm and the adaptive shortest path algorithm, which were compared with Dijkstra's algorithm.

The mathematical model suggested in this paper denoted the routing path in uppercase link parameters in lower case. Prefixes and suffixes used represented links, node and routes. The network was assumed to be without wavelength conversion.

N   =   number of nodes in the optical network (14 nodes), N ε V.
L   =   length of the path or number of links or path selected, *r*
r   =   number of routers; *r* ε R
PB   =   blocking probability due to insufficient wavelengths
Nλ   =   free wavelengths available

Algorithm Description

The adaptive shortest path algorithm was computed on the base of the network state information and the wavelength was utilised effectively. The topology structure ring, random and tree are presented in Figures 5, 6 and 7, respectively. Network simulation using MATLAB was performed using a different topology structure that had 14 nodes and 50 wavelengths that were assigned for transmission of data packets.

The following procedures were followed:

1) Initial creation of nodes was carried out and the nodes were placed in the structure according to the topology.

2) Wavelength to be assigned was defined.

3) Source and destination nodes in the structure were defined.

4) The algorithm was used to compute and find the available path in the network.

5) The initial shortest path was calculated using Dijkstra's algorithm.

6) The wavelength from the table for the best path chosen was assigned.

7) Bandwidth constraints were used to ensure that data transmission followed the best path to avoid loss of data and that connection was available.

8) If data transmission is completed, the total number of data transferred in the network is calculated.

9) If data transmission is not completed, the number of data packets are calculated and free wavelength available in the network is assigned.

10) After the completion of data transmission, blocking probability in the network was estimated.
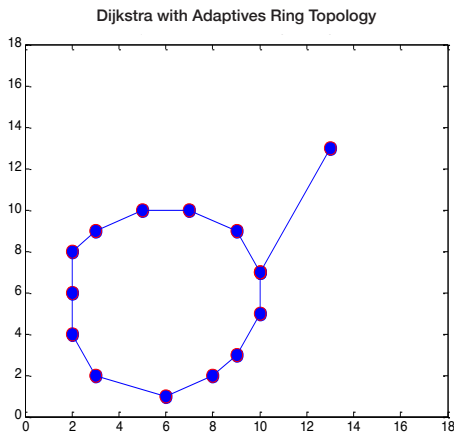
Figure 2. Ring topology with 14 nodes
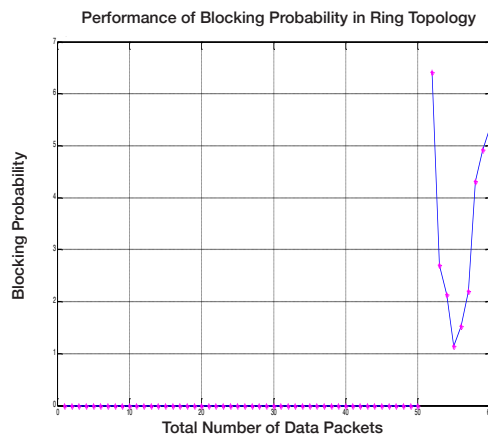
Figure 3. Blocking probability analysis
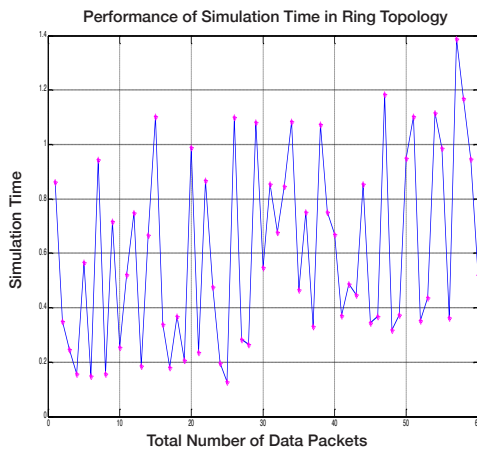
Figure 4. Simulation time
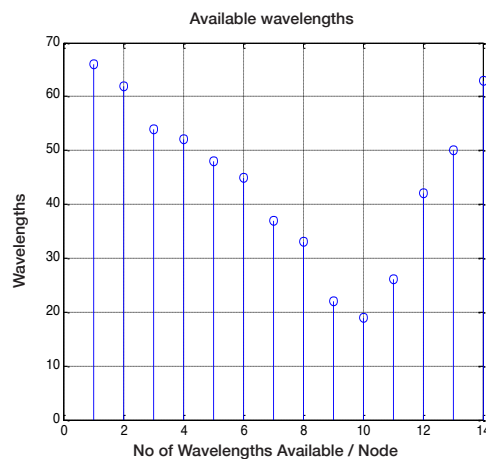
Figure 5. Wavelength utilisation

SIMULATION RESULTS AND DISCUSSION

Figure 2 represents the structure of ring topology in the network. Nodes were placed according to the ring structure as defined in the programme. The advantage of ring topology is that it enhances data transmission throughout the nodes in the network. Figure 3 shows the estimation of blocking probability, which increased after the transmission of 50 data packets due to the

