PROCEEDINGS E-BOOK

THE 2ND INTERNATIONAL CONFERENCE ON SCIENCE TECHNOLOGY & INNOVATION-MAEJO UNIVERSITY

ICSTI-MJU 2022

MARCH 18, 2022, CHIANG MAI, THAILAND

FACULTY OF SCIENCE, MAEJO UNIVERSITY
SCIENCE TECHNOLOGY AND INNOVATION INTERNATIONAL CONFERENCE

FACULTY OF SCIENCE, MAEJO UNIVERSITY

PROCEEDINGS E-BOOK

The 2nd International Conference on

Science Technology & Innovation-Maejo University

(2nd ICSTI-MJU) Chiang Mai, Thailand

March 18, 2022

Proceeding e-book publish online: July 28, 2022

Conference Chairman: Faculty of Science, Maejo University, Thailand

MESSAGE FROM THE DEAN OF FACULTY OF SCIENCE, MAEJO UNIVERSITY

The 2nd International Conference on Science Technology & Innovation-Maejo University (2nd ICSTI-MJU) will provide an excellent international forum for sharing knowledge and results in theory, methodology and applications of Science Technology and Innovation. The aim of the Conference is to provide a platform to the researchers and practitioners from both academia as well as industry to meet and share cutting-edge development in the field.

I am confident that 2nd ICSTI-MJU2022 will play an important role in encouraging activities in research and development of Innovations in Sciences and Technology, Biological Science and Technology, Mathematics, Statistics, Computer, Data Science, and related fields. It gives an excellent opportunity to forge collaborations between research institutions both within the country and with international partners.

On behalf of the Faculty of Science, Maejo University, I would like to express my sincere gratitude to the distinguished invited speakers for their presence and contributions to the conference. I also thank all the program committee members for their efforts in ensuring a rigorous review process to select high quality papers. In addition, I am also grateful to all our faculty members for organizing this very successful conference.

We expect that the ideas that have emerged in 2nd ICSTI-MJU will result in the development of further innovations for the benefit of scientific, industrial, and social communities.

T. Cheunbarn

Assistant Professor Dr. Tapana Cheunbarn
Dean of Faculty of Science, Maejo University

PREFACE

The 2nd ICSTI-MJU is hosted by Faculty of Science, Maejo University, Thailand. The conference looks for significant contributions to all major fields of Innovations in Science and Technology, Biological Science and Technology, Mathematics, Statistics, Computer, Data Science and related fields. The aim of the conference is to provide a platform to the researchers and practitioners from both academia as well as industry to meet and share cutting-edge development in the field. It is held online conference on March 18, 2022. It is honored by four keynote speakers: Professor Dr. Shigeo Maruyama, Department of Mechanical Engineering, School of Engineering, The University of Tokyo, Japan; Emeritus Professor Don McNeil, School of Mathematical and Physical Sciences, Macquarie University, Australia; Professor Dr. rer. nat. Andreas Schäffer, Director of the Institute for Environmental Research Chair of Environmental Biology and Chemodynamics Aachen University, Germany; and Professor Dr. Gomathi Velu, Department of Agricultural Microbiology, Directorate of Natural resource Management Tamil Nadu Agricultural University, India. A total of 37 papers from Thailand, Lao, India, Canada, United Kingdom, and Germany have undergone a strict peer-review process, and they are invited for oral presentation at the conference.

We would like to thank the organizing committees, session chairs, reviewers, and all technical staffs for their continuous hard work and professional efforts in organizing this conference. We would also like to extend our best gratitude to keynote speakers for their invaluable contribution and worthwhile ideas shared in the conference.

Conference Chairman of the 2nd ICSTI-MJU

July 28, 2022

Keynote Speakers:

1. **Professor Shigeo Maruyama**
   Department of Mechanical Engineering, School of Engineering,
   The University of Tokyo, Japan

2. **Emeritus Professor Don McNeil**
   School of Mathematical and Physical Sciences, Macquarie University,
   Australia

3. **Professor Dr. rer. nat. Andreas Schäffer**
   Director of the Institute for Environmental Research Chair of
   Environmental Biology and Chemodynamics Aachen University,
   Germany

4. **Professor Dr. Gomathi Velu**
   Department of Agricultural Microbiology, Directorate of Natural resource
   Management Tamil Nadu Agricultural University, India

Conference committees

Advisory Board

Associate Professor Dr. Weerapon Thongma
President of Maejo University, Chiang Mai, Thailand

Assistant Professor Dr. Tapana Cheunbarn
Dean of Faculty of Science, Maejo University, Chiang Mai, Thailand

Associate Professor Dr. Dara Phusanga
Faculty of Science, Maejo University, Chiang Mai, Thailand

Assistant Professor Dr. Sayan Unankard
Faculty of Science, Maejo University, Chiang Mai, Thailand

International Advisory Board

Professor Dr. Dankmar Böhning	UK
Professor Dr. Oliver Weichold	Germany

Conference Chairman

Associate Professor Dr. Wasin Charerntantanakul	Thailand
Associate Professor Dr. Viruntachar Kruefu	Thailand
Assistant Professor Dr. Suparut Narksitipan	Thailand
Assistant Professor Dr. Krisana Lanumteang	Thailand
Assistant Professor Dr. Surasak Kuimalee	Thailand
Dr. Panwad Sillapawattana	Thailand
Dr. Pornpan Uttamang	Thailand

International Committees

Professor Dr. Don McNeil	Australia
Professor Dr. Dankmar Bohning	United Kingdom
Professor Dr. Raja Ramachandran	India
Professor Dr. Chee Peng Lim	Australia
Professor Dr. Saravanakumar Ramasamy	Japan
Professor Dr. Yeol Je Cho	Korea
Professor Dr. Nicolae Barsan	Germany
Professor Dr. Juvy J. Monserate	Philippines
Professor Nicolae Barsan	Germany
Professor.Dr.Juvy J. Monserate	Philippines
Professor Dr. Oliver Weichold	Germany
Professor Dr. James A. Roth	USA
Professor Dr. Caroline Fossum	Sweden
Professor Dr. Dusan Palic	Germany
Associate Professor Dr. Akira Baba	Japan
Associate Professor Dr. Henry N. Roa	Ecuador
Associate Professor Dr. R.Sivaperegasam P.Rajanthiran	Malaysia
Associate Professor Dr. Akira Baba	Japan
Assistant Professor Dr. Saeid Hosseini	Oman
Assistant Professor Dr. Sanad Al Maskari	Oman
Assistant Professor Dr. Ling Chen	Taiwan
Assistant Professor Dr. Mosae Selvakumar Paulraj	Bangladesh
Dr. Martin C. H. Gruhlke	Germany
Dr. Alberth Nahas	Indonesia
Dr. Sen Wang	Australia

Local Committees

Professor Dr. Vilai Rungsardthong	Thailand
Associate Professor Dr. Wararit Panichkitkosolkul	Thailand
Associate Professor Dr. Anamai Na-udom	Thailand
Associate Professor Dr. Nipaporn Chutiman	Thailand
Associate Professor Dr. Piyapatr Busababodhin	Thailand
Associate Professor Dr. Lakhana Watthanacheewakul	Thailand
Associate Professor Dr. Kanit Mukdasai	Thailand
Associate Professor Dr. Kamsing Nonlaopon	Thailand
Associate Professor Dr. Thongchai Botmart	Thailand
Associate Professor Dr. Prasit cholamjiak	Thailand
Assistant Professor Dr. Urailuk Singthong	Thailand
Associate Professor Dr. Wanwilai Vittayakorn	Thailand
Associate Professor Dr. Siwapom Meejoo Smith	Thailand
Associate Professor Dr. Thitiphan Chimsook	Thailand
Associate Professor Dr. Chupong Pakpum	Thailand
Associate Professor Dr. Theerapol Thurakitseree	Thailand
Associate Professor Dr. Viruntachar Kruefu	Thailand
Associate Professor Dr.Dheerawan Boonyawan	Thailand
Associate Professor Dr. Wasin Charerntantanakul	Thailand
Associate Professor Dr. Panuwat Suppakul	Thailand
Associate Professor Dr. Phanchana Sanguansermsri	Thailand
Associate Professor Dr. Donruedee Sanguansermsri	Thailand
Associate Professor Dr. Suriyan Supapvanich	Thailand
Associate Professor Dr. Saengtong Pongjaroenkit	Thailand
Assistant Professor Dr. Sumaman Buntoung	Thailand
Assistant Professor Dr. Atchar Punya Jaroenjittichai	Thailand
Assistant Professor Dr. Waraporn Piyawit	Thailand
Assistant Professor Dr. Orada Chumphukam	Thailand
Associate Professor Dr. Sulawan Kaowphong	Thailand
Assistant Professor Dr. Anakhaorn Srisaipet	Thailand
Assistant Professor Dr. Napat Chantaramee	Thailand
Assistant Professor Dr. Navavan Thongmee	Thailand

Assistant Professor Dr. Tanin Tangkuaram	Thailand
Assistant Professor Dr. Ratchadaporn Puntharod	Thailand
Assistant Professor Dr. Chokchai Yatongchai	Thailand
Assistant Professor Dr. Surasak Kuimalee	Thailand
Assistant Professor Dr. Sumaman Buntoung	Thailand
Assistant Professor Dr. Atchar Punya Jaroenjittichai	Thailand
Assistant Professor Dr. Nareeluk Nakaew	Thailand
Assistant Professor Dr. Juangjun Jumpathong	Thailand
Assistant Professor Dr. Pattana Kakumyan	Thailand
Assistant Professor Dr. Srikanjana Klayraung	Thailand
Assistant Professor Dr. Thanwanit Thanyasiriwat	Thailand
Assistant Professor Dr. Yuppayao Kophimai	Thailand
Assistant Professor Dr. Rattana Leadsuwansri	Thailand
Assistant Professor Dr. Krisana Lanumteang	Thailand
Assistant Professor Dr. Chanikan Emharuethai	Thailand
Assistant Professor Dr. Jerawan Saerour	Thailand
Dr. Panwad Sillapawattana	Thailand
Dr. Pichamon Limcharoenchat	Thailand
Dr. Pornpan Uttamang	Thailand
Dr. Parkpoom Choomanee	Thailand
Dr. Radshadaporn Janta	Thailand
Dr. Ponlkrit Yeesin	Thailand
Dr. Suphat Phongthai	Thailand
Dr. Supajit Sraphet	Thailand
Dr. Noodchanath Kongchouy	Thailand
Dr. Julárat Chumnaul	Thailand
Dr. Orasa Anan	Thailand
Dr. Jenjira Thipcha	Thailand
Dr. Narong Chanlek	Thailand
Dr. Chaipattana Saisa-ard	Thailand
Dr. Narong Chanlek	Thailand

Conference Agenda

Time	Friday 18 March 2022
8.00-8.30	**Registration**
8.30-9.30	**Opening Ceremony**
9.30-9.45	**Coffee Break**
9.45-10.30	**Professor Dr. Shigeo Maruyama** Department of Mechanical Engineering, School of Engineering, The University of Tokyo, **Japan**
10.30-11.15	**Emeritus Professor Dr. Don McNeil** School of Mathematical and Physical Sciences, Macquarie University, **Australia**
11.15-11.15	Oral Presentation
12.15-13.00	**Break**
13.00-13.45	**Professor Dr. rer. nat. Andreas Schäffer** Director of the Institute for Environmental Research Chair of Environmental Biology and Chemodynamics Aachen University, **Germany**
13.45-14.30	**Professor Dr. Gomathi Velu**; Department of Agricultural Microbiology, Directorate of Natural resource Management Tamil Nadu Agricultural University, **India**
14.30-17.30	Oral Presentation (continue)
17.30	Conference Closing

Note : THAILAND/Bangkok; Time Zone: UTC/GMT+7

Conference Schedule

2nd ICSTI-MJU 2022, Maejo University, Chiang Mai, Thailand

Time	18 March 2022	
8.00-8.30	Registration	
8.30-9.30	Opening ceremony	
9.30-9.45	Coffee break	
9.45-10.30	**Keynote speaker: Professor Dr. Shigeo Maruyama Japan**	**CVD synthesis and application of 1D vdW hetero-structures based on single-walled carbon nanotubes**
10.30-11.15	**Keynote speaker: Emeritus Professor Dr. Don McNeil Australia**	**Mapping, Graphing and Forecasting Global Land Surface Temperature using Data from NASA**

Note : THAILAND/Bangkok; Time Zone: UTC/GMT+7

Oral Presentation Session			
	Session 1: Mathematics, Statistics, Computer & Data Science	**Session 2: Innovations in Science & Technology**	**Session 3: Biological Science and Environmental**
11.15-11.30	iAB65002: Effects of educational campaigns on mathematical models to control the spread of rotavirus infection (Thailand)	iAB65001: Gallic acid, the isolation and method development for the quantitative determination from Phyllanthus emblica L. extract (Thailand)	iAB65004: On-site diagnosis of acute hepatopancreatic necrosis disease in shrimp farms using visual detection of isothermal nucleic acid amplification (Thailand)
11.30-11.45	iAB65005: Estimation of population size based on zero-truncated, one-inflated, and covariate information (Thailand)	iAB65003: Blood glucose biosensor based on glucose oxidase co-immobilized with copper(I) oxide at manganese (IV) oxide on graphene quantum dots in chitosan scaffold (Thailand)	iAB65006: Development of hybrid tofu with dietary fiber supplementation as an alternative for dietary protein (Thailand)
11.45-12.00	iAB65008: Construction of Bivariate Copulas for Multivariate Statistical Process Control: a review (Thailand)	iAB65009: Facile synthesis of platinum-gold nanoparticles@carbon dot-graphene oxide and characterization by deposition with chitosan on electrode surface (Thailand)	iAB65007: The integration of passive and active satellite sensors for water resources extraction and mapping for sugarcane plantation area in Northeastern, Thailand: Geoinformatics approaches (Thailand)
12.00-12.15	iAB65010: Modelling Road Accidents Injuries and Fatalities in Suratthani province of Thailand using Conway-Maxwell-Poisson Regression (Thailand)	iAB65013: Drying Kinetics Equation of Curcuma comosa (Curcuma xanthorrhiza Roxb.) (Thailand)	iAB65012: Effect of various extraction solvents on efficiency of natural pigment based dye sensitized solar cell (Thailand)
12.15-13.00	**Break**		
13.00-13.45	**Keynote speaker: Professor Dr. rer. nat. Andreas Schäffer Germany**		**Pesticide mixtures - Gaps in environmental risk assessment**
13.45-14.30	**Keynote speaker: Professor Dr. Gomathi Velu India**		**Role of biopolymers in sustainable environment**