non-availability of wavelengths from the wavelength table. Data packets are transmitted once a free wavelength is assigned. Figure 4 shows the calculation of the simulation time, which varied accordingly with respect to the data packets to be transmitted. The graph showed increases and decreases in time since the data transfer took place once the free wavelength was available in the network. Finally, the estimation of wavelength usage is shown in Figure 5; the number of wavelength utilisation varied with respect to the nodes as the nodes came one after another. Wavelength was therefore assigned to the individual combination of source and destination nodes.
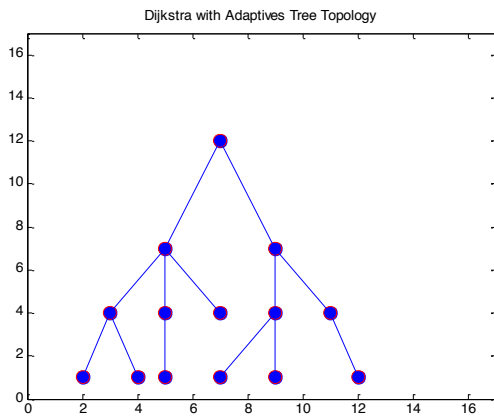
Figure 6. Tree topology with 14 nodes
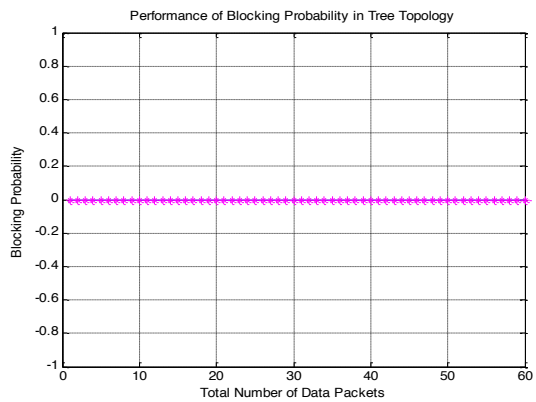
Figure 7. Blocking probability analysis
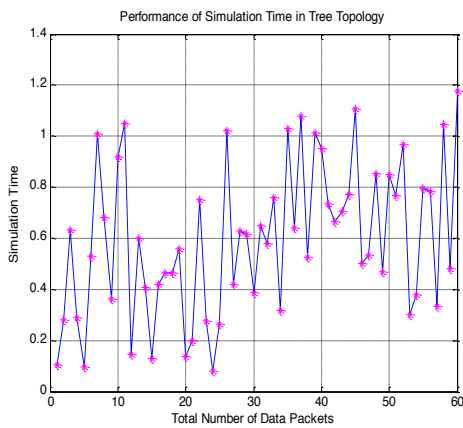
Figure 8. Simulation time
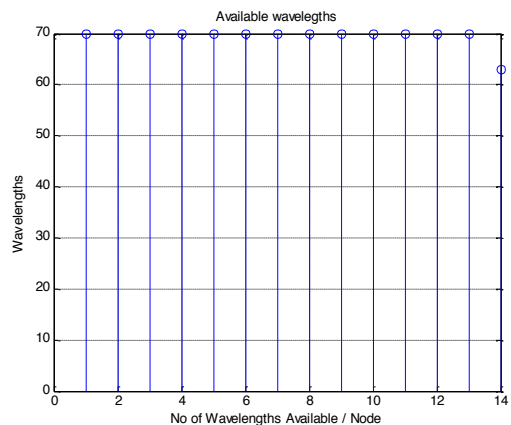
Figure 9. Wavelength utilization

Figure 6 shows the tree topology of the network. The node structure indicates that there were multiple links placed in between the different pairs of nodes. Figure 7 indicates zero blocking probability, but due to the multiple links placed between the nodes, connection cost may have been increased, as is the case with wavelength converters if different wavelengths are used in the programme. Figure 8 shows the simulation time obtained during the transmission of data in the ring topology. Simulation time was calculated as CPU time as it showed how fast the data packets were transmitted through the network. Figure 9 shows the estimation of the blocking probability node by node. The tree structure created multiple pairs of nodes in the network. It had various branches in the network structure so it assigned the required wavelength and lightpath to individual pairs of nodes. Wavelength was utilised effectively throughout the network structure.
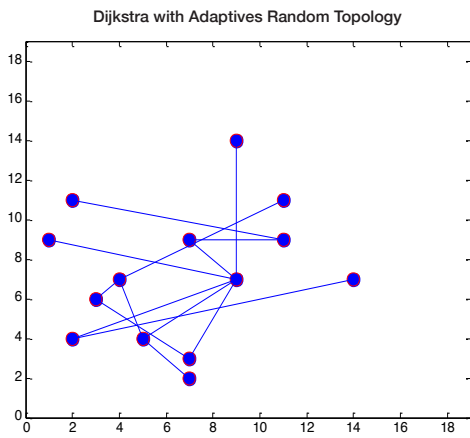
Figure 10. Random topology with 14 nodes
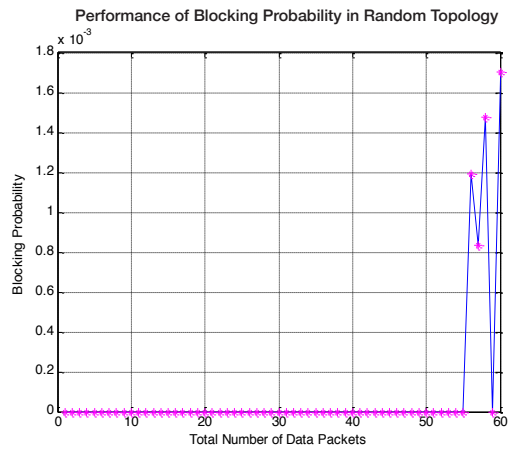
Figure 11. Blocking probability analysis
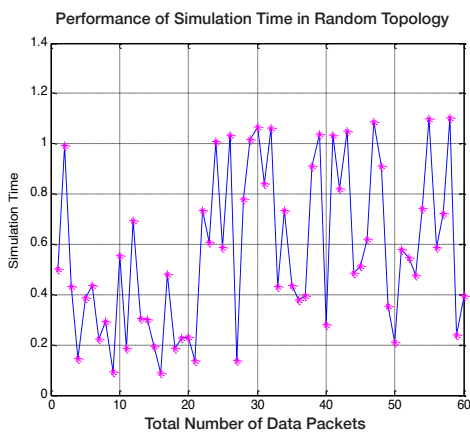
Figure 12. Simulation time
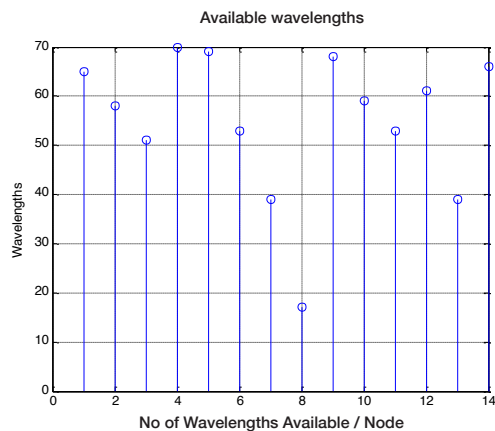
Figure 13. Wavelength utilisation

The random topology shown in Figure 10 indicates that the nodes were placed invariably in the network. This created multiples of node pairs. Figure 11 shows the simulation of the random topology as the same mesh network structure in which wavelength was initially assigned to all the nodes. After the number of data packets reached 50, blocking probability was increased due to the non-availability of wavelengths. Once free wavelength was released, the remaining data packets were transferred. The graph shows that the blocking probability was increased. Figure 12 shows the estimation of the simulation time required for the completion of the transmission of the data packets. Wavelength usage was calculated for the random topology. The study showed that the wavelength utilisation was non-uniform as there was mixing of the routing path in the network. The shortest path network was attended to as soon as there was a wavelength available for the service.