Oral Presentation Session (Continue)			
	Session 1: Mathematics, Statistics, Computer & Data Science	**Session 2: Innovations in Science & Technology**	**Session 3: Biological Science and Environmental**
14.30-14.45	iAB65011: Travel time impact in multi-modal mode: a case study of the new railway line Ban Phai to Nakhon Phanom (Thailand)	iAB65018: Smart Application for Finding Parking and Car Owners (Thailand)	iAB65015: A low-cost alternative photosensitizer for dye-sensitizied solar cells using Sandoricum koetjape natural dye (Thailand)
14.45-15.00	iAB65016: Reward-Based Crowdfunding under Blockchain for Beauty Industry (Thailand)	iAB65027: Improving the properties of paper from rice straw (Lao)	iAB65020: The Natural resources management on basis of biodiversity and community utilization in Mae Hong Son Province: A case study at Ban Na Pla Chad (Thailand)
15.00-15.15	iAB65017: Question Generation from Thai Wiki paragraphs by using Transformer (Thailand)	iAB65029: Development of electrochemical dopamine sensor based on conducting polymer-gold nanoparticle composites (Thailand)	iAB65021: The Reduction of Forest Fire by the Local Community Participation on Resources Management in Mae Tha Lu Village, Mae Hong Son Province (Thailand)
15.15-15.30	iAB65022: Forecasting air pollution particulate matter (PM2.5) in Chiangmai, Thailand (Thailand)	iAB65031: Synthesis of Biphasic calcium phosphate (BCP) from Lime Mud (Thailand)	iAB65023: Application of needle flowers, portulaca flowers and teak leaves as natural sensitizers for dye-sensitized solar cells (Thailand)
15.30-15.45	**Coffee Break**		
15.45-16.00	iAB65026: Estimating under-reporting of COVID-19 in infected cases in Chiangmai (Thailand) by capture-recapture methods (Thailand)	iAB65034: Biomass derived activated porous carbon from Manilkara kauki L bark as potential adsorbent for the removal of Congo red dye from wastewater: adsorption isotherm and kinetic studies (India)	iAB65024: Plant-growth media from hydrogels based on poultry-feather waste (Germany)
16.00-16.15	iAB65028: The Parameters Estimation for the Discrete Weibull Regression Model With Type-I Right Censored Data (Thailand)	iAB65035: Purification humic acid as precursor to synthesize reduced graphene by hydrothermal method with microwave assisted method (Thailand)	iAB65025: The reduction of unpleasant odor in pseudo ceramic by using natural filler (Thailand)

	Session 1: Mathematics, Statistics, Computer & Data Science	Session 2: Innovations in Science & Technology	Session 3: Biological Science and Environmental
16.15-16.30	iAB65030: On the asymptotic normality of the estimator for ratio of Binomial proportions (Thailand)	iAB65037: Characterization of bioplastic (Poly Hydroxy Butyrate) producing bacteria from termite mound soil (India)	iAB65036: Decolorization of Methylene Blue (MB) by Biosorption of Termite Fungus Termitomyce sp. (India)
16.30-16.45	iAB65032: Estimating the number of cannabis abuse in Chiangmai (Thailand) using single source data (Thailand)	-	iAB65039: Organic soil amendments' impact on the fate of the herbicide MCPA and microbial communities (Germany)
16.45-17.00	iAB65033: application of the Horvitz Thompson estimator based upon Poisson Shanker model for capture-recapture data (Thailand)	-	-
17.00-17.15	iAB65038: Impacts of Road Characteristics and Weather Conditions on Fatal Accidents in Surat Thani Province of Thailand (Thailand)	-	-
17.15-17.30	iAB65040: Mathematical Modeling on Transmission and Optimal Control Strategies of Corruption Dynamics (Thailand)	-	-
17.30	**Conference Closing**		

CONTENT

Paper ID		Page
iAB65001	Gallic acid, the isolation and method development for the quantitative determination from Phyllanthus emblica L. extract *Chiramet Auranwiwat, Wiriyaporn Sumsakul, Sinee Siricoon, Thongkorn Ploypetchara, Waraporn Sorndech, Siriporn Butseekhot*	1-7
iAB65002	Effeck of educational campaigns on mathematical models to control the spread of rotavirus infection *Anuwat Jirawattanapanit*	8-20
iAB65003	Blood glucose biosensor based on glucose oxidase co-immobilized with copper(I) oxide at manganese (IV) oxide on graphene quantum dots in chitosan scaffold *Onanan Injumnong, Anchana Preechaworapun, Tanin Tangkuarum*	21-27
iAB65005	Estimation of population size based on zero-truncated, one-inflated, and covariate information *Tita Jongsomjit, Rattana Lerdsuwansri, Kn"sana Lanumteang*	29-35
iAB65006	Development of hybrid tofu with dietary fiber supplementation as an alternative for dietary protein *Sarod Chaiworn, Suphat Phongthai, Churdsak Jaikang, Jetsada Ruangsuriya*	36-50
iAB65012	*Effect of Various Extraction Solvents on Efficiency of Natural Pigment Based Dye-Sensitized Solar Cell* *Ponnambalam Sabarikirishwaran, Pattranan Junluthin, Yuwalee Unpaprom and Rameshprabu Ramaraj*	51-59
iAB65011	Travel time impact in multi-modal mode: a case study of the new railway line Ban Phai to Nakhon Phanom *Khanuengnj Prakhammin, Nawapon Nakharutai , Manad Khamkong*	60-73
iAB65017	Question Generation from Thai Wiki paragraphs by using Transformer *Napat Srichantapong, Wachirawut Thamvise*	74-88
iAB65018	Smart Application for Finding Parking and Car Owners *Wiyada Yawai*	89-102

CONTENT (CONTINUE)

Paper ID		Page
iAB65025	The reduction of unpleasant odor in pseudo ceramic by using natural filler *Tithinun Rattanaplome, Napat Chantaramee*	103-111
iAB65027	Improving the Tensile Strength of Rice Straw Paper by Different Paper Additive *Leulee Nortoualee, Siraporn Cheunbarn, Tapana Cheunbarn, Tawat Soitong and Supattra Wongsaenmai*	112-116
iAB65028	The Parameters Estimation for the Discrete Weibull Regression Model with Type-I Right Censored Data *Dusit Chaiprasithikul, Monthira Duangsaphon*	113-130
iAB65029	Development of Electrochemical Dopamine Sensor Based on Conducting Polymer-Gold Nanoparticle Composites *Rapiphun Janmanee, Sosista Chatwet, Nutsarin Sukonpakdee, Anchana Preechaworapun, Kulwadee Pinwattana, Padarat Ninjiaranai, Saengrawee Sn'wichai*	131-136
iAB65035	Purification humic acid as precursor to synthesize reduced graphene by hydrothermal method with microwave assisted method *Kanyawee Kaewpradub, Ratchadaporn Puntharod*	137-143

Gallic acid, the isolation and method development for the quantitative determination from *Phyllanthus emblica* L. extract

Chiramet Auranwiwat[1,*], Wiriyaporn Sumsakul[1], Sinee Siricoon[2], Thongkorn Ploypetchara[2], Waraporn Sorndech[2], Siriporn Butseekhot[2]

[1]Expert Center of Innovative Herbal Products, Thailand Institute of Scientific and Technological Research, Pathum Thani, Thailand
[2]Expert Center of Innovative Health Food, Thailand Institute of Scientific and Technological Research, Pathum Thani, Thailand

**Corresponding Author Email: chiramet@tistr.or.th*

Abstract. The phytochemical investigation of *Phyllanthus emblica* L. fruits extract led to the isolation of gallic acid. The structure of this compound was characterized by spectroscopic data especially ^{1}H and ^{13}C NMR data. The method validation for quantitative determination of gallic acid was developed by using RP-HPLC . The chromatographic separation was archieved with VDSpher PUR C18-E column (4.6 x 250 mm, 5 μm), injection volume 10 μL. The mobile phase consisting of ACN and 0.05% H_3PO_4 was flowed with 1.0 mL/min and detected at 272 nm. The resulting of calibration curve showed a good linearity with r^2 = 0.9995. LOD and LOQ values were 8.36 and 25.34 μg/mL, respectively. The accuracy showed the %recovery between 96.73±1.53-103.34±0.97. while precision exhibited %RSD at 1.42%. This developed method was applied to determine gallic acid in *P. emblica* extract and the result was calculated to contain 2.67 %w/w.

Keywords: Gallic acid; *Phyllanthus emblica*; HPLC; method validation

1. Introduction

Gallic acid, a phenolic compound which contained carboxy and three hydroxy units at position 1 and 3, 4, 5 of aromatic ring, respectively. This compound has been found in various kinds of plant such as the leaves of *Labisa pumila,* roots and bark of *Phyllanthus niruri* [1]. The putative biosynthesis pathway of gallic acid has been reported with difference routes. First pathway was shikimate pathway, shikimic acid was oxidation to dehydroshikimic acid that can tautomerized to enol-form then dehydroshikimate dehydrogenase to obtain gallic acid. These pathways also investigated using enzymatic process using HPLC monitoring of shikimic acid, dehydroshikimic acid, $NADP^+$ and gallic acid [2,3]. Furthermore, the initiation of L-

phenylalanine then conversed to caffeic acid and followed by protocatechuic acid oxidation to be gallic acid [1]. The numerous biological activities of gallic acid have been published especially antioxidant, anticancer, anti-inflammatory, antimicrobial and antiviral activities [3].

India gooseberry, as a common name of *Phyllanthus emblica* L. is a member of Euphorbiaceae family. This plant is distributed in subtropical and tropical areas such as Thailand, Pakistan, Uzbekistan, Sri Lanka, China and Malaysia. This plant has been used in Thai herbal remedy such as TriPala recipe for prevention of diarrhea, hemorrhage, jaundice, dyspepsia and sickness [4]. Several publications were reported on phytochemical investigation of *P. emblica* found various groups of secondary metabolites including flavonoids [5-7], phenolic compounds [4,8,9] and terpenoids [10-12]. Gallic acid is one of major compound which found in *P. emblica* so it can be marker of the extract. Some of these compounds exhibited the interesting pharmaceutical activities, for example, antioxidant activity [8,9], antiviral activity [11] and cytotoxicity [7,11]. Herein, we isolated gallic acid from Indian gooseberry fruits and developed the method for quantitative determination.

2. Experimental

2.1 General experimental procedures

The NMR spectra were measure on a 400 MHz Bruker NMR spectrometer. Chemical shift were recorded in part per million (δ) in methanol-d_4 (δ_H 3.31 and δ_C 49.0 ppm). The HRESIMS data was obtained on a MS Q-TOF Agilent Technologies mass spectrometer. Thin-layer chromatography (TLC) was performed on silica gel $60GF_{254}$ (Merck). Column chromatography (CC) was carried out on Sephadex LH-20 with methanol. Solvents for extraction and chromatography were distilled pior to use. The acetonitrile was purchased in HPLC grade from RCI Labscan, Thailand while acetic acid was obtained from Merck, Germany. Gallic acid was purchased from Sigma, Switzerland. The ultrasonic bath was Bandelin Sonorex, Banoelin.

2.2 Plant materials

The fruits of *P. emblica* was collected from Kanchanaburi, Thailand in 2020. This plant was identified and has been deposited at Bangkok Herbarium, Bangkok, Thailand (specimen no. Wiriyaporn02)

2.3 Extraction and isolation

The fruits of *P. emblica* (900 g) were cut to small pieces and then dried in the oven at 50 °C. The drird fruits (82.7 g) were extracted by EtOH for 3 hr and 3 times then solvent was concentrated under reduced pressure to obtain EtOH extract as brown viscous oil. Crude EtOH extract (8.52 g) was subjected to Sephadex LH-20 with MeOH to give 5 fractions (A-E). Fraction C (920.0 mg) was further isolated by Sephadex LH-20 by elution with MeOH to provide 6 subfractions (C1-C6). Compound **1** (11.2 mg) as a brown solid, was obtained from subfraction C5 (28.2 mg).

2.4 Method validation

2.4.1 HPLC method condition

The HPLC was performed on Water 600 controller, equipped with 486 detector, USA using VDSpher PUR C18-E column (250 x 4.6 mm, 5 μm), injection loop 10 μL, flow rate 1 mL/min and detection wavelength 272 nm. The gradient condition was followed ACN (A) and 0.05% H_3PO_4 (B) as mobile phase: 0-5 min, 10%B, 6-15 min, 40%B, 16-25 min, 40%B, 26-30 min, 90%B and 35-40 min, 100%B.

2.4.2 Preparation of standard stock solution

Accurately weight 5 mg of standard gallic acid then transfer to a 5 mL volumetric flask and dissolved in 2 mL MeOH. The stock standard solution was sonicated for 5 min and adjust volume with MeOH to 5 mL and obtained concentration 1 mg/mL of stck solution.

2.4.3 Calibration curve

The calibration curves were built by plotting the peak area with concentration of each standard solution. The gallic acid standard solution were prepared as concentration 12.5, 25, 50, 100 and 250 μg/mL from stock solution into volumetric flask and adjust with MeOH. The linear regression equation was explained by y = ax ± b, where x and y are the concentration and peak area, respectively. The linearity was established in term of correlation coeddicient (r^2). The study of each concentration was measured in triplicate.

2.4.4 Limite of detection (LOD) and limit of quantification (LOQ)

The limite of detection (LOD) was the lowest concentration of compound which this method can detect while limit of quanlification (LOQ) was the lowest concentration that can detected and reliable result. The values of LOD and LOQ were calculated from the calibration curve using formula as LOD = 3.3SD or 3.3δ/S whereas LOQ = 10SD or 10δ/S.

2.4.5 Accuracy

The accuracy was perfomed by a recovery study. The addition of standard gallic acid in sample then determined total gallic acid content in sample. All samples were analyzed in triplicate and recovery were calculated. The equation used to analyte the recovery percentage was (detected amount - original amount)/spike x 100

2.4.6 Precision

Precision study was determined the six spots of sample which injected in the same condition as gallic acid. The precisions were expressed as relative standard deviation (RSD, %)

2.5 Sample analysis

Accurately weighted 25 mg of EtOH extract of *P. emblica* and transfer to a 25 mL volumetric flask and dissolved in 10 mL MeOH. The sample solution was sonicated for 5 min and adjusted with MeOH to 25 mL.

3. Results and Discussion

*3.1 Characterization of compound **1***

The ethanolic extract of P. emblica fruits led to the isolation a compound **1**. The structural elucidation was determined using NMR spectroscopic technique and mass spectroscopy. The HRESIMS of compound **1** showed sodiated molecular ion peak at *m/z* 169.0144 $[M-H]^-$ consisted with molecular formula of $C_7H_6O_5$ (Figure 1). The ^1H NMR data exhibited at δ_H 7.11 (2H, *s*, H-2, H-6) while ^{13}C NMR spectrum found 5 resonances for 7 carbons at δ_C 170.1 (C=O), 144.4 (C-3 and C-5), 138.1 (C-4), 120.8 (C-1) and 110.0 (C-2 and C-8) (Figure 2). A comparision of NMR data with the literature, compound **1** was characterized as gallic acid as shown in Table 1 [13].