CONCLUSION

In this paper, various approaches of routing algorithms were explained to route the path and assign wavelength in the optical fibre network. The different routing types studied were the fixed, fixed alternate and adaptive algorithm. The algorithm implemented in this work showed better performance in the aspect of blocking probability. Simulation time, packets transferred and traffic load were the design parameters that justified the performance level of blocking probability. The node structure was obtained for ring, random and tree topology using 14 nodes as defined in the NSFNET structure.

The results stated that the usage of the ASP routing algorithm enhances the performance analysis of the RWA algorithm by reducing the usage of wavelengths for the corresponding level of blocking probability based on the adaptive routing when compared to other two routing algorithms. However, this is limited only with respect to the routing technique implemented in different topologies compared with the use of adaptive routing in various topologies like tree, ring and random. Tree topology had reduced wavelength utilisation compared with the other two since it had multiple links and wavelengths were assigned to the corresponding node pairs so it had no blocking probability due to the effective utilisation of wavelengths. We conclude that the use of adaptive algorithm showed better performance in choosing the best path and in lower blocking using tree topology.

ACKNOWLEDGEMENT

REFERENCES

Humblet, P. A. (1996). Models of blocking probability with and without wavelength converters, in all-optical network. *IEEE Journal on Selected Areas in Comm, 14*(5), 858 - 867.

Hurai, H. M. (1997). Performance of alternate routing method in switching networks. In *Proceedings of Conference INFOCOMM.* Kobe, Japan: IEEE.

Karmi, G. (1992). Light-path communications – An approach to high bandwidth. *IEEE transactions on communications, 40*(7), 1171–1182.

Karmi, G., & Chlamtec, A. (1989). Pure optical communication for TB comm. In *Proceedings of IEEE Infocom, III* (pp. 887–896). Washington D. C.

Kershenbaum, A. B. (1995). RWA method in singlehop optical networks with blocking. In *INFOCOM 14th Proceedings of the Conference of Computer and Comm.* Boston, USA: IEEE.

Ribeiro, L. P. (2011). Performance of wavelength assignment heuristics in a dynamic optical network with adaptive routing and traffic grooming. In *Microwave and Optoelectronics Conference (IMOC), 2011 SBMO/IEEE MTT-S International* (pp. 435-439). IEEE.

Subramaniam, S. (2014). Distributed grooming RWA for dynamic optical networks using ACO. *IEEE/OSA Journal of Optical Communications and Networking, 6*(6), 578–589.

Sun, X. (2003). Performance anlysis of first-fit wavelength assignment algorithm in optical network. In *Proceeidngs of the 7th International conference on Telecomms.* Zagreb, Croatia: IEEE.

Wason, A., & Kaler, R. S. (2011). Wavelength assignment algorithms for WDM optical networks. *Optik-International Journal for Light and Electron Optics, 122*(10), 877-880.

Zang, H. (2000, January). A review of RWA approaches for wavelength-routed optical WDM networks. *Optical Networks Magazine, 1*(1), 47–60.

Short Communication

The Effectiveness of the Quick Coherence Technique using Heart Rate Variability-Biofeedback Technology on the Recovery of Heart Coherence among University Students

Abdul Qahar Sarwari[1,2]* and Mohammad Nubli Wahab[1]

[1]*Center for Modern Languages and Human Sciences, University Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia*
[2]*Faculty of Journalism, Alberuni University of Afghanistan, Afghanistan*

ABSTRACT

Literature has established the effectiveness of self-regulatory techniques using Heart Rate Variability (HRV)-Biofeedback Technology in improving individual's heart coherence. The current study aims to evaluate the effectiveness of Quick Coherence Technique, a self-regulatory technique, through the application of HRV-biofeedback technology on the level of heart coherence among university students. A total of 20 students of a technical public university participated in the study. The Quick Coherence Technique (QCT) and the emWave device and software of the HeartMath Institute were used to collect the data. According to the HRV power spectrum, the HRV data is divided under very low frequency (VLF), low frequency (LF), and high frequency (HF) frequency ranges. Based on the results, the use of HRV-biofeedback technology and the QCT had helped to increase the levels of HRV scores and heart coherence of the participants.

Keywords: Biofeedback, coherent heart, heart coherence, Heart Rate Variability, Quick Coherence Technique

INTRODUCTION

The coherent heart may help individuals to have more successful personal and social lives in their societies. According to the HeartMath. org (2016), coherence refers to the "clarity of thought, speech and emotional composure". The heart rate variability (HRV) and the HRV-biofeedback technology which were developed by the HeartMath Institute and

are available in the portable devices and software (Ross, 2011), help researchers to assess the levels of HRV and heart coherence of individuals. The HRV has quite a long history, but the recent developments through the durable efforts of scholars and researchers had more effects on HRV-related issues and studies. According to McCraty and Shaffer (2015), there are some evidences from the assessment of multifaceted rhythms of heart known as heart rate variability (HRV) from 1960s and 1970s, it has gained serious attention recently.