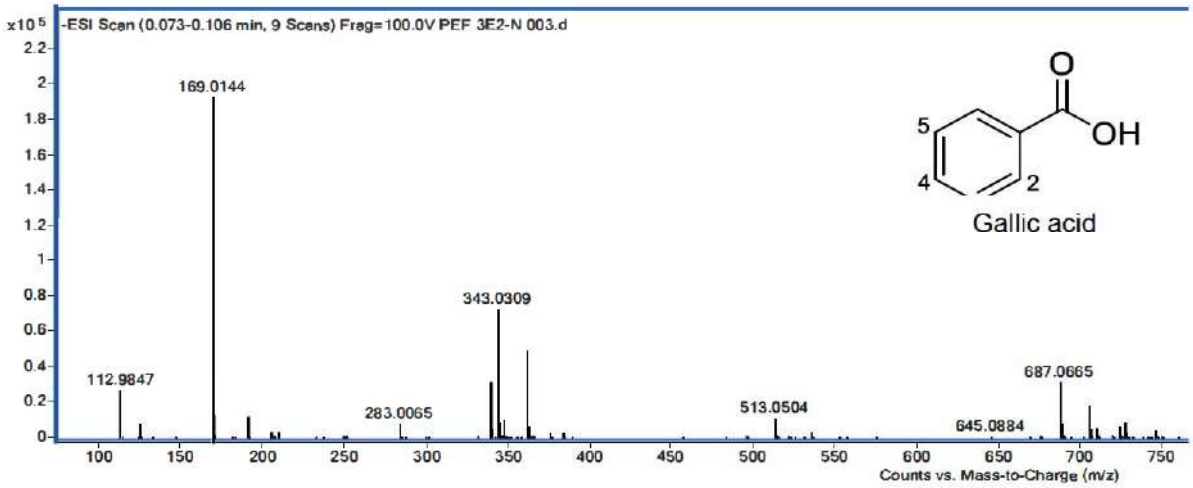


Figure 1. HRESIMS spectrum of compound **1**

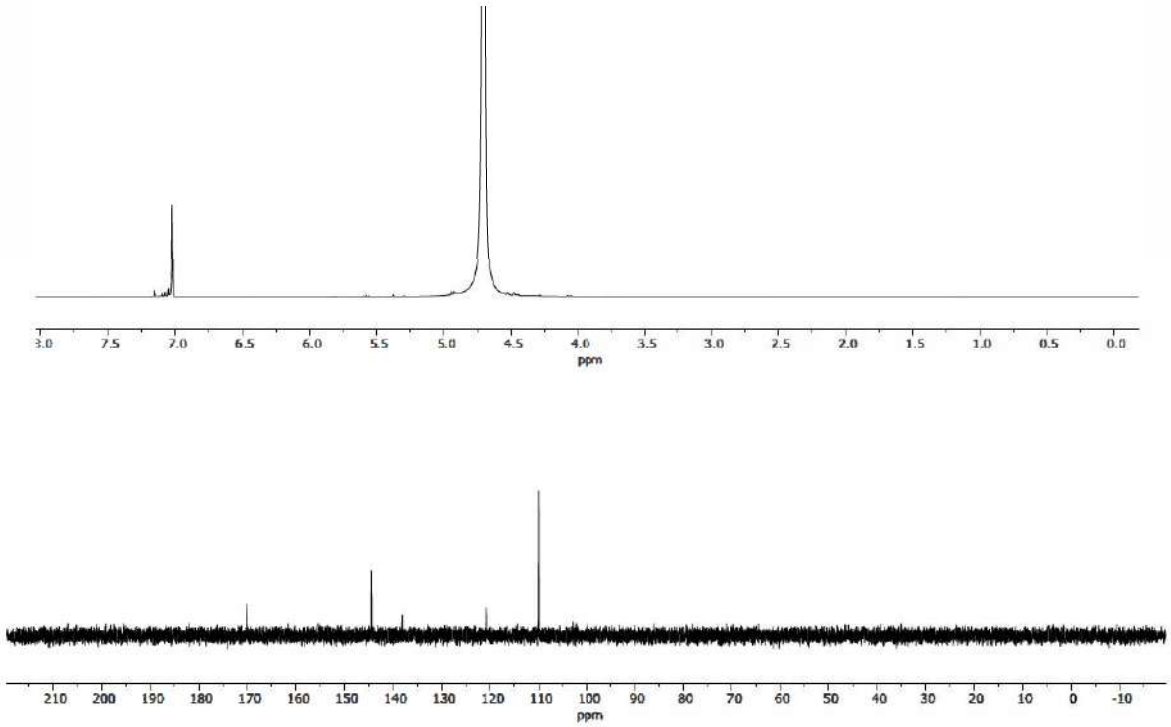


Figure 2. ^{1}H (400 MHz) and ^{13}C (100 MHz) spectra of compound **1** in D_2O

Table 1. A comparision NMR data of compound **1** with gallic acid

Compound **1**[a]		Gallic acid[b]	
Postion	δ_H (ppm)	Postion	δ_H (ppm)
2/6	7.11	2/6	7.03
		OH	1.29
Postion	δ_C (ppm)	Postion	δ_C (ppm)
1	120.8	1	119.9
2/6	110.0	2/6	108.1
3/5	144.4	3/5	146.5
4	138.1	4	139.6
C=O	170.1	C=O	170.5

[a] recorded in D_2O
[b] recorded in methanol-d_4

3.2 Method validation

3.2.1 Calibration curve, LOD and LOQ

The calibration curve for gallic acid presented a good correlation coefficient (r^2 = 0.9995) and the formula y = 32172x + 21035 in the range of concentration between 12.5-250 μg/mL (Figure 3). The LOD and LOQ values were calculated by equation then results showed LOD and LOQ were 8.36 and 25.34 μg/mL, respectively.

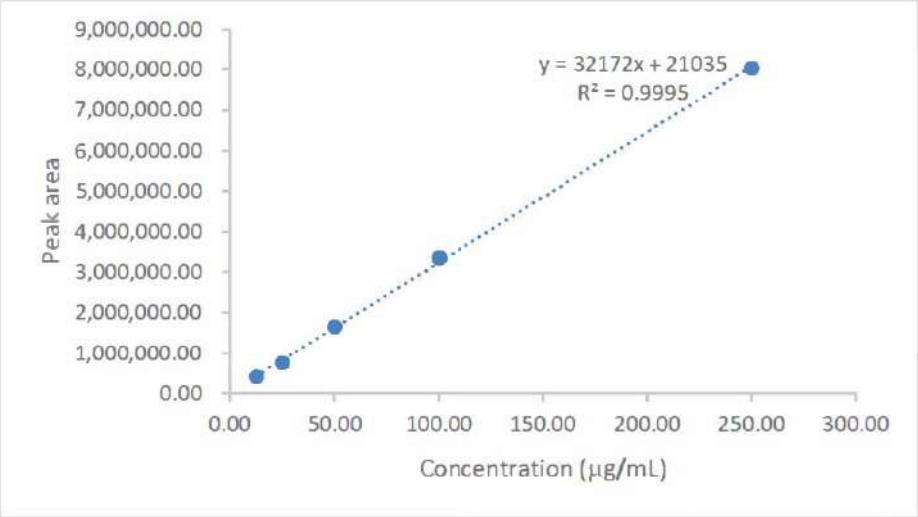


Figure 3. Calibration curve of gallic acid in the range 12.50-250.00 µg/mL

3.2.2 Accuracy and precision

The accuracy test was determined %recovery of standard gallic acid using the standard addition method.The results showed %recovery between 96.73±1.53-103.34±0.97. Moreover, precison of this method was observed six replicate analyses of extract. The relative standard value for this method was 1.42% that indicates this method is good precision as shown in Table 2.

Table 2. The results of accuracy and precision of method validation

Samples	%Recovery	%RSD
P. emblica extract		1.42
P. emblica extract+GA 10 ug/mL	96.73±1.53	1.58
P. emblica extract+GA 20 ug/mL	97.23±0.74	0.76
P. emblica extract+GA 30 ug/mL	103.34±0.97	0.93

3.3 The determination of gallic acid in extract

The crude extract of *P. emblica* was measured the gallic acid content with the optimized condition and validated method. The calculation of average peak area compared with the calibration curve of gallic acid was performed. HPLC chromatogram establish peak of gallic acid at 14.6 min when compared with standard compound. Thus, gallic acid content in the ethanolic extract as 27.53 µg/mL or 2.67% w/w in sample (Figure 4).

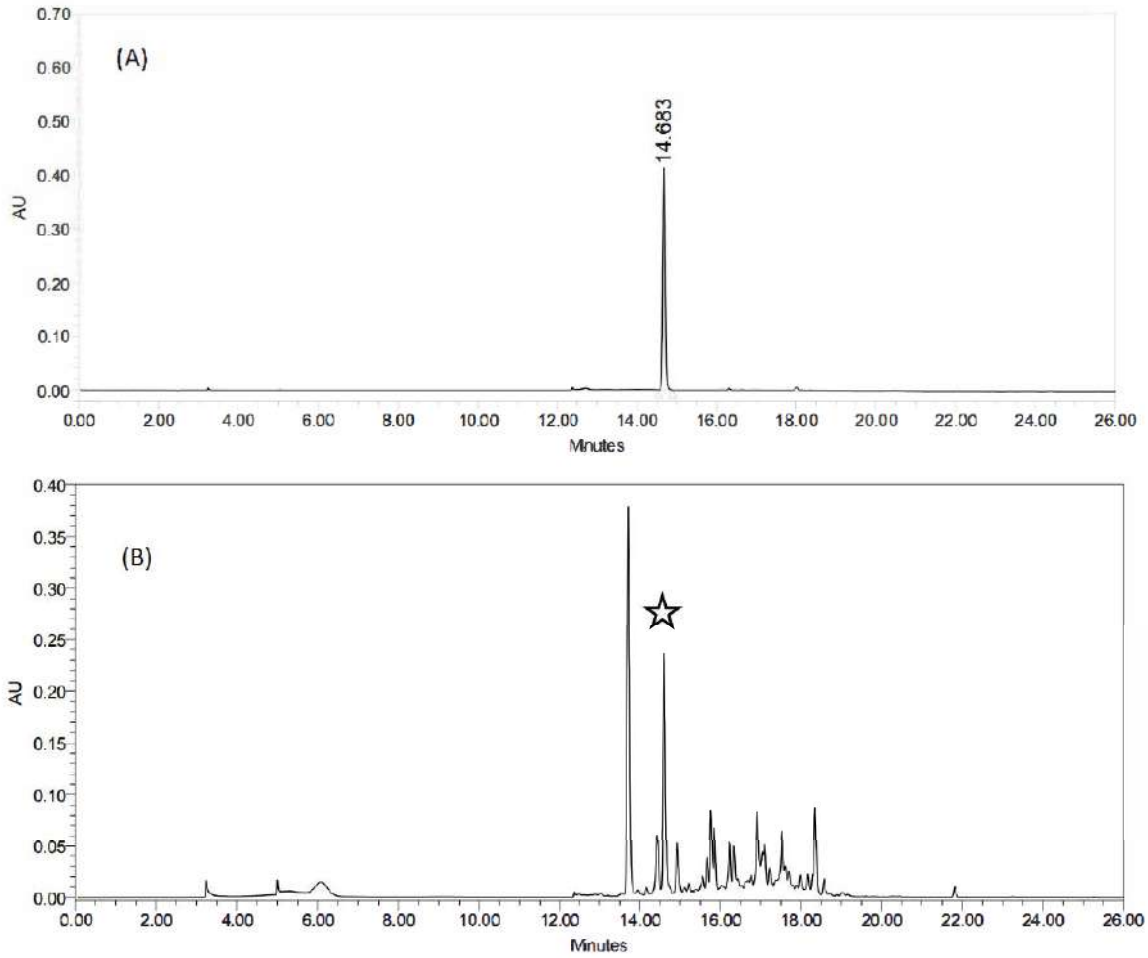


Figure 4. HPLC chromatogram of standard gallic acid (A) and *P. emblica* extract (B)

4. Conclusion

The study of ethnolic extract of *P. emblica* the resulting the isolated of gallic acid. The method for determination gallic acid has been developed and validated. The result of method validation exhibited a good linearity with r^2 = 0.9995 while LOD and LOQ values were 8.36 and 25.34 μg/mL, respectively. The test of accuracy observed the %recovery of gallic acid between 96.73±1.53-103.34±0.97. The good precision of the method was establish in precision test with %RSD at 1.42%. To apply in *P. emblica* extract, the determination of gallic acid in extract was 27.53 μg/mL or 2.67% w/w. This method contains the results in the range of ICH guideline so it can apply for quality control of product in the industry. This method can be choice for analysis gallic acid in real sample due to the difference of instrument might get different chromatogram.

Acknowledgements

We would like to thank Thailand Institute of Sciencetific Research for facilitites supporting. CA. acknowledge Assoc. Prof. Dr. Thunwadee Limtharakul for recording NMR spectra.

References

[1] Fernandas, F.H.A. and Salgado, H.R.N. (2016) Gallic acid: Review of methods of determination and quantification, *Critical Review in Analytical Chemistry 46*, 257-265.

[2] Ossipov, V., Salminen, J.-P., Ossipova, S., Haukioja, E., and Pihlaja, K. (2003) Gallic acid and hydrolysable tannins are formed in birch leaves from an intermediate compound of the shikimate pathway, *Biochemical Systematics and Ecology 31*, 3-16.

[3] Badhani, B., Sharma, N., and Kakkar, R. (2015) Gallic acid: a versalite antioxidant with promising therapeutic and industrial applications, *RSC Advances 5*, 27540-27557.

[4] Kumaranm, A., and Karunakaran, R.J. (2006) Nitric oxide radical scavenging active components from *Phyllanthus emblica* L., *Plant Foods for Human Nutrition 61*, 1-5

[5] Zhang, Y.-J., Abe, T., Tanaka, T., Yang, C.-R., and Kouno, I. (2002) Two new acylated flavanone glycosides from the leaves and brances of *Phyllanthus emblica*, *Chemical and Pharmaceutical Bulletin 50*, 841-843.

[6] Zhang, Y.-J., Nagao, T., Tanaka, T., Yang, C.-R., Okabe, H., and Kouno, I. (2004) Antiproliferative activity of the main constituents from *Phyllanthus emblica*, *Biological and Pharmaceutical Bulletin 27*, 251-255.

[7] El-Desouky, S.K., Ryu, S.Y., and Kim, Y.-K. (2008) A new cytotoxic acylated apigenin glucoside from *Phyllanthus emblica* L., *Natural Product Research 22*, 91-95.

[8] Luo, W., Zhao, M., Yang, B., Ren, J., Shen, G., and Rao, G. (2011) Antioxidant and antiproliferative capacities of phenolics purified from *Phyllanthus emblica* L. fruits, *Food Chemistry 126*, 277-282.

[9] Zhang, J., Miao, D., Zhu, W.-F., Xu, J., Liu, W.-Y., Kitdamrongthan, W., Manosroi, J., Abe, M., Akihisa, T., and Feng, F. (2017) Biological activities of phenolics from fruits of *Phyllanthus emblica* L. (Euphorbiaceae), *Chemistry & Biodiversity 14*, e1700404.

[10] Khatum, M., Billah, M., and Quader, Md. A. (2012) Sterols and sterol glucoside from *Phyllanthus* species, *Dhaka University Journal of Science 60*, 5-10.

[11] Lv, J.-J., Yu, S., Xin, Y., Cheng, R.-R., Zhu, H.-T., Wang, D., Yang, C.-R., Xu, M., and Zhang, Y.-J. (2015a) Anti-viral and cytotoxic norbisabolane sesquiterpenoid glycosides from *Phyllanthus emblica* and their absolute configurations, *Phytochemistry 117*, 123-134.

[12] Lv, J.-J., Yu, S., Xin, Y., Zhu, H.-T., Wang, D., Cheng, R.-R., Yang, C.-R., Xu, M., and Zhang, Y.-J. (2015b) Stereochemistry of cleistanthane diterpenoid glucosides from *Phyllanthus emblica, RSC Advances 5*, 29098-29107.

[13] Lubis, M.Y., Marpaung, L., Nasution, M.P., and Simanjuntak, P. (2018) Gallic acid from pods of jiringa (*Archidendron jiringa* [Jack] I.C. Nielsen) and its antioxidant, *Asian Journal of Pharmaceutical and Clinical Research 11*, 114-117.

Effects of educational campaigns on mathematical models to control the spread of rotavirus infection

Anuwat Jirawattanapanit[1*]
[1]Department of Mathematics, Faculty of Education, Rajabhat Phuket University, Phuket, Thailand

**Corresponding Author Email: anuwat.j@pkru.ac.th*

Abstract. This research is to develop and evaluate the stability of a mathematical model for controlling the spread of rotavirus infection. The model is analyzed using standard methods including the equilibrium point and the stability of the equilibrium points. The education campaign rate in the model, the basic reproductive number and numerical solutions are studied. We found that the education campaign rate is the factor affecting the model. If the infection's population is educated and follows the hypothesis of this model, then the spread of rotavirus infection will decrease and there will be no epidemic. It concluded that when the value of the education campaign rate decreases, the number of infected human increases. Therefore, the basic reproductive number is greater than one, meaning the Rotavirus infection will occur in the community. On the other hand, when the value of the education campaign increases, the number of infected human decreases. The basic reproductive number is less than one, meaning that the Rotavirus infection will have died out the community.

Keywords: Mathematical model, Rotavirus infection, Education Campaign

Introduction

Studying the mathematical models of epidemics makes it possible to know the epidemic and the results obtained from the model, help researchers understand the factors that can control the spread of the disease. Including having a correct understanding of the transmission of the disease. The study also identifies the strengths of a mathematical model capable of altering the characteristics of an epidemic. Analyzing the model, data shows the effectiveness of understanding the evolution of the epidemic and understanding disease control measures. Therefore, the results of this study are highly beneficial in reducing the risk of infection. Infection transmission and epidemic control. Diarrhoea is a disorder of the gastrointestinal tract. It would call liquid stools three times a day in each year. There are up to 1.7 billion diarrhoea patients worldwide. There will be distension, abdominal pain, nausea, vomiting and often taken. If it is longer than three weeks, diarrhoea will be called chronic. If it healed within three weeks, it is called acute diarrhoea, bacteria and viruses. Countries in the United Kingdom report 13,000 diarrhoea patients with Rotavirus a year. Australia has the number of infected people.It is as high as 32,000 people per year, and in Africa, the death rate of children are 15%. Thailand and Southeast Asia were found that 43-56% of younger children more than five years with diarrhoea who need to be hospitalized, are caused by a virus called Rotavirus,

which is a virus in RNA Group (Double-stranded RNA virus) in the family Reoviridae 7 species (A, B, C, D, E, F, G). When receiving the Rotavirus into the body, a short incubation period of fewer than 48 hours (duration from 1-7 days), When entering the digestive tract, the small intestine will destroy the small intestine wall, resulting in water absorption, reduced water and minerals and Enzyme for indigestion of carbohydrates causing diarrhoea no mucus or blood. Therefore, Rotavirus Is the cause of diarrhoea worldwide, resulting in approximately 527,000 deaths per year. Diarrhoea from Rotavirus is usually found in young children under five years, but in the past two years, it increased in adults. It may be infected by caring for a sick child or infected by contamination in the environment, and the Rotavirus is the most common in the winter. Rotavirus diarrhoea is caused by an infection with food and water directly through the mouth or indirectly after exposure to contaminated faeces or toys contaminated with faeces [3]. Diarrhoea from Rotavirus is usually found in groups with the symptoms within 1-3 days, and a patient can spread the infection for more than a week after the onset. The disease symptoms are fever, vomiting, diarrhoea, and diarrhoea in patients with severe symptoms. They need to be hospitalized due to dehydration, and if the treatment is unsuitable, it may be dangerous. It can have recurrent infections, do many times later. Symptoms will be less. From the mathematical model, the spread of diarrhoea from Rotavirus is known. The model and the spread of disease help the researchers understand the factors that can control the spread of the disease and have a correct understanding of the transmission. In addition, the strengths of mathematical models can change the characteristics of the disease and the parameters related to the disease. The Mathematical study epidemics of Rotavirus, keeping in mind the consequences of the epidemic and helping researchers better understanding the factors that control the spread of the disease [2]. In addition, the strengths of the mathematical model can also modify the characteristics of the epidemic and various parameters associated with the disease [5]. So, the results of this study would be highly beneficial in reducing the risk of infection and Rotavirus infection control. This research aims to develop and evaluate the stability of mathematical modelling for controlling the spread of Rotavirus infection on the Education Campaign. The model is analyzed using standard methods, the equilibrium point, stability, and analytic solutions [6]. The effectiveness of the Education Campaign in mathematical modelling and numerical solutions is studied.

Model Formulation

In our model, we assume that the human population is one constant because the birth rates and the death rates of the human population are equal. Therefore, the total number of human people denote by. The human N population are divided into four classes; the susceptible human (S), the exposed human (E), the infected human (I) and the recovered human (R). The diagram of four classes of the human population and the crucial parameters are used, representing the Rotavirus infection dynamics model. That is shown in Figure 1.

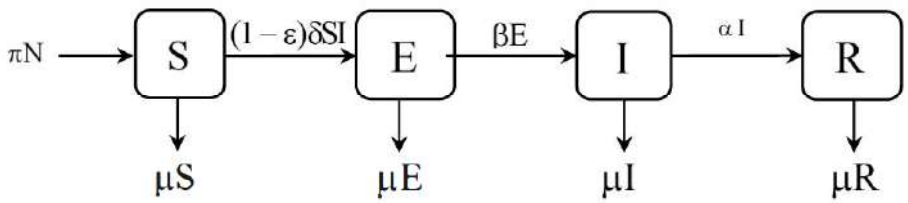


Figure 1. Flow chart of the dynamical transmission of Rotavirus infection.

The transitions between model classes can be now expressed by the following system of first order differential equations:

$$\frac{dS}{dt} = \pi N - (1-\varepsilon)\delta SI - \mu S \quad (1)$$

$$\frac{dE}{dt} = (1-\varepsilon)\delta SI - E(\beta+\mu) \quad (2)$$

$$\frac{dI}{dt} = \beta E - I(\mu+\alpha) \quad (3)$$

$$\frac{dR}{dt} = \alpha I - \mu R \quad (4)$$

with $N = S + E + I + R$

Where;

$S(t)$ is the susceptible human populations at time t

$E(t)$ is the exposed human populations at time t

$I(t)$ is the infected human populations at time t

$R(t)$ is the recovered human populations at time t

N is the total number of human populations

π is the birth rate of human populations

ε is the effectiveness of education campaign

δ is the probability that virus transmitted from infected human to susceptible human

β is the proportional rate for people exposed to the infected human populations

α is the recovery rate of infected human populations

μ is the natural death rate of human populations

Model Analysis

Since the model monitors human population, all the associated parameters and state variables are non-negative is $t \geq 0$. It is easy to show that the state variables of the model remain non-negative for all non-negative initial conditions [1]. The biological feasible region

$$\Psi = \left\{ (S, E, I, R) \in R_+^4 : N \to \frac{\pi}{\mu} \right\}$$

Lemma 1. The closed ψ is positively invariant and attracting.

Proof. Adding (1)-(4) give the rate of change of the total population.

$$\frac{dN}{dt} = \frac{dS}{dt} + \frac{dE}{dt} + \frac{dI}{dt} + \frac{dR}{dt}$$

$$\frac{dN}{dt} = \pi N - (1-\varepsilon)\delta SI - \mu S + (1-\varepsilon)\delta SI - E(\beta+\mu) + \beta E - I(\mu+\alpha) + \alpha I - \mu R$$

$$\frac{dN}{dt} = \pi N - \mu S - E\mu - I\mu - \mu R$$

$$\frac{dN}{dt} = \pi N - \mu(S + E + I + R)$$

$$\frac{dN}{dt} = \pi N - \mu N$$

$$\frac{1}{N}dN = (\pi - \mu)dt$$
$$\int \frac{1}{N}dN = \int (\pi - \mu)dt$$

$$\ln N = (\pi - \mu)t + c$$
$$N = e^{(\pi-\mu)t+c}$$
$$N(t) = N_0 e^{(\pi-\mu)t} \qquad \because N_0 = e^c$$

Thus, the total human population (N) are bounded by $\pi - \mu$, so that $\frac{dN}{dt} = 0$ whenever $\pi = \mu$. It can be shown that $N(t) = N_0 e^{(\pi-\mu)t}$. In particular $N(t) = N_0 e^{(\pi-\mu)t}$, if $N_0 e^{(\pi-\mu)t} > 0$. Hence, the region Ψ is positively invariant and attracts all solutions in R_+^4

Basic Reproductive Number

The basic reproductive number (R_0) is defined as the expected number of secondary cases produced by a single infection in a completely susceptible population, by using the next generation method and used spectral radius [8]. We have rewritten the system in matrix form $\frac{dx}{dt} = F(x) - V(x)$

Here $F(x)$ gives the rate of appended of new infections in a compartment and $V(x)$ gives the transferring of individuals. We obtained,

$$F(X) = \begin{bmatrix} 0 \\ (1-\varepsilon)\delta SI \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad V(X) = \begin{bmatrix} -\pi N + (1-\varepsilon)\delta SI + \mu S \\ \beta E + \mu E \\ -\beta E + \mu I + \alpha I \\ -\alpha I + \mu R \end{bmatrix}$$

where:

$$FV^{-1}(E_0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{(1-\varepsilon)\delta\beta N}{(\beta+\mu)(\mu+\alpha)} & \frac{(1-\varepsilon)\delta N}{\gamma+\varphi} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ and Spectral Radius from $FV^{-1}(E_0)$,

that is $\rho[FV^{-1}(E_0)] = \frac{(1-\varepsilon)\delta\beta N}{(\beta+\mu)(\mu+\alpha)}$. We have the basic reproductive number as shown,

$$R_0 = \frac{(1-\varepsilon)\delta\beta N}{(\beta+\mu)(\mu+\alpha)}$$

Stability Analysis

In this section, the stability of equilibrium can be analyzed by using the Jacobian matrix of the model at the disease free equilibrium. Referring to the results of Vanden Driessche and Watmough (2002), the stability of this system as shown in the follow theorem.

Theorem 1: The disease free equilibrium of the system about the equilibrium E_0, is local asymptotically stable if R_0 >1 and unstable if R_0 <1.

Proof. The Jacobian matrix of the model (Eqs. 1-4) evaluated a $E_0(S,E,I,R) = E_0(N,0,0,0)$ is obtained the local stability of equilibrium point is determined from the Jacobian matrix of the system of ordinary differential equation. The equation (1), (2), (3) and (4) evaluated at the equilibrium point. The Jacobian matrix is

$$J_0 = \begin{bmatrix} -\mu & 0 & -(1-\varepsilon)\delta N & 0 \\ 0 & -(\beta+\mu) & (1-\varepsilon)\delta N & 0 \\ 0 & \beta & -(\mu+\alpha) & 0 \\ 0 & 0 & \alpha & -\mu \end{bmatrix}$$

$$\det(J_0 - \lambda I) = \begin{vmatrix} -\mu-\lambda & 0 & -(1-\varepsilon)\delta N & 0 \\ 0 & -(\beta+\mu+\lambda) & (1-\varepsilon)\delta N & 0 \\ 0 & \beta & -(\mu+\alpha+\lambda) & 0 \\ 0 & 0 & \alpha & -\mu-\lambda \end{vmatrix}$$

$$\det(J_0 - \lambda I) = (\mu - \lambda)^2 [\lambda^2 + (\beta + 2\mu + \alpha)\lambda + (\beta + \mu)(\mu + \alpha) - (1 - \varepsilon)\beta\delta N]$$

The eigenvalues of the Jacobian matrix J_0 are obtained by solving $\det(J_0 - \lambda I) = 0$, then the characteristic equation as follows: $(\mu - \lambda)^2 [\lambda^2 + (\beta + 2\mu + \alpha)\lambda + (\beta + \mu)(\mu + \alpha) - (1 - \varepsilon)\beta\delta N] = 0$

where; $\lambda_{1,2} = -\mu < 0$ and $\lambda_{3,4} = \dfrac{-(\beta + 2\mu + \alpha) \pm \sqrt{(\beta+2\mu+\alpha)^2 - 4[(\beta+\mu)(\mu+\alpha)-(1-\varepsilon)\beta\delta N]}}{2}$

The two roots of $\lambda^2 + A\lambda + B = 0$ will be negative real part if they satisfy the Routh-Hurwitz criteria.

1) $A = \beta + 2\mu + \alpha$
2) $B = (\beta+\mu)(\mu+\alpha) - (1-\varepsilon)\beta\delta N$
3) A>0
4) B>0

Theorem 2: The endemic equilibrium of the system Eqs.(1)-(4) for the equilibrium

$$E_1(S^*, E^*, I^*, R^*) = \left(\frac{\pi N}{(1-\varepsilon)\delta I^* + \mu}, \frac{(1-\varepsilon)\delta S^* I^*}{\beta + \mu}, \frac{\beta E}{\mu + \alpha}, \frac{\alpha I^*}{\mu}\right),$$

is local asymptotically stable if $R_0 > 1$, and unstable if $R_0 < 1$.