The heart rate variability which indicates the time interval changes between the adjoining heartbeats is an embryonic property of mutually-dependent regulatory mechanisms that works on diverse dimensions of time to become accustomed to psychological and environmental challenges (McCraty & Shaffer, 2015; Porges, 2007; Task Force, 1996). According to them, a good range of HRV frequency is associated with the well-performance, intrinsic and natural self-regulatory capacity, and adaptability and flexibility of an individual. They also argued that the heart coherence is measurable through HRV and HRV analyses. The heart coherence is measured through the identification of the higher peak of 0.04 HZ to 0.26 HZ frequency range under the HRV power spectrum. Moreover, coherence involves association, correlations, steadiness and efficient energy use. For instance, the individuals' speeches or opinions can be considered coherent if the ideas and words fitted together and will be considered incoherent if the words and ideas are unconnected and meaningless. The physic and physiology fields also use coherence term to describe the situation of harmonisation between different erratic mechanisms (McCraty & Shaffer, 2015; Sutarto, Nubli & Zin, 2012; Porges, 2007).

Scholars (e.g. McCraty & Shaffer, 2015; Porges, 2007; Ross, 2011; Jacob, 2010) have focused on the effectiveness of the use of self-regulatory techniques on the improvement of the human physical and psychological well-being. As pointed out, the techniques of self-regulation have shown positive impacts on the production of physical and psychological well-being among members of different groups. The HRV and self-regulatory techniques are effective in improving human performance and decrease of anxiety among individuals, especially among university students in a multicultural university environment (Lagos et al., 2008; Thurber, 2006). Moreover, according to McCraty and Zayas (2014), self-regulation includes the skills to manage impulses and control one's own behaviours through awareness of one's emotions and thoughts at the same time. McCraty and Shaffer (2015) on a study on new perspectives of physiological mechanisms and assessment of self-regulatory capacity argued that, a favourable level of HRV reflects well social performance and healthy function, a natural self-regulatory capability, and flexibility. At the same time, higher HRV frequency range of above from 0.04 of LF and below the 0.26 score of HF, which is favourable frequency, indicates a good psychophysiological situation which is friendly and well-matched with social connections and public communication (Bernston, Cacioppo, & Quigley, 1997; Quintana et al., 2012; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

However, the scholars conducted their studies in the Western parts of the world. There are many differences between the Western and Eastern environments and norms of human performance (Abu Bakar & Mustaffa, 2013; Gudykunst, 2001). At the same time, there are no evidences in the literature to show that whether the cited arguments are helpful or not among university students in an Asian university environment. Thus, this study used the Quick

Coherence Technique through the application of HRV-biofeedback technology to evaluate its impacts on the levels of heart coherence of university students at a Malaysian university campus.

LITERATURE REVIEW

The heart coherence (the balanced level of HRV) l decreases psychological distresses, problematic behaviours and concerns, and improves classroom behaviours, academic achievements and social performance among students (McCraty & Shaffer, 2015; Ross, 2011; Childre & Rozman, 2003). Moreover, heart coherence, psychological well-being, flexibility and well-performance skills are associated with the desirable HRV scores of individuals (McCraty & Shaffer, 2015; McCraty & Childe, 2010). The results of a study by McCraty and Shaffer (2015) confirm that a good rate of HRV reflects proper social performance among individuals. Based on results of a study on the role of vagal function in the risk for cardiovascular disease and mortality, Musselman, Evans and Nemeroff (1998) argued that, to guide well-organised portion of cognitive and awareness resources, HRV is helpful for competent performance in a demanding environment where a delayed reply and behavioural embarrassment are the key. However, low HRV increases the risk of stress disclosure (Thayer, 2007).

Furthermore, the heart actions and reactions of individuals are associated with their autonomic nervous system. The autonomic nervous system (ANS) which affects the human heart rhythmic actions has two parts: sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The levels of SNS were associated with the *fight-and-flight* situation and cognitive and psychological stress, and the high levels of activities of PNS are associated with too much relaxation and laziness (Jacob, 2010; McCraty & Childre, 2010; Bernston, Cacioppo, & Quigley, 1991). The HRV frequency bands are divided under three bands which are: very low frequency (VLF), low frequency (LF), and high frequency (HF). The VLF frequency and LF frequency of lower than 0.4 Hz are associated with the SNS, and the HF frequency of above 0.15 is associated with the PNS. The desirable and good HRV frequency range which is associated with coherent heart, good performance, and psychological flexibility is from 0.4 of LF to 0.26 of HF (Tanis, 2015; McCraty & Shaffer, 2015; Jacob, 2010; McCraty & Childre, 2010; Bernston et al., 1991).

Thayer (2007) argued that HRV is associated with executive and nonexecutive function tasks, and physical detraining also affects HRV. According to Ross (2011), the HRV-biofeedback technology tools help people to control their HRV and enter in the state of coherence, to have coherent heart and stable HRV. The heart rate variability (HRV) biofeedback as the promising mechanism for the rising HRV has become more popular in the HRV setting of studies (Rene, 2008; Pignotti & Steinberg, 2001). Rene (2008) also argued that the findings of a study on the effectiveness of a handy biofeedback tool regarding mental problems have confirmed that the HRV-biofeedback is a safe and non-invasive device.

Researchers and scholars (e.g. McCraty & Shaffer, 2015; Reyes, 2014; Ross, 2011; Porges, 1997) have used the HRV-biofeedback technologies, including emWave device and software, in their studies on different fields, such as health, psychology, education, sports, and military. The results from their studies confirmed the effectiveness of the use of emWave, the QCT and other biofeedback technologies and techniques on human performance and enable individuals to assess the actions and reactions of their hearts. However, to see whether the

above-mentioned techniques and technologies are useful in an Asian collegiate environment, this study assess the effectiveness of the quick coherence technique using heart rate variability-biofeedback technology on the recovery of heart coherence among university students at a Malaysian public university.

STUDY DESIGN

The Quick Coherence Technique (QCT) of the HeartMath.org (2016) was used to assess its relationship with the probable developments in the levels of heart coherence of the participants. The QCT includes three stages: heart focus, heart breathing, and heart feeling (HeartMath.org, 2016; Ross, 2011; Childre, Martin, & Childoe, 2000). During the data collection process, the heart focus stage asks the participants to focus on the area of their hearts. The heart breathing stage asks the participants to breathe slowly and deeply and imagine the process of their breathing. Finally, the heart feeling stage asks the participants to try to remember a good moment of their lives and try to re-experience that moment (HeartMath.org, 2016; Ross, 2011; Childre, Martin, & Childoe, 2000).