Proof.

$$J_1 = \begin{bmatrix} -(1-\varepsilon)\delta I^* - \mu & 0 & -(1-\varepsilon)\delta S^* & 0 \\ (1-\varepsilon)\delta I^* & -(\beta+\mu) & (1-\varepsilon)\delta S^* & 0 \\ 0 & \beta & -(\mu+\alpha) & 0 \\ 0 & 0 & \alpha & -\mu \end{bmatrix}$$

$$\det(J_1 - \lambda I) = \begin{vmatrix} -(1-\varepsilon)\delta I^* - \mu - \lambda & 0 & -(1-\varepsilon)\delta S^* & 0 \\ (1-\varepsilon)\delta I^* & -(\beta+\mu+\lambda) & (1-\varepsilon)\delta S^* & 0 \\ 0 & \beta & -(\mu+\alpha+\lambda) & 0 \\ 0 & 0 & \alpha & -\mu-\lambda \end{vmatrix}$$

$$\begin{aligned}\det(J_1 - \lambda I) = (\mu+\lambda)[&(\lambda^3 + ((1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu)\lambda^2 + (((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) \\ &+ ((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*)\lambda \\ &+ ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* \\ &- ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)]\end{aligned}$$

The eigenvalues of the Jacobian matrix J_1 are obtained by solving $\det(J_1 - \lambda I) = 0$, we provided the characteristic equation as follows:

$$\begin{aligned}&(\mu+\lambda)[(\lambda^3 + ((1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu)\lambda^2 + (((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) + \\ &((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*)\lambda \\ &+((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* - ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)] = 0\end{aligned}$$

The three roots of $\lambda^3 + A\lambda^2 + B\lambda + C = 0$ will be negative real part if they satisfy the Routh-Hurwitz criteria.

1) $A = (1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu$

2) $B = ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) + ((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*$

3) $C = ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* - ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)$

4) $AB > C$

Equilibrium Points

The model will be analyzed to investigate the equilibrium points by using the standard method for analyzing our model. The system has two possible equilibrium points. In the case of the absence of the disease, that is $I=0$. Given,

$$X = \begin{bmatrix} S \\ E \\ I \\ R \end{bmatrix}, \quad F(X) = \begin{bmatrix} 0 \\ (1-\varepsilon)\delta SI \\ 0 \\ 0 \end{bmatrix} \text{ and } \quad V(X) = \begin{bmatrix} -\pi N + (1-\varepsilon)\delta SI + \mu S \\ \beta E + \mu E \\ -\beta E + \mu I + \alpha I \\ -\alpha I + \mu R \end{bmatrix}$$

From equations (1) - (4) finding the Jacobian max of F(x) and V(x) evaluated at $E_*(S, E, I, R)$. We follow that,

$$J = \begin{bmatrix} -(1-\varepsilon)\delta I-\mu & 0 & -(1-\varepsilon)\delta S & 0 \\ (1-\varepsilon)\delta I & -(\beta+\mu) & (1-\varepsilon)\delta S & 0 \\ 0 & \beta & -(\mu+\alpha) & 0 \\ 0 & 0 & \alpha & -\mu \end{bmatrix}$$

The disease-free equilibrium point

We used the Jacobian max of F(x) and V(x) evaluated at $E_0(S,E,I,R) = E_0(N,0,0,0)$

We obtained $E_0(N,0,0,0)$, then

$$J_0 = \begin{bmatrix} -\mu & 0 & -(1-\varepsilon)\delta N & 0 \\ 0 & -(\beta+\mu) & (1-\varepsilon)\delta N & 0 \\ 0 & \beta & -(\mu+\alpha) & 0 \\ 0 & 0 & \alpha & -\mu \end{bmatrix}$$

$$\det(J_0 - \lambda I) = \begin{vmatrix} -\mu-\lambda & 0 & -(1-\varepsilon)\delta N & 0 \\ 0 & -(\beta+\mu+\lambda) & (1-\varepsilon)\delta N & 0 \\ 0 & \beta & -(\mu+\alpha+\lambda) & 0 \\ 0 & 0 & \alpha & -\mu-\lambda \end{vmatrix}$$

$$\det(J_0 - \lambda I) = (\mu-\lambda)^2[\lambda^2 + (\beta+2\mu+\alpha)\lambda + (\beta+\mu)(\mu+\alpha) - (1-\varepsilon)\beta\delta N]$$

where: $\det(J_0 - \lambda I) = 0$

$$(\mu-\lambda)^2[\lambda^2 + (\beta+2\mu+\alpha)\lambda + (\beta+\mu)(\mu+\alpha) - (1-\varepsilon)\beta\delta N] = 0$$

where; $\lambda_{1,2} = -\mu < 0$ and

$$\lambda_{3,4} = \frac{-(\beta+2\mu+\alpha) \pm \sqrt{(\beta+2\mu+\alpha)^2 - 4[(\beta+\mu)(\mu+\alpha) - (1-\varepsilon)\beta\delta N]}}{2}$$

The endemic equilibrium point

In the case of the disease in present, that is $I^* > 0$. Hence,

$$E_1(S^*, E^*, I^*, R^*) = \left(\frac{\pi N}{(1-\varepsilon)\delta I^* + \mu}, \frac{(1-\varepsilon)\delta S^* I^*}{\beta+\mu}, \frac{\beta E}{\mu+\alpha}, \frac{\alpha I^*}{\mu}\right)$$

The Jacobian is defined as follows:

$$J_1 = \begin{bmatrix} -(1-\varepsilon)\delta I^* - \mu & 0 & -(1-\varepsilon)\delta S^* & 0 \\ (1-\varepsilon)\delta I^* & -(\beta+\mu) & (1-\varepsilon)\delta S^* & 0 \\ 0 & \beta & -(\mu+\alpha) & 0 \\ 0 & 0 & \alpha & -\mu \end{bmatrix}$$

$$\det(J_1 - \lambda I) = \begin{vmatrix} -(1-\varepsilon)\delta I^* - \mu - \lambda & 0 & -(1-\varepsilon)\delta S^* & 0 \\ (1-\varepsilon)\delta I^* & -(\beta+\mu+\lambda) & (1-\varepsilon)\delta S^* & 0 \\ 0 & \beta & -(\mu+\alpha+\lambda) & 0 \\ 0 & 0 & \alpha & -\mu-\lambda \end{vmatrix}$$

$$\begin{aligned}\det(J_1 - \lambda I) = (\mu+\lambda)[&(\lambda^3 + ((1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu)\lambda^2 + (((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) \\ &+ ((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*)\lambda \\ &+ ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* \\ &- ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)]\end{aligned}$$

where: $\det(J_1 - \lambda I) = 0$

$$\begin{aligned}&(\mu+\lambda)[(\lambda^3 + ((1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu)\lambda^2 + (((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) + \\ &((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*)\lambda \\ &+((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* - ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)] = 0\end{aligned}$$

where: $\lambda^3 + A\lambda^2 + B\lambda + C = 0$ and $AB > C$

1) $A = (1-\varepsilon)\delta I^* + \beta + \alpha + 3\mu$

2) $B = ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta) + ((1-\varepsilon)\delta I^* + \mu)(\mu+\alpha) + (\mu+\beta)(\mu+\alpha) - (1-\varepsilon)\beta\delta S^*$

3) $C = ((1-\varepsilon)\delta I^* + \mu)(\mu+\beta)(\mu+\alpha) + (1-\varepsilon)^2\beta\delta^2 S^2 I^* - ((1-\varepsilon)\beta\delta S^*)((1-\varepsilon)\delta I^* + \mu)$

Numerical Analysis

In this section, we would like to present the numerical simulation of our model. The parameter values that we used in the numerical simulation are given in Table 1.

Table 1. Parameters values used in numerical simulation at disease free state

Description	Parameters	Values
The total number of human populations	N	1,500 persons
The birth rate of human populations	π	22 day^{-1}
The probability that virus transmitted from infected human to susceptible human	δ	1.59×10^{-4} day^{-1}
the proportional rate for people exposed to the infected human populations	β	2.5×10^{-1} day^{-1}
The natural death rate of human populations	μ	1.2×10^{-2} day^{-1}
The recovery rate of infected human populations	α	1.4×10^{-1} day^{-1}
The effectiveness of education campaign	ε	0 – 1

By solving the system of differential equations. The numerical results showed the relationship between the parameters of education campaign rate and basic reproductive Number in Table 2.

Table 2. The relationship between the parameters of effectiveness of Education Campaign and Basic Reproductive Number

Education campaign rate (ε)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	1
Basic reproductive Number (R_0)	1.497	1.348	1.198	1.048	0.898	0.749	0.599	0.449	0.299	0.150	0

The analysis model was found that the stability of equilibrium points when the Education Campaign $\varepsilon = 0.4$, have basic reproductive number $R_0 = 0.8980$, and the Education Campaign $\varepsilon = 0$, the disease endemic equilibrium $R_0 = 1.4970$. The Education Campaign rate is the factor affecting to the mathematical modeling.

Stability of the endemic state

We changed the values of Education Campaign $\varepsilon = 0.999$, and the values of the other parameters are given in Table 1. We have obtained the eigenvalues, and basic reproductive numbers are: λ_1 =-0.01200000, λ_2 =-0.01200000, λ_3 = -0.304000181, λ_4 =-0.523999819 and R_0 = 0.002360<1. All of the eigenvalues are negative, and the basic reproductive is less than one, the endemic equilibrium point will be local asymptotically stable [4]. as shown in Figure 2.

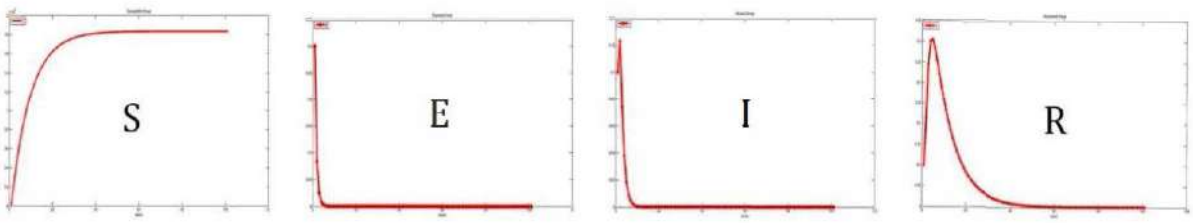


Figure 2. Time series (S) Susceptible human, (E) Exposed human, (I) Infected human, (R) recovered human. The solutions approach to the disease-free equilibrium state.

We changed the values of Education Campaign $\varepsilon = 0.4$, and kept the values of the other values of parameters to be those given in Table 1, we obtained the eigenvalues and basic reproductive numbers are: λ_1 =-0.01200000, λ_2 =-0.01200000, λ_3 =-0.15181961, λ_4 =-0.26218039 and $R_0 = 0.8980 < 1$. Since all of eigenvalues have been negative and the basic reproductive has been less than one, the endemic equilibrium point will be local asymptotically stable. as shown in Figure 3.

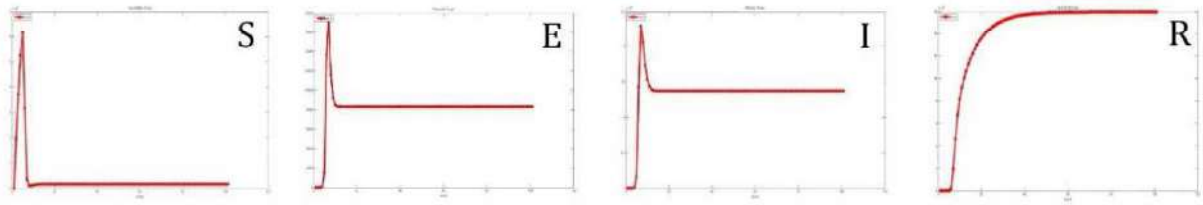


Figure 3. Time series (S) Susceptible human, (E) Exposed human, (I) Infected human, (R) recovered human. The solutions approach the endemic equilibrium state to the disease-free state.

We changed the values of Education Campaign ε =0.5, 0.6, 0.7, 0.8, and 0.85. Also, kept the values of the other values of parameters to be those given in Table 1, we have the eigenvalues and basic reproductive numbers are: $\lambda^4 + A\lambda^3 + B\lambda^2 + C\lambda + D = 0$,A=16.89534009, B=7.02090688, C=0.73721679, D=0.00786477, ABC>C^2+A^2D and R_0=0.898, 0.749, 0.599, 0.449, 0.299 and 0.150 (R_0 <1). Since all of eigenvalues have been negative and the basic reproductive has been less than one, the endemic equilibrium point will be local asymptotically stable [4] as shown in Figure 4-7.

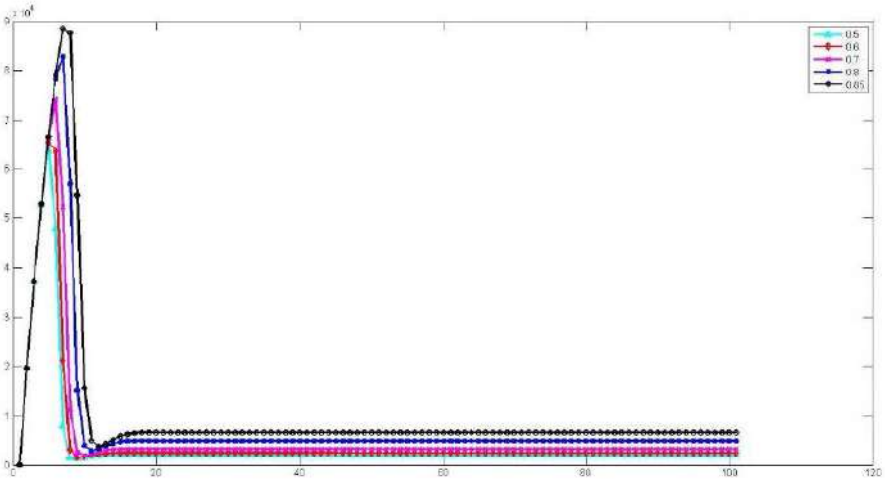


Figure 4. Time series of Susceptible human on Education Campaign rate ε =0.5, 0.6, 0.7, 0.8 and 0.85, respectively. The values of parameters are in the text. The solutions approach the endemic equilibrium state to the disease-free state.

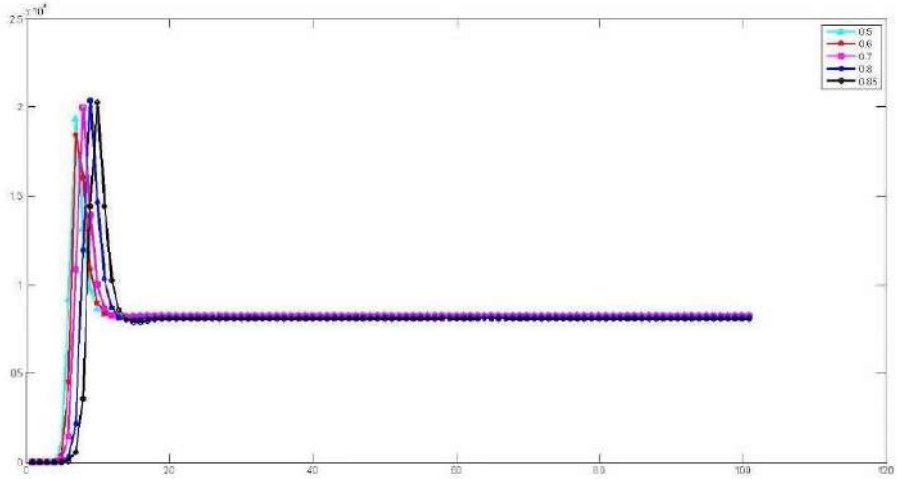


Figure 5. Time series of Exposed human on Education Campaign rate ε =0.5, 0.6, 0.7, 0.8 and 0.85, respectively. The values of parameters are in the text. The solutions approach the endemic equilibrium state to the disease-free state.

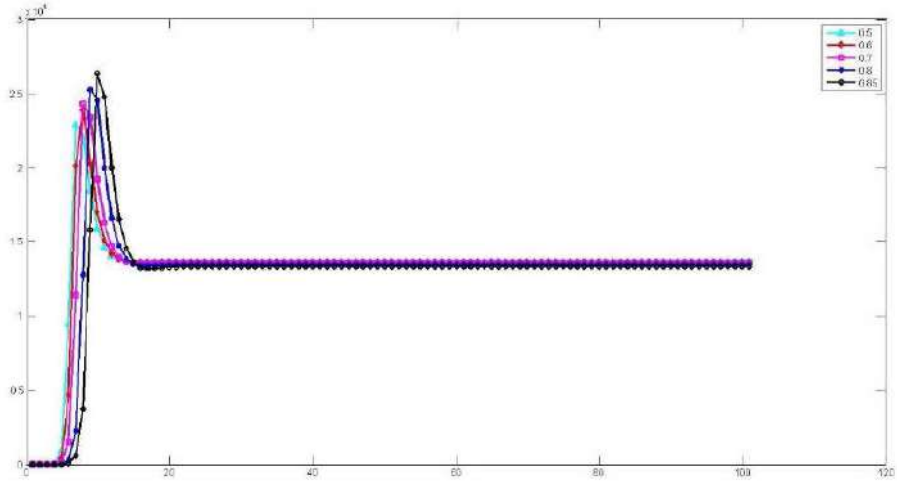


Figure 6. Time series of Infected human on Education Campaign rate ε =0.5, 0.6, 0.7, 0.8 and 0.85, respectively. The values of parameters are in the text. The solutions approach the endemic equilibrium state to the disease-free state.

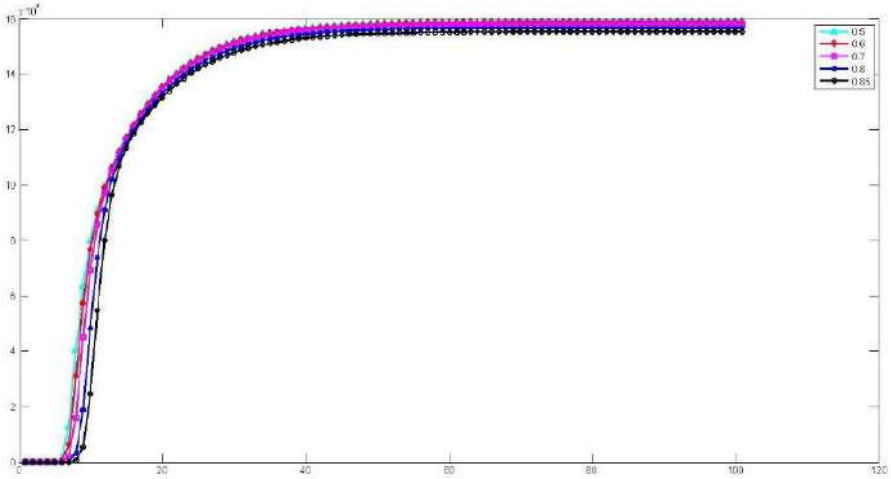


Figure 7. Time series of Recovery human on Education Campaign rate ε =0.5, 0.6, 0.7, 0.8 and 0.85, respectively. The values of parameters are in the text. The solutions approach the endemic equilibrium state to the disease-free state.

This study found that the Education Campaign rate is one of the factors affecting the dynamics of a mathematical model SEIR for controlling the spread of Rotavirus infection. It was explored that if the population at risk of infection with Rotavirus infection knows the prevention of Rotavirus infection less will contribute to the spread of the disease increases. If the most population at risk of infection with Rotavirus infection knows the prevention of Rotavirus infection, then it will contribute to the spread of the decreased disease until there is no further spread of Rotavirus infection. The population at risk of infection with knowledge about the prevention of Rotavirus infection, is not less than 40 percent of the total population, and will contribute to the spread of the disease until there is no further spread of infection.

Discussion

In this study, we proposed the dynamics model of Rotavirus infection by considering the education campaign. We analyzed the model by a standard method in which we determined equilibrium points and investigated the stability of the model. The basic reproductive number is obtained through the next generation method. The Education Campaign rate is the factor affecting the mathematical modelling. In epidemiology, the basic reproductive number is the number of secondary cases generated by a primary infectious cause for the mathematic model, the basic reproductive number is the threshold parameter for determining the stability of the model at each equilibrium point. The stability of the system is investigated using the Roth-Hurwitz criteria. The qualitative behaviours of this model are shown in Fig. 3. We found that when Education Campaign rate is 0.4, 0.5, 0.6, 0.7, 0.8 and 0.85, respectively. The population at risk of infection with knowledge about the prevention of Rotavirus infection is not less than 40 percent of the total population will contribute to the spread of the disease until there is no further spread of infection.

Conclusion

It concluded that when the value of the education campaign rate decreases, the number of infected human increase. The basic reproductive number is greater than one, meaning the Rotavirus infection will occur in the community. On the other hand, when the value of the education campaign increase, the number of infected human decrease. The basic reproductive number is less than one, meaning that the Rotavirus infection will have died out the community.

Acknowledgments
The authors are grateful to the Department of Mathematics, Faculty of Science and Technology, Phuket Rajabhat University, for providing the facilities to carry out the research.

References

[1] Anderson, R.M., and May, R.M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press.
[2] Biophysics Group. (2009). *Mathermatics Model of Transmission*. Faculty of science, Mahidol University.
[3] Bureau of Epidemiology. (2020). *Diarrhea*. (Online).http://rnnvw.boe.moph.go.th /facchickenpox, 13 August 2020.
[4] Fred Brauer, Pauline den Driessche and Jianhong Wu (Eds.). (2008). *Mathematical Epidemiology*. Vancouver, B.C. V6T 1Z2, Canada: Springer-Verlag Berlin Heidelberg.
[5] Kribs-Zaleta, C.M. and Valesco-Hernández, J.X. (2000). *A simple vaccination model with multiple endemic states*. Mathematical Biosciences, 164 (2): 183–201.
[6] Kermack, W. O. and McKendrick, A. G. (1927). *A Contribution to the Mathematical Theory of Epidemics*. Proc. Roy. Soc. Lond. A 115: 700-721.
[7] L. Prihutami. (2009). *Stability Analysis of Tuberculosis Transmission Model*, Thesis, Diponegoro University, Semarang.
[8] Van den Driessche, P. and J. Watmough. (2002). *Reproductive numbers and sub-threshold endemic equilibria for compartment models of disease transmission*. Math. Biosci, 180: 29-48.

Blood glucose biosensor based on glucose oxidase co-immobilized with copper (I) oxide at manganese (IV) oxide on graphene quantum dot in chitosan scaffold

Onanan Injumnong[1], Anchana Preechaworapun[2], Tanin Tangkuarum[1,*]
[1]Chemistry program, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand
[2]Chemistry program, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok 65000, Thailand

**Corresponding Author Email: tanin.tang@gmail.com*

Abstract. The fabrication of glucose biosensorfor real blood based on glucose oxidase (GOx) co-immobilized with copper (I) oxide at manganese (IV) oxide ($Cu_2O@MnO_2$) on the graphene quantum dot (GQD) in chitosan scaffold (CHIT) on the glassy carbon electrode (GCE) (GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE) was developed. The cyclic voltammetric and amperometric techniques were used for the current detection from the produced H_2O_2. The optimized parameters as applied potential, GOx loading, GQD content, and $Cu_2O@MnO_2$ were found at +0.80 V, 15 units, 300 µg, and 300 µg, respectively. The GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE biosensor showed 2 linear ranges of 0.30 mM – 1.0 mM ($y = 201.72x - 8.4483$, $R^2 = 0.9991$) and 1.0 mM – 50.0 mM ($y = 113.39x + 66.306$, $R^2 = 0.9955$) with a limit of detection (S/N=3) at 33 µM. This GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE biosensor was applied to detect glucose in real blood samples and it was validated with the commercial glucose biosensor.

Keywords: glucose biosensor; graphene quantum dot; copper (I) oxide at manganese (IV) oxide

1. Introduction

Diabetes is a disease occurring in the human organs that do not produce enough insulin. Alternatively, it happens in the body that cannot use insulin to regulate blood sugar. Hyperglycaemia is an effect of uncontrolled diabetes and leads to serious damage to body systems such as the nerves and blood vessels. The glucose in the blood is a major substance that causes diabetes and it needs a detection device that is small in size, simple, and quick analysis. Biosensors are one of the devices that fulfill the requirement and they have been continuous increasingly developed and improved to have better performance using different types of nanomaterials. The graphene quantum dot (GQD) is one type of carbon molecule that possesses increasing electron transfer properties [1]. There is much research on utilizing it in photovoltaics, organic light-emitting diodes, fuel cells, bioimaging, biosensing, oligonucleotides, environmental monitoring, and thermal interface materials [2]. The copper (I) oxide at manganese (IV) oxide ($Cu_2O@MnO_2$) is one type of oxygen combinate bimetallic that utilizes electrocatalytic response to the oxidation of generated hydrogen peroxide (H_2O_2) from the biosensing system [3]. Chitosan (CHIT) is one type of polysaccharide that can cross-link with a glutaraldehyde appearance scaffold. It has strong toughness, good adsorbent, high

affinity in aqueous, excellent composite with nanomaterial, and enhanced electrode lifetime usage [4]. Glucose oxidase (GOx) is a low-cost oxidoreductase enzyme that catalyzes glucose and oxygen, obtaining H_2O_2 that makes it widely used as a generic enzyme for glucose biosensors [5]. This research aimed to fabricate an amperometric glucose biosensor using GOx co-immobilized $Cu_2O@MnO_2$ on GQD in CHIT scaffold on the glassy carbon electrode (GCE) (GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE). The conceptual model of GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE is shown in Figure 1. GOx is introduced in CHIT to catalyze the oxidation of glucose in the presence of oxygen and gave gluconolactone and H_2O_2. The H_2O_2 is catalyzed by $Cu_2O@MnO_2$ in oxidized form and changes into the reduced form promptly giving two electrons and enhancing transfer to the electrode by the GQD. The measured current is proportional to the glucose concentration.

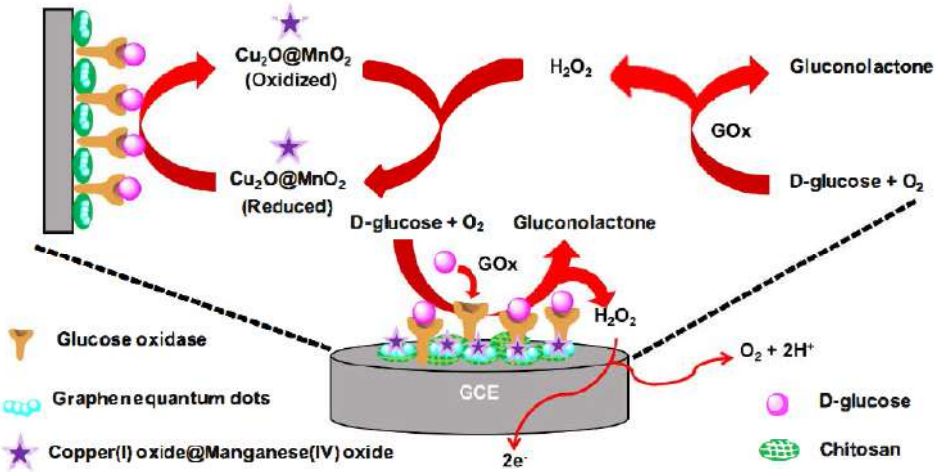


Figure 1. The conceptual model of GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE.

2. Methodology

2.1 Materials and apparatus

The citric acid (99%) was provided from Ajex Finechem (NSW, Australia), glutaraldehyde (GA, 50.0%), sodium hydroxide (NaOH, 97.0%), potassium permanganate ($KMnO_4$, 99.0%), manganese (II) sulfate ($MnSO_4$, 98.0%), copper (II) sulfate pentahydrate ($CuSO_4$ $5H_2O$, 98.0%) and glucose oxidase (117 U/mg) were supplied from Sigma-Aldrich, USA. Sodium borohydride (97%) was provided by Loba Chemie Co. Ltd, Thailand. Cyclic voltammetry and amperometry were operated with CHI1230A (CH instrument, Inc., USA) controlled by a personal computer. Three electrodes system comprised of a bare GCE (CH instrument, Ø 3 mm) or the modified GCE as working electrodes, a platinum wire electrode (CH instrument, Ø 0.5 mm) as a counter electrode, and an Ag/AgCl electrode (3 M KCl internal solution, CH instrument) as a reference electrode. All solutions used in this study were prepared with 18.2 MΩ-cm ultrapure water (ZMQ50007, Milli-Q Academic, USA). The operating condition was set up at room temperature in 0.05 M phosphate buffer solution (PBS).