PARTICIPANTS

The participants of this study were 20 undergraduate students of a Malaysian technical public university, namely University Malaysia Pahang. The participants were between 18 and 22 of age and they voluntarily participated in the study.

INSTRUMENTS

The emWave tool and the emWave PC biofeedback (1.0) software which is developed by the HearthMath Institute for the heart rhythm variations were used. According to Reyes (2014), the emWave is a portable device that helps individuals to monitor HRV, and to rehearse biofeedback techniques. The HeartMaath Institute through conducting of various studies has provided evidences for the effective role of hearts of individuals on their wellbeing and balanced lives. The emWave tool and other biofeedback technologies of HeartMath Institute are valid based on various academic studies on biofeedback, stress and emotions for around 17 years (HeartMath.org, 2016). During the data collection sessions, the emWave PC tool, which is a portable device, was connected into a computer to access its energy from the computer. Based on the HRV data collection procedure, the device's sensor will be placed on the ears. Figure 1 shows the emWave PC tool and its ear sensor. The pictures were taken from the HeartMath.org (2016).

Figure 1. The emWave PC tool and its ear sensor

DATA COLLECTION PROCEDURE

Before the data collection sessions, the letter of consent which included information about the application of HRV-technology and techniques, data collection procedure, time and sessions was given to the participants . The data were collected individually from the participants, and they were allowed to look on the screen in the end of each data collection session. The data collection had four sessions, which were the simple baseline session without intervention of any technique, and the three stages of the QCT, which are: heart focus, heart breathing, and heart feeling, and the exact time for each session was two minutes.

DATA ANALYSIS

The data were analysed based on the HRV power spectrum. According to the HRV power spectrum, generally there are three types of HRV frequencies which affect the autonomic nervous system through either the sympathetic or parasympathetic nervous system. The frequency types are: very low frequency (VLF), low frequency (LF) and high frequency (HF). The VLF frequency is below the 0.05 HZ scoring band, the LF frequency is from 0.05 up to 0.15 HZ, and the HF is the range of 0.15 – 0.4 HZ (Jacob, 2010; McCraty & Childre, 2010; Bernston et al., 1991).

Based on the HRV power spectrum, a good HRV score is above 0.04 HZ of VLF to below 0.26 HZ of HF. The VLF band and the LF band of below 0.04 HZ belongs to the sympathetic nervous system and too low and not good, and also above the 0.26 HZ of HF belongs to the parasympathetic nervous system and too high and it is also not good (McCraty & Shaffer, 2015; Appelhans & Lueccken, 2006). The emWave PC biofeedback (1.0) software and device used for the purpose of this study, show scores for the frequency ranges are from 0 to 100 instead of 0 to 1, and in this scoring system, 0.15 means 15 and 0.4 means 40 (HeartMath.org 2016).

FINDINGS

The findings from this study showed that there were some positive changes in the levels of heart coherence of the participants from the baseline session to the three other data collection stages which were conducted under the Quick Coherence Technique (QCT). Based on the coherence ratio scoring results, in the baseline session which was conducted without the intervention of any self-regulatory technique (any stages of CQT), most of the participants had higher scores for the very low frequency (VLF) band which indicates the incoherent heart. But, through the application of three stages of the QCT, the low frequency (LF) scores of the participants increased significantly connected to the coherent heart.

Based on the coherence ratio scoring, the mean score of all 20 participants for the low frequency (LF) band of the baseline session was 18.6. But, their mean score for the heart focus was 24.6, for the heart breathing session was 37.3, and for the heart feeling session it was 35.9. The results showed the positive effects of the use of the QCT on the increase of heart coherence among university students. Table 1 shows the coherence ratio mean scores of all participants for the four sessions of HRV data.

Table 1
Illustrates the mean scores of HRV data collections sessions

Data Collection Sessions	Scores for each Session		
	VLF	LF	HF
Baseline	70.2	18.6	11.1
Heart Focus	60.3	24.6	15
Heart Breathing	51.9	37.3	10.6
Heart Feeling	47.2	35.9	16.8

DISCUSSION AND CONCLUSION

This study was carried out to evaluate the effectiveness of the use of the Quick Coherence Technique (QCT), as a self-regulatory technique, on the improvement of heart coherence of university students through the application of the heart rate variability (HRV) - biofeedback technology. The levels of heart coherence of the participants were assessed through four different stages, including the baseline. The data collection stages were the baseline session in which the data were collected without involvement of any self-regulatory technique, and the three other stages with the intervention of the QCT.

During the baseline session, the participants were mostly in *fight-and-flight* situation with their higher scores of very low frequency (VLF) band; because their HRV scores were in vibration and could not relax well. Based on the related work (e.g. McCraty & Shaffer, 2015; Jacob; 2010; Bernston et al., 1991), higher scores in VLF belong to the sympathetic nervous system (SNS) that associated with fight-and-flight situation and psychological stress. However, during the QCT stages, their heart coherence improved and their scores for the low frequency (LF) band which indicates the coherence situation increased significantly. As the low frequency (LF) scores significantly changed from baseline which was 18.6 to 24.6 for the Heart Focus session of the QCT, 37.3 for the Heart Breathing session, and 35.9 for the Heart Feeling session, and based on McCraty and Shaffer (2015) higher LF scores associated with flexibility, psychological well-being and good performance; thus, the results support the effectiveness of the use of the self-regulatory techniques, especially the QCT, on the improvement of psychological well-being.

The results from this study support McCraty and Shaffer (2015), Porges (2007), and McCraty and Zayas (2014) regarding the effectiveness of the use of HRV-technology and self-regulatory techniques on the increase of heart coherence of individuals. Based on the results from this study, the use of QCT had positive effects on the increase of low frequency (LF) scores and heart coherence of the participants. Moreover, the application of the HRV-biofeedback technology and self-regulatory techniques, including the QCT, would help university students to improve the levels of their heart coherence.

The main contribution of this study is to show the effectiveness of CQT self-regulatory technique and Heart Rate Variability (HRV)-Biofeedback Technology in the improvement of heart coherence among university students from different backgrounds in an Asian university

environment. However, as this study had a small number of participants and also used self-regulatory techniques, thus a wider sample would be useful and through the application of different self-regulatory techniques for better validity of the outcome.