2.2 Synthesis of graphene quantum dot (GQD)

2.0 grams of citric acid was heated at 200 °C. It was melted and converted into pale yellow within 5 minutes. It turned to orange color within 30 minutes and then it was added dropwise into 50 mL of 10 mg mL^{-1} NaOH solution at constant stirring for 15 minutes, GQD was obtained. The GQD solution was neutralized with 3 M NaOH and stored in a refrigerator until use. [6]

2.3 Synthesis of copper (I) oxide at manganese (IV) oxide ($Cu_2O@MnO_2$)

A 0.32 M $KMnO_4$ solution was prepared and then $CuSO_4$ $5H_2O$ was added into 100 mL of the 0.32 M $KMnO_4$ solution (at a stoichiometric ratio of Cu:Mn as 1:5), noted as solution A. 50 mL

of a 0.88 M $MnSO_4$ solution was prepared and then 8.5 mL HNO_3 was added, noted as solution B. Solution B was added drop-wise into solution A under continuous stirring resulting in the formation of a dark precipitate. The resultant slurry was refluxed at 120 °C for 12 hours, then washed with ultrapure water and dried at 120 °C for 12 hours. Finally, the obtained $Cu_2O@MnO_2$ powder was kept in a desiccator at room temperature for further work. [7]

2.4 Electrode modification

The GCE was polished on a cloth polishing pad with an alumina slurry diameter 0.05 µm. 0.5% CHIT solution was prepared in acetic acid for dissolution and then adjusted the CHIT solution to pH 5 by 0.1 M NaOH. Glutaraldehyde (GA) was cross-linked to CHIT by well mixed, obtaining a solid in solution. It was then kept sonication for 12 hours until the solid in the solution melted to liquid, and the cross-linked of GA-CHIT was obtained. Weigh 4 mg of $Cu_2O@MnO_2$, pipet 0.5 mL of GQD, and 0.5 mL of GA-CHIT were mixed up and sonicated for 3 hours, and the CHIT-GQD@$Cu_2O@MnO_2$ slurry was achieved. The obtained slurry was dropped on the carbon grid for characterization by the transmission electron microscope (TEM, JED-2300, USA) and energy-dispersive x-ray spectroscopy (EDX, JED-2300, USA). The electrode was modified by drop-casting 20 µL of CHIT-GQD@$Cu_2O@MnO_2$ on the GCE surface and dried at room temperature. Then, it was dropped with 20 µL of 3 mg mL^{-1} GOx and dried at room temperature in the air. After that, the modified GOx/CHIT-GQD@$Cu_2O@MnO_2$/GCE was rinsed thoroughly with water before use. Other modified GCEs were prepared in the same manner but without the unwanted substances.

3. Results and discussion

3.1 Characterization of synthesized materials

The CHIT-GQD@$Cu_2O@MnO_2$ morphology and elemental composition were confirmed by TEM and EDX, result as in figure 2A. With the ability of TEM mapping features from the TEM technique, it shows carbon (C) (Figure 2B), oxygen (O) (Figure 2C), copper (Cu) (Figure 2D), and manganese (Mn) (Figure 2E) which each element appears from the precursor substances. TEM zoom-in image (Figure 2F) shows it looks like a nanorod. The semiqualitative composition of the particles was examined by the EDX in Figure 2G. There are the presence of C, O, Mn, and Cu for 64.36%, 25.80%, 3.43%, and 3.43% by mass, respectively. These results are corresponding to the chemical used to synthesize CHIT-GQD@$Cu_2O@MnO_2$.

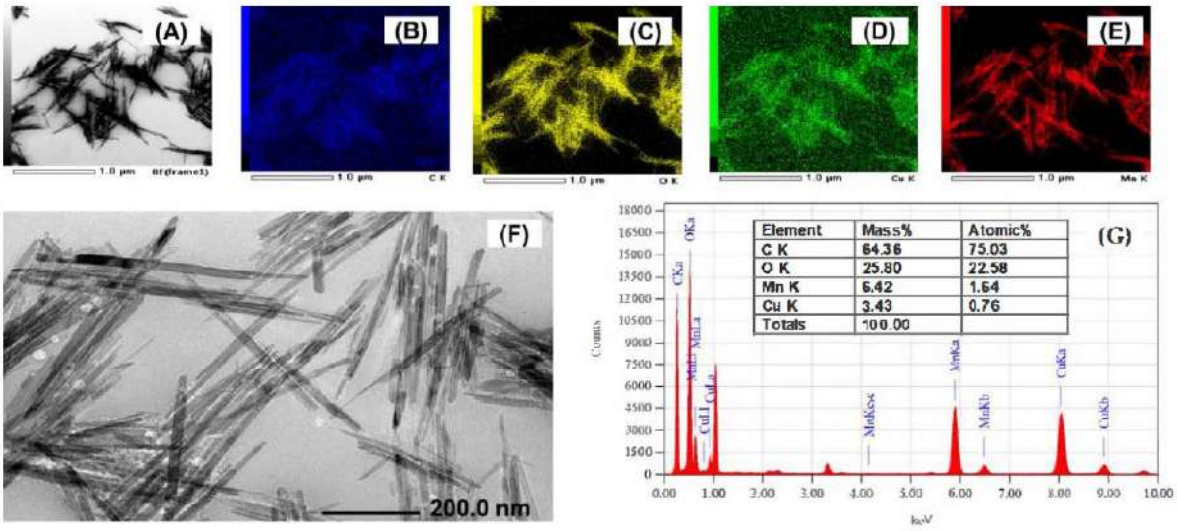


Figure 2. TEM images of CHIT-GQD@$Cu_2O@MnO_2$ (A) and zoom in (F). The TEM mapping shows C (B), O (C), Cu (D), and Mn (E). EDX spectrum (G) shows the element contents.

3.2 Electrochemical behavior of modified electrodes

Modified electrodes were studied using cyclic voltammetry and it showed cyclic voltammograms in Figure 3. They were operated without (a) and with 1.0 mM glucose (b) in 0.05 M PBS pH 7.4, and in the potential range from -0.4 to +1.0 V. Cyclic voltammograms of seven different kinds of modified and unmodified electrodes were bare GCE (Figure 3A), CHIT/GCE (Figure 3B), GOx/CHIT/GCE (Figure 3C), GOx/GCE (Figure 3D), GOx/CHIT-GQD/GCE (Figure 3E), GOx/CHIT-Cu_2O@MnO_2/GCE (Figure 3F) and GOx/CHIT-GQD@Cu_2O@MnO_2/GCE (Figure 3G). The bare GCE (Figure 3A) and the CHIT/GCE (Figure 3B) were not active to the glucose while modified with GOx/CHIT (Figure 3C) and GOx (Figure 3D) showed glucose active at the +0.90 V. The effect of GOx combination with CHIT-GQD (Figure 3E) showed oxidation peak at +0.65 V in a glucose solution, this meaning that the GQD does enhance the catalytic oxidation of generated H_2O_2. In the addition of Cu_2O@MnO_2 into GOx/CHIT (Figure 3F), the oxidation peak of generated H_2O_2 was shifted to +0.80 V and a higher oxidative current appeared. The addition of GQD as CHIT-GQD-Cu_2O@MnO_2 (Figure 3G) shows the highest oxidative currents at 10.83 μA with the peak shifted to +0.74 V. From these results, the GQD serves to increase the electron transfer property while Cu_2O@MnO_2 composite materials do a synergetic effect.

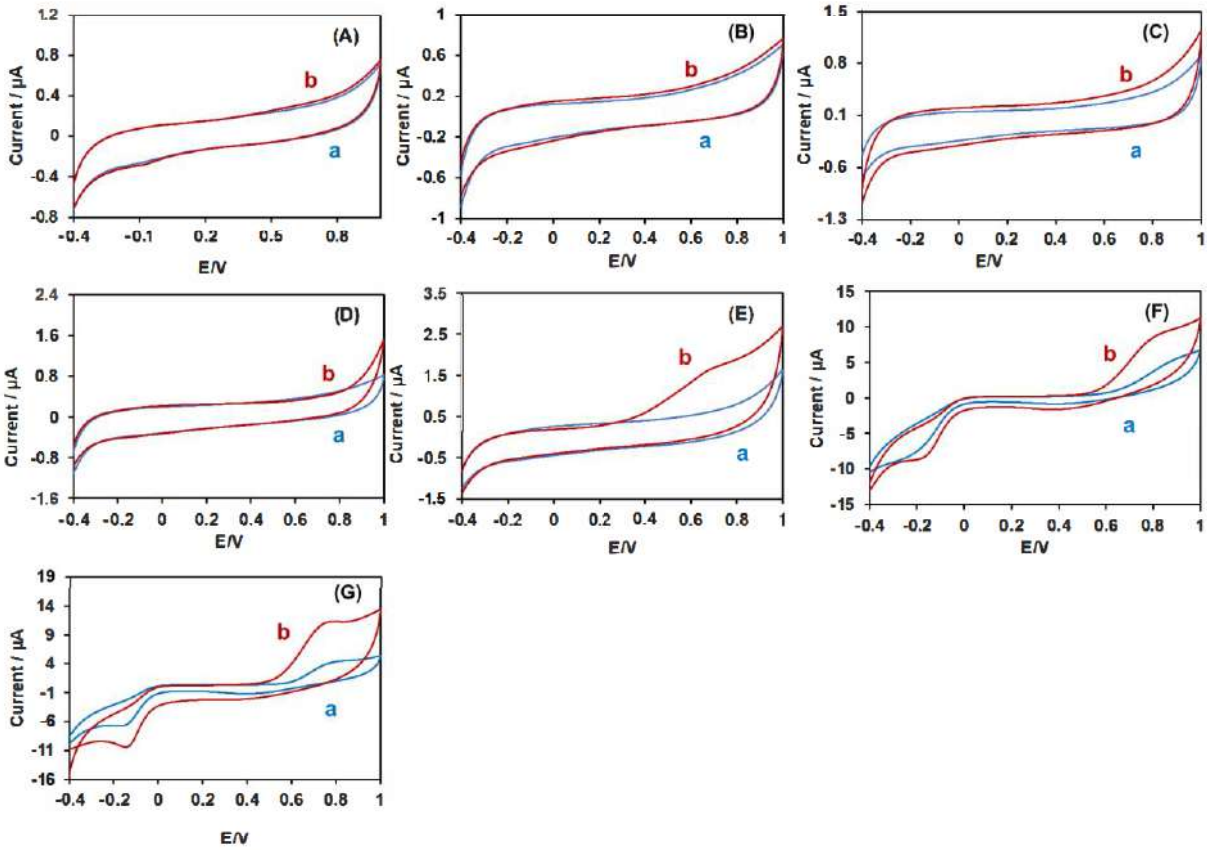


Figure 3. Cyclic voltammograms of the bare GCE (A), CHIT/GCE (B), GOx/CHIT/GCE (C), GOx/GCE (D), GOx/CHIT-GQD/GCE (E), GOx/CHIT-Cu_2O@MnO_2/GCE (F), and GOx/CHIT-GQD@Cu_2O@MnO_2/GCE (G) in 0.05 M PBS pH 7.4 containing without (a) and with 1.0 mM glucose (b) at a scan rate of 50 mV s^{-1}.

3.3 Optimization of experimental parameters

The optimization of the developed GOx/CHIT-GQD@Cu_2O@MnO_2/GCE biosensor for glucose detection was done for obtaining the highest performance. The study was done by spiking 1.0 mM glucose into stirred PBS and detected by the amperometry. The applied potential was done in the range of +0.50 to +0.90 V, the obtained current was plotted versus the corresponding applied potential as in Figure 4A. The highest current response is obtained at +0.80 V compared to the other potential. It was chosen at +0.80 V for further work. Figure 4B showed the effect of enzyme loading in GOx/CHIT-GQD@Cu_2O@MnO_2/GCE modified electrode. Various GOx units were varied from 0 units to 20 units. At 0 units, it is no current occurred. While increasing enzyme units, the current keep increasing and saturating at 15 units, therefore it is chosen for further work. Figure 4C is showed the effect of Cu_2O@MnO_2, it is varied from 0 to 500 µg. The current is steeply increasing and is highest at 300 µg, this point is chosen for further work. The GQD effect is shown in Figure 4D, the highest current is obtained at 300 µg and it is selected at this amount. Figure 4E showed the variation of pH solution (0.05 M PBS) from pH 6.2 to 8.2. The results showed that the current response linearly increased at pH from 6.2 to 7.4. At pH 6.2, the current is a little appearance. Increasing the pH to higher, the current is increasing and it is much high at a pH 7.4. Since the pH affects the active site of the enzyme, from this point the acidic pH affects more denature enzymes than the basic pH. Therefore, a 0.05 M PBS solution of pH 7.4 is suitable for this experiment.

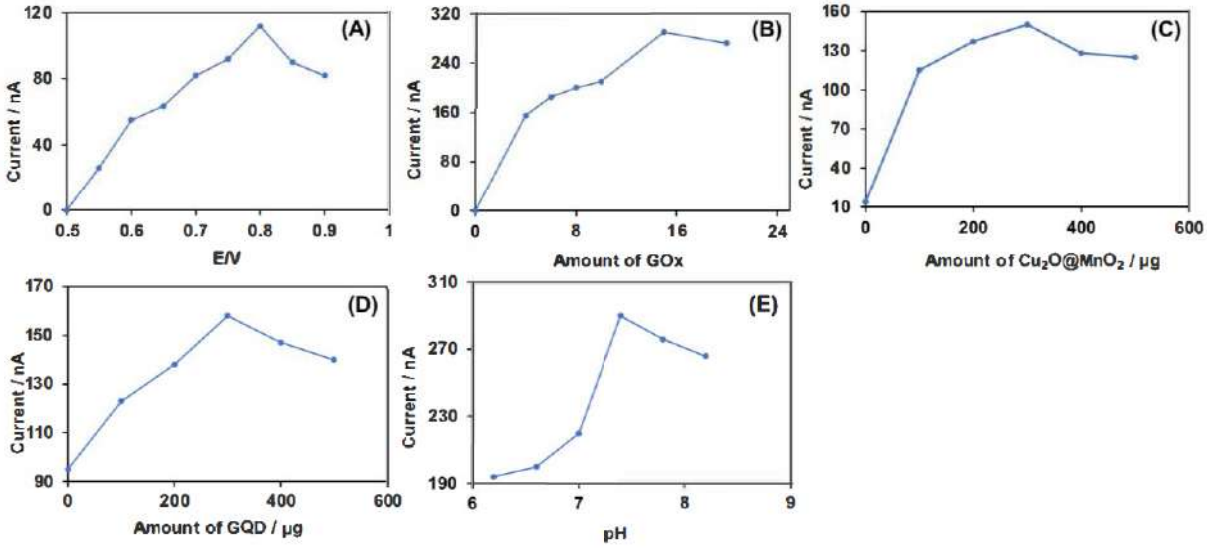


Figure 4. Optimization of GOx/CHIT-GQD@Cu_2O@MnO_2/GCE parameter as applied potential (A), amount of GOx (B), amount of Cu_2O@MnO_2 (C), amount of GQD (D), and the pH (E).

3.4 Linearity

The developed GOx/CHIT-GQD@Cu_2O@MnO_2/GCE was tested by spiking glucose into stirred PBS for linearity under the optimal condition (applied potential +0.80 V, 15 units of GOx, 300 µg of Cu_2O@MnO_2, 300 µg of GQD and 0.05 M PBS pH 7.4). Figure 5A shows the current of the modified electrode corresponding to the glucose concentration. It shows two linear ranges of 0.3 to 1.0 mM (R^2 = 0.9991) and 1.0 to 50.0 mM (R^2 = 0.9955). At the low concentration linear range (Figure 5B), the limit of detection is calculated at S/N=3 and found to be 33 µM. Both linearity and limit of detection are good to support the detection ability for diabetes diseases in whole blood samples.