ACKNOWLEDGEMENTS

REFERENCES

Bakar, H. A., & Mustaffa, C. S. (2013). Organizational communication in Malaysia organizations. *Corporate Communications: An International Journal, 18*(1), 87-109.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic Determinism: The modes of Autonomic Control, the Doctrine of Autonomic Space, and the Laws of Autonomic Constraint. *Psychological Review, 98*(4), 459-487.

Childre, D., & Rozman, D. (2003). *Transforming anger: The Heartmath solution for letting go of rage, frustration, and irritation.* Oakland: New Harbinger Publications.

Childre, D., Martin, H., & Childoe, D. (2000). *HeartMath solution: The Institute of HeartMath's revolutionary program for engaging the power of the heart's intelligence.* San Francisco: HarperOne.

Gudykunst, W. B. (2001). *Asian American ethnicity and communication.* Thousand Oaks, CA: Sage.

HeartMath.org. (2016). *Quick coherence technique.* Retrieved from http://cdn.quantumintech.com/emwave-pro-library/asmts-hrv.html.

Jacob, K.K. (2010). *Separating attention from arousal during TV viewing: using heart rate variability to track variations in sympathetic and parasympathetic activation* (Ph.D. Thesis). Indiana University Press, USA.

Lagos, L., Aschillo, E., Vaschill, B., Lehrer, P., Bates, M., & Pandina, R. (2008). Heart rate variability biofeedback for dealing with competitive anxiety: A case study. *Applied Psychophysiology and Biofeedback, 36*(3), 109–115.

McCraty, R., & Childre, D. (2010). Coherence: bridging personal, social, and global health. *Alternative Therapies in Health and Medicine, 1*6(4), 10-24.

McCraty, R., & Shaffer, F. (2015). Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Global Advances in Health and Medicine, 4*(1), 46-61.

McCraty, R., & Zayas, M. (2014). Cardiac coherence, self-regulation, autonomic stability, and psychosocial well-being. *Frontiers in Psychology, 5*, 1-13.

Musselman, D. L., Evans, D. L., & Nemeroff, C. B. (1998). The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Archives of General Psychiatry, 55*(1), 580-592.

Pignotti, M., & Steinbegr, M. (2001). Heart rate variability as an outcome measure for thought field therapy in clinical practice. *Journal of Clinical Psychology, 57*(10), 1193-1206.

Porges, S. W. (1997). Emotion: An evolutionary by product of the neural regulation of the autonomic nervous system. In C. S. Carter, B. Kirkpatrick & I. I. Lederhendler (Eds.), *Annals of the New York Academy of Sciences: Vol. 807. The integrative neurobiology of affiliation* (pp. 62–77). New York: New York Academy of Sciences.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74*(2), 116-43.

Quintana, D. S., Guastella, A. J., Outhred, T., Hickie, I. B., &Kemp, A. H. (2012). Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition. *International Journal of Psychophysiology, 86*(2), 168–172.

Rene, R. (2008). *The efficacy of a portable HRV feedback device in conjunction with mental health treatment of clients with major depressive disorder enrolled in a county welfare-to-work program.* (Ph.D. Thesis). Alliant International University Press, USA.

Reyes, F. J. (2014). Implementing heart rate variability biofeedback groups for veterans with posttraumatic stress disorder. *Biofeedback, 42*(4), 137-142.

Ross, M. W. (2011). *The evolution of education: Use of biofeedback in developing heart intelligence in a high school setting.* (PhD. Thesis). University of Calgary Press, USA.

Sutarto, A. P., Nubli A. W., & Zin, N. M. (2012). Resonant breathing biofeedback training for stress reduction among. *International Journal of Occupational Safety and Ergonomics, 18*(4), 549–561.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*(3), 354–381.

Thayer, J. F. (2007). What the heart says to the brain (and vice versa) and why we should listen. *Psychological Topics, 16*(2), 241-250.

Thurber, M. R. (2006). *Effects of heart-rate variability biofeedback training and emotional regulation on music performance anxiety in university students.* (Ph.D. Thesis). University Of North Texas Press, USA.

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 26 (3) JUL. 2018

The Editorial Board of the Journal of Science and Technology wishes to thank the following for acting as referees for manuscripts published in this issue of JST.

Nik Abdul Hadi Sapiaa
(*UTP, Malaysia*)

Nik Azuan Nik Ismail
(*UKM, Malaysia*)

Nisha Pandey
(*Gyancity Research Labs, India*)

Nur Azida Che Lah
(*UniKL, Malaysia*)

Obie Farobie
(*Hiroshima University, Japan*)

Ong Hong Choon
(*USM, Malaysia*)

Osi Arutanti
(*Hiroshima University, Japan*)

P. Ravi Kumar
(*JBCE, Brunei*)

Papita Das
(*Jadavpur University, India*)

Paul Sathiyan
(*KU, India*)

Perumal Kumar
(*Curtin University, Malaysia*)

Poonam Jindal
(*NITKKR, India*)

Pradeep Kumar Gupta
(*JUIT, India*)

Rajani V Honnungar
(*RNSIT, India*)

Rajesh Nithyanandam
(*Taylors University, Malaysia*)

Ravi Prakash
(*AIT, India*)

Rizka Zulhijah
(*UB, Indonesia*)

Rosni Abdullah
(*USM, Malaysia*)

Rusli Daik
(*UKM, Malaysia*)

Sandeep Dhariwal
(*LPU, India*)

Sandeep Kumar
(*IIT Roorkee, India*)

Saratha Sathasivam
(*USM, Malaysia*)

Sathya Shankara Sharma
(*MIT, India*)

Satnam Kaur
(*NITJ, India*)

Somasundaram Raghavan
(*UTHCT, USA*)

Sulong Ibrahim
(*UMT, Malaysia*)

Syamsul Rizal Abd Shukor
(*USM, Malaysia*)

T. Srinivas
(*IISC, India*)

Tanuj Kumar
(*LPU, India*)

Tay Chia Chay
(*UiTM, Malaysia*)

Tedi Kurniawan
(*UMP, Malaysia*)

Thamer Ahmed Mohamed
(*UPM, Malaysia*)

Uma Reddy
(*AMC, India*)

Viv Djanat Prasita
(*HTU, Indonesia*)

Yusuf Abba
(*UNIMAID, Nigeria*)

Zulfan Adi Putra
(*UTP, Malaysia*)

Pertanika Journals

Our goal is to bring high quality research to the widest possible audience

INSTRUCTIONS TO AUTHORS
(Manuscript Preparation & Submission Guide)

Revised: June 2016

Please read the Pertanika guidelines and follow these instructions carefully. Manuscripts not adhering to the instructions will be returned for revision without review. The Chief Executive Editor reserves the right to return manuscripts that are not prepared in accordance with these guidelines.

MANUSCRIPT PREPARATION

Manuscript Types

Pertanika accepts submission of mainly **four** types of manuscripts for peer-review.

1. **REGULAR ARTICLE**

   Regular articles are full-length original empirical investigations, consisting of introduction, materials and methods, results and discussion, conclusions. Original work must provide references and an explanation on research findings that contain new and significant findings.

   Size: Generally, these are expected to be between 6 and 12 journal pages (excluding the abstract, references, tables and/or figures), a maximum of 80 references, and an abstract of 100–200 words.

2. **REVIEW ARTICLE**

   These report critical evaluation of materials about current research that has already been published by organizing, integrating, and evaluating previously published materials. It summarizes the status of knowledge and outline future directions of research within the journal scope. Review articles should aim to provide systemic overviews, evaluations and interpretations of research in a given field. Re-analyses as meta-analysis and systemic reviews are encouraged. The manuscript title must start with "Review Article:".

   Size: These articles do not have an expected page limit or maximum number of references, should include appropriate figures and/or tables, and an abstract of 100–200 words. Ideally, a review article should be of 7 to 8 printed pages.

3. **SHORT COMMUNICATIONS**

   They are timely, peer-reviewed and brief. These are suitable for the publication of significant technical advances and may be used to:

   (a) report new developments, significant advances and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope;

   (b) report/discuss on significant matters of policy and perspective related to the science of the journal, including 'personal' commentary;

   (c) disseminate information and data on topical events of significant scientific and/or social interest within the scope of the journal.

   The manuscript title must start with "*Brief Communication*:".

   Size: These are usually between 2 and 4 journal pages and have a maximum of three figures and/or tables, from 8 to 20 references, and an abstract length not exceeding 100 words. Information must be in short but complete form and it is not intended to publish preliminary results or to be a reduced version of Regular or Rapid Papers.

4. **Others**

Brief reports, case studies, comments, concept papers, Letters to the Editor, and replies on previously published articles may be considered.

PLEASE NOTE: NO EXCEPTIONS WILL BE MADE FOR PAGE LENGTH.

Language Accuracy

Pertanika **emphasizes** on the linguistic accuracy of every manuscript published. Articles must be in **English** and they must be competently written and argued in clear and concise grammatical English. Contributors are strongly advised to have the manuscript checked by a colleague with ample experience in writing English manuscripts or a competent English language editor.

Author(s) **must provide a certificate** confirming that their manuscripts have been adequately edited. A proof from a recognised editing service should be submitted together with the cover letter at the time of submitting a manuscript to Pertanika**. All editing costs must be borne by the author(s)**. This step, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

Linguistically hopeless manuscripts will be rejected straightaway (e.g., when the language is so poor that one cannot be sure of what the authors really mean). This process, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

MANUSCRIPT FORMAT

The paper should be submitted in one column format with at least 4cm margins and 1.5 line spacing throughout. Authors are advised to use Times New Roman 12-point font and *MS Word* format.

1. *Manuscript Structure*
Manuscripts in general should be organised in the following order:

Page 1: *Running title*
This page should **only** contain the running title of your paper. The running title is an abbreviated title used as the running head on every page of the manuscript. The running title should not exceed 60 characters, counting letters and spaces.

Page 2: *Author(s) and Corresponding author information.*
This page should contain the **full title** of your paper not exceeding 25 words, with name(s) of all the authors, institutions and corresponding author's name, institution and full address (Street address, telephone number (including extension), hand phone number, and e-mail address) for editorial correspondence. First and corresponding authors must be clearly indicated.

The names of the authors may be abbreviated following the international naming convention. e.g. Salleh, A.B.[1], Tan, S.G[2]*., and Sapuan, S.M[3].

Authors' addresses. Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers:

George Swan[1] and Nayan Kanwal[2]
[1]Department of Biology, Faculty of Science, Duke University, Durham, North Carolina, USA., [2]Office of the Deputy Vice Chancellor (R&I), Universiti Putra Malaysia, Serdang, Malaysia.

A **list** of number of **black and white / colour figures and tables** should also be indicated on this page**.** Figures submitted in color will be printed in colour. See "*5. Figures & Photographs"* for details.

Page 3: *Abstract*
This page should **repeat** the **full title** of your paper with only the **Abstract** (the abstract should be less than 250 words for a Regular Paper and up to 100 words for a Short Communication), and **Keywords**.
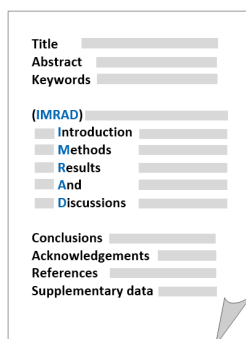
Keywords: Not more than eight keywords in alphabetical order must be provided to describe the contents of the manuscript.

Page 4: *Introduction*
This page should begin with the **Introduction** of your article and followed by the rest of your paper.

2. Text
Regular Papers should be prepared with the headings *Introduction, Materials and Methods, Results and Discussion*, *Conclusions, Acknowledgements, References, and Supplementary data* (if avavailble) in this order.

Title
Abstract
Keywords

(**IMRAD**)
 Introduction
 Methods
 Results
 And
 Discussions

Conclusions
Acknowledgements
References
Supplementary data

MAKE YOUR ARTICLES AS CONCISE AS POSSIBLE

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words Introduction, Materials and Methods, Results, And, Discussion. It indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: Title, Authors, Keywords, Abstract, Conclusions, and References. Additionally, some papers include Acknowledgments and Appendices.

The Introduction explains the scope and objective of the study in the light of current knowledge on the subject; the Materials and Methods describes how the study was conducted; the Results section reports what was found in the study; and the Discussion section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the Journal's instructions to authors.

3. Equations and Formulae
These must be set up clearly and should be typed double spaced. Numbers identifying equations should be in square brackets and placed on the right margin of the text.

4. Tables
All tables should be prepared in a form consistent with recent issues of Pertanika and should be numbered consecutively with Roman numerals. Explanatory material should be given in the table legends and footnotes. Each table should be prepared on a new page, embedded in the manuscript.

When a manuscript is submitted for publication, tables must also be submitted separately as data - .doc, .rtf, Excel or PowerPoint files- because tables submitted as image data cannot be edited for publication and are usually in low-resolution.

5. Figures & Photographs
Submit an **original** figure or photograph. Line drawings must be clear, with high black and white contrast. Each figure or photograph should be prepared on a new page, embedded in the manuscript for reviewing to keep the file of the manuscript under 5 MB. These should be numbered consecutively with Roman numerals.

Figures or photographs must also be submitted separately as TIFF, JPEG, or Excel files- because figures or photographs submitted in low-resolution embedded in the manuscript cannot be accepted for publication. For electronic figures, create your figures using applications that are capable of preparing high resolution TIFF files. In general, we require **300 dpi** or higher resolution for **coloured and half-tone artwork**, and **1200 dpi or higher** for **line drawings** are required.

Failure to comply with these specifications will require new figures and delay in publication.
NOTE: Illustrations may be produced in colour at no extra cost at the discretion of the Publisher; the author could be charged Malaysian Ringgit 50 for each colour page.

6. References
References begin on their own page and are listed in alphabetical order by the first author's last name. Only references cited within the text should be included. All references should be in 12-point font and double-spaced.

NOTE: When formatting your references, please follow the **APA reference style** (6th Edition). Ensure that the references are strictly in the journal's prescribed style, failing which your article will **not be accepted for peer-review**. You may refer to the *Publication Manual of the American Psychological Association* for further details (http://www.apastyle.org/).

7. General Guidelines

Abbreviations: Define alphabetically, other than abbreviations that can be used without definition. Words or phrases that are abbreviated in the introduction and following text should be written out in full the first time that they appear in the text, with each abbreviated form in parenthesis. Include the common name or scientific name, or both, of animal and plant materials.

Acknowledgements: Individuals and entities that have provided essential support such as research grants and fellowships and other sources of funding should be acknowledged. Contributions that do not involve researching (clerical assistance or personal acknowledgements) should **not** appear in acknowledgements.

Authors' Affiliation: The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved to another institution, the current address may also be stated in the footer.

Co-Authors: The commonly accepted guideline for authorship is that one must have substantially contributed to the development of the paper and share accountability for the results. Researchers should decide who will be an author and what order they will be listed depending upon their order of importance to the study. Other contributions should be cited in the manuscript's Acknowledgements.

Copyright Permissions: Authors should seek necessary permissions for quotations, artwork, boxes or tables taken from other publications or from other freely available sources on the Internet before submission to Pertanika. Acknowledgement must be given to the original source in the illustration legend, in a table footnote, or at the end of the quotation.

Footnotes: Current addresses of authors if different from heading may be inserted here.

Page Numbering: Every page of the manuscript, including the title page, references, tables, etc. should be numbered.

Spelling: The journal uses American or British spelling and authors may follow the latest edition of the Oxford Advanced Learner's Dictionary for British spellings.

SUBMISSION OF MANUSCRIPTS

Owing to the volume of manuscripts we receive, we must insist that all submissions be made electronically using the **online submission system ScholarOne™**, a web-based portal by Thomson Reuters. For more information, go to our web page and click "**Online Submission**".

Submission Checklist

1. **MANUSCRIPT**: Ensure your MS has followed the Pertanika style particularly the first four pages as explained earlier. The article should be written in a good academic style and provide an accurate and succinct description of the contents ensuring that grammar and spelling errors have been corrected before submission. It should also not exceed the suggested length.

   COVER LETTER: All submissions must be accompanied by a cover letter detailing what you are submitting. Papers are accepted for publication in the journal on the understanding that the article is **original** and the content has **not been published** either **in English** or **any other language(s)** or **submitted for publication elsewhere**. The letter should also briefly describe the research you are reporting, why it is important, and why you think the readers of the journal would be interested in it. The cover letter must also contain an acknowledgement that all authors have contributed significantly, and that all authors have approved the paper for release and are in agreement with its content.

   The cover letter of the paper should contain (i) the title; (ii) the full names of the authors; (iii) the addresses of the institutions at which the work was carried out together with (iv) the full postal and email address, plus telephone numbers and emails of all the authors. The current address of any author, if different from that where the work was carried out, should be supplied in a footnote.

   The above must be stated in the cover letter. Submission of your manuscript will not be accepted until a cover letter has been received.

2. **COPYRIGHT**: Authors publishing the Journal will be asked to sign a copyright form. In signing the form, it is assumed that authors have obtained permission to use any copyrighted or previously published material. All authors must read and agree to the conditions outlined in the form, and must sign the form or agree that the corresponding author can sign on their behalf. Articles cannot be published until a signed form *(original pen-to-paper signature)* has been received.

Please do **not** submit manuscripts to the editor-in-chief or to any other office directly. Any queries must be directed to the **Chief Executive Editor's** office via email to executive_editor.pertanika@upm.my.

Visit our Journal's website for more details at http://www.pertanika.upm.edu.my/home.php.

HARDCOPIES OF THE JOURNALS AND OFF PRINTS
Under the Journal's open access initiative, authors can choose to download free material (via PDF link) from any of the journal issues from Pertanika's website. Under "**Browse Journals**" you will see a link, "*Current Issues*" or "*Archives*". Here you will get access to all current and back-issues from 1978 onwards.

The **corresponding author** for all articles will receive one complimentary hardcopy of the journal in which his/her articles is published. In addition, 20 off prints of the full text of their article will also be provided. Additional copies of the journals may be purchased by writing to the Chief Executive Editor.

Contents