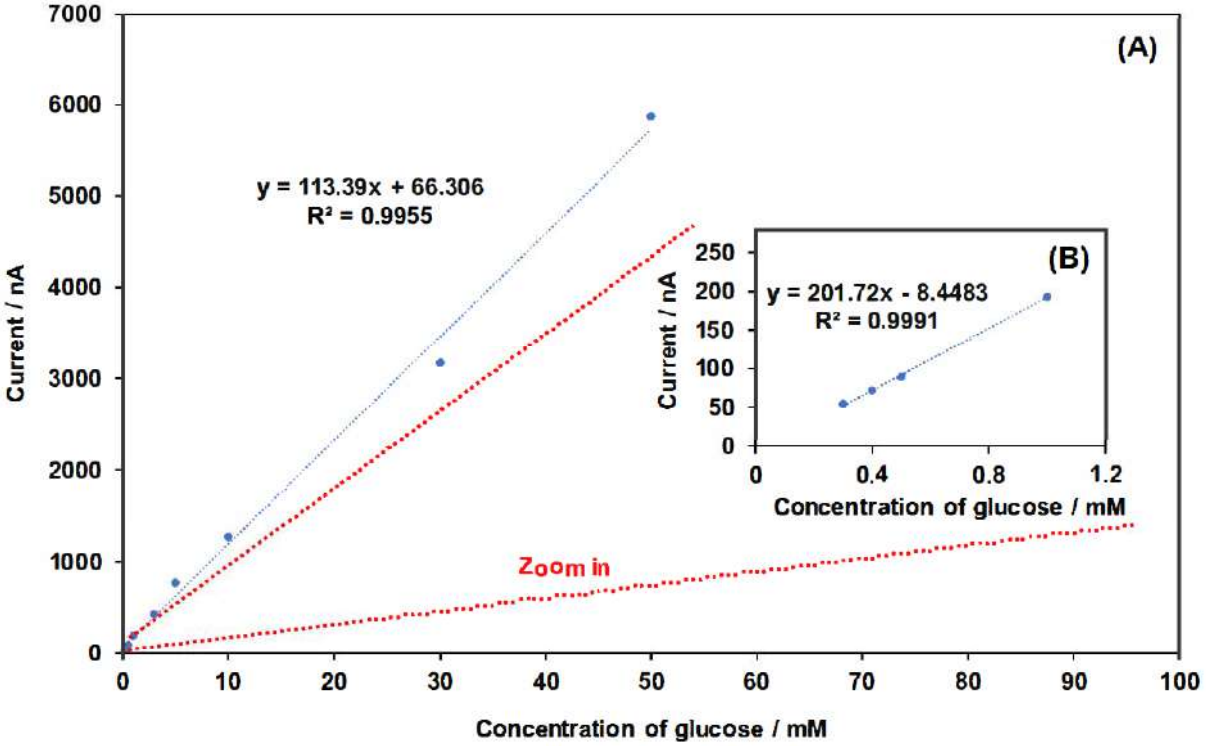


Figure 5. Calibration curve between the response current of the modified electrode and glucose concentration from 1.0 to 50.0 mM (A) and zoom in from 0.3 to 1.0 mM (B).

3.5 Sample analysis

The GOx/CHIT-GQD@Cu_2O@MnO_2/GCE was successfully applied for the detection of glucose in 5 human blood samples from Sansai hospital, Chiang Mai, Thailand. The standard addition method was applied to examine glucose. The obtained results were compared to the results from commercial glucose biosensors as shown in Table 1. The calculated T-test (column 4) and F-test (column 5) were less than the critical T-test and F-test at a 95% confidence interval, respectively. This indicated that both techniques were not significantly different at a confidence level of 95% to statistical F-test and T-test values. The performance of developed biosensors was compared to earlier reports as in table 2. We found that our developed biosensor has a comparative performance to the previous reports.

Table 1. Glucose detection in human blood samples.

Sample	Developed glucose biosensor (mM)	Commercial glucose biosensor (mM)	T-test* (95% confidence limit)		F-test (95% confidence limit)	
			Crit	Cal	Crit	Cal
Sample 1	5.7	7.0	2.78	0.76	6.39	4.07
Sample 2	6.0	7.2				
Sample 3	4.8	6.3				
Sample 4	9.6	9.2				
Sample 5	9.4	8.0				

* Cal = Calculated, Crit= Critical

Table 2. Comparison of glucose biosensor.

Biosensor*	Detection techiques	LOD/mM	Linearity	[Refer]
GOx-SiO_2/Lig/CPE	Cyclic voltammetry	0.145	0.5 – 9.0 mM	[8]
GOx/GA/CP/GCE	Cyclic voltammetry	0.200	1.0 – 25.0 mM	[9]
MWCNT/GO/GOx	Amperometry	0.028	0.05 – 23.2 mM	[10]
Ppy/GOx/AuNPs/GR	Amperometry	0.200	up to 19.9 mM	[11]
AuNPs-GOx-MWCNTs-PVA/GCE	Amperometry	0.200	0.5 – 8.0 mM	[12]
GNP/MWCNTs/GOx	Amperometry	0.128	Up to 9.0 mM	[13]
GR-MWNTs/AuNPs/GCE	Amperometry	0.950	2.0 – 5.2 mM	[14]
GOx/CHIT-GQD@Cu_2O@MnO_2/GCE	Amperometry	0.033	0.30 – 50.0 mM	This work

* AuNPs = Gold nanoparticles, CHIT = Chitosan, CP = Chalcopyrite, Cu_2O@MnO_2 = Copper (I) oxide at manganese (IV) oxide, GA = Glutaraldehyde, GCE = Glassy carbon electrode, GNP = Gold nanoparticles, GO = Graphene oxide, GOx = Glucose oxidase, GQD = Graphene quantum dot, GR = Graphite rod, SiO_2 = Silica, Lig = Lignin, MWCNTs = Multi-walled carbon nanotubes, Ppy = Polypyrole, PVA = Polyvinyl alcohol

4. Conclusions

This work successfully synthesized the copper (I) oxide at manganese (IV) oxide on graphene quantum dots (GQD@Cu_2O@MnO_2) and used in glucose biosensors application. It showed a nanorod shape containing carbon (64.36%), oxygen (25.80%), manganese (3.43%), and copper (3.43%). The developed glucose biosensor based on glucose oxidase co-immobilized synthesized material in chitosan scaffold (GOx/CHIT-GQD@Cu_2O@MnO_2/GCE) was completely developed. It showed wide linearity from 0.30 mM – 50.0 mM with a low detection limit (S/N=3) of 33 μM. It also showed quick analysis, simple operation, and the ability to detect glucose in real blood samples.

Acknowledgments
The authors would like to thank the Sansai hospital for donating real blood samples, and also Dr. Tik Ouiram, NANOTEC, Thailand for TEM images.

References

[1] Tabish, T A and Zhang, S 2016 *Graphene quantum dots: syntheses, properties, and biological applications* (Oxford: Academic Press). pp 171-192
[2] Bacon, M, Bradley, S J and Nann, T 2014 *Particle and Particle Systems Characterization* 31(4) pp 415-428
[3] Ouiram, T, Moonla, C, Preechaworapun, A, Muangpil, S, Maneeprakorn, W and Tangkuaram, T 2021 *Electroanalysis* 33(2) pp 455-463
[4] Moonla, C, Nontapha, C, Ouiram, T, Preechaworapun, A and Tangkuaram, T 2019 *Electroanalysis* 31(8) pp 1588-1597
[5] Wong, C M, Wong, K H and Chen, X D 2008 *Applied Microbiology and Biotechnology* 78(6) pp 927-938
[6] Trinadh, T, Khuntia, H, Anusha, T, Bhavani, K S, Kumar, J V S and Brahman, P K 2020 *Diamond and Related Materials* 110
[7] Ouiram, T, Moonla, C, Preechaworapun, A and Tangkuaram, T 2019 *Electroanalysis* 31(7) pp 1356-1362
[8] Jędrzak, A, Rębiś, T, Klapiszewski, Ł, Zdarta, J, Milczarek, G and Jesionowski, T 2018 *Sensors and Actuators B: Chemical* 256 pp 176-185
[9] Wang, Y, Zhao, J, Yang, T, Zhang, Y, Tao, D, Hasebe, Y and Zhang, Z 2021 *Journal of Electroanalytical Chemistry* 894
[10] Palanisamy, S, Cheemalapati, S and Chen, S M 2014 *Materials Science and Engineering C* 34 pp 207-213
[11] German, N, Ramanavicius, A and Ramanaviciene, A 2017 *Electroanalysis* 29(5) pp 1267-1277
[12] Zhang, H, Meng, Z, Wang, Q and Zheng, J 2011 *Sensors and Actuators B: Chemical* 158(1) pp 23-27
[13] Liu, Y, Wu, S, Ju, H and Xu, L 2007 *Electroanalysis* 19(9) pp 986-992
[14] Devasenathipathy, R, Mani, V, Chen, S M, Huang, S T, Huang, T T, Lin, C M, Hwa, K Y, Chen, T Y and Chen, B J 2015 *Enzyme and Microbial Technology* 78 pp 40-45

Estimation of population size based on zero-truncated, one-inflated and covariate information

Tita Jongsomjit[1,*], Rattana Lerdsuwansri[1], Krisana Lanumteang[2]
[1] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand
[2] Department of Statistics, Faculty of Science, Maejo University, Thailand

**Corresponding Author Email: tt.jsj@hotmail.com*

Abstract. To estimate the unknown size of the population that is difficult or hidden to enumerate, the capture-recapture method is widely used for this purpose. We propose the one-inflated, zero-truncated geometric (OIZTG) model to deal with three important characteristics of some capture–recapture data: missing zero counts, excess ones, and observed heterogeneity. The OIZTG is generated by two distinct processes, one from a zero-truncated geometric (ZTG) process, and the other one-count producing process. To explain heterogeneity at an individual level, the OIZTG provides a simple way to link the covariate information. A simulation study shows that there are differences among population size estimates. The OIZTG gives asymptotic unbiased estimator whereas existing population size estimators seem to be quite biased. Finally, we applied the OIZTG to estimate the number of heroin users in Chiang Mai from 2013 to 2018, and the likelihood ratio test was examined for the presence of one-inflation.

Keywords: Capture-recapture; Zero-truncation; One-inflation; Observed heterogeneity; Geometric regression.

1. Introduction

A capture-recapture approach aims to estimate an elusive target population size N. This approach assumes that the target population is closed with no births, deaths, or migration over the observational period. Some identification mechanisms, such as registration or trapping systems, are used to identify observed units. A unit may be identified exactly once, or it may be observed twice, three times, or more. We denote the number of units observed y times by f_y and $n = f_1 + f_2 + \cdots + f_m$ is the observed sample size where m is the largest observed count. The number of unobserved or missing units , f_0, needs to be estimated in order to get the estimates of population size, $\hat{N} = \hat{f}_0 + n$.

In the following we illustrate the situation that is estimating the number of heroin users in Chiang Mai, Thailand from 2013 to 2018. The count variable of interest is the number of occasions that a specific heroin user contacted the treatment centers y times presented in

Table 1. Moreover, individual information such as gender will also be collected. From Table 1, a total of 843 heroin users were observed, 537 were treated once, 152 twice, 80 three times

Table 1. Frequencies of heroin users in Chiang Mai, Thailand from 2013 to 2018.

Group	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	Total
Male	482	134	73	30	13	7	5	7	0	1	1	0	0	1	**754**
Female	55	18	7	4	2	1	1	1	0	0	0	0	0	0	**89**
Total	**537**	**152**	**80**	**34**	**15**	**8**	**6**	**8**	**0**	**1**	**1**	**0**	**0**	**1**	**843**

and so on. As can be seen, there is an unobserved zero count as well as a huge number of ones.

So far, two modeling solutions for one-inflation have been proposed. Godwin and Böhning (2017) [1] add an excess probability of observing one counts in the positive Poisson (PP) distribution and propose the one-inflated positive Poisson (OIPP) distribution. In order to allow unobserved heterogeneity as well, Godwin (2017) [2] proposes the one-inflated, zero-truncated negative binomial (OIZTNB) model. Since the OIZTNB suffers the boundary problem implying the high estimate for the population size, geometric model may serve as an alternative model. The advantage of the geometric distribution is that it is a mixture of a Poisson with an exponential distribution, thus it incorporates some form of heterogeneity.

In this paper, we propose the one-inflated, zero-truncated geometric (OIZTG) model and a simple way to link covariate information to estimate an unknown population size.

2. The one-inflated, zero-truncated geometric model

The one-inflated and zero-truncated geometric model (OIZTG) arises as a generalization of the zero-truncated geometric model (ZTG). The data might be thought to be generated by two distinct processes, one from a ZTG process and the other from one-count producing process.

The probability mass function for a random variable Y that follows a truncated geometric distribution is

$$P^{ZTG}(Y_i = y_i) = (1-\theta_i)^{y_i-1}\theta_i \; ; y_i = 1,2,\ldots$$

where θ_i, $0 < \theta_i < 1$, is the geometric parameter. The OIZTG distribution is obtained by adding an excess probability of observing "one counts", w, in the ZTG distribution and we assume w to be the same for all samples (i.e. $w_i \equiv w$, for all $i, i = 1,2,\ldots n$):

$$P^{OIZTG}(Y_i = y_i) = \begin{cases} w + (1-w)\theta_i & ; y_i = 1 \\ (1-w)(1-\theta_i)^{y_i-1}\theta_i & ; y_i = 2,3,\ldots \end{cases}$$

The mean of this distribution is $w + (1-w)(1/\theta_i)$ and the variance is $w + (1-w)\left[\left(2/\theta_i^{\,2}\right) + (1/\theta_i)\right] - [w + (1-w)(1/\theta_i)]^2$.

Reparameterization

In a regression model framework, typically the mean of the response variable is modeled as a linear function of the predictors. The convention has incorporated covariates into the truncated models in the same manner as untruncated counterparts, and this convention is upheld for the OIZTG model. To obtain a regression model for the mean of the OIZTG

distribution, we need to re-parameterized the OIZTG distribution by letting $E(Y_i) = \mu_i = 1/\theta_i$. Hence, we have

$$\theta_i = \frac{1}{\mu_i}$$

where $\mu_i = exp\,(\beta^T x_i)$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ is a matrix of coefficients and x_i is a vector of covariate values for subject i or $x_i = (1, x_{i1}, \dots, x_{ip})^T$. The OIZTG distribution can then be written in term of new parameterization as follows:

$$P^{OIZTG}(Y_i = y_i) = \begin{cases} w + (1-w)\left(\frac{1}{\mu_i}\right) & ; y_i = 1 \\ (1-w)\left(1 - \frac{1}{\mu_i}\right)^{y_i - 1}\left(\frac{1}{\mu_i}\right) & ; y_i = 2,3, \dots \end{cases} \quad (1)$$

For reparameterizing the inflation parameter, if one-deflation is discarded, then the logistic link $w = 1/(1 + exp[-\varphi])$ may be used and φ may be estimated without limited range, ensuring that $0 < w < 1$ and that (1) is a proper mass function. However, if $w = \varphi/(1 + \varphi)$, then $-\infty < w < 1$ for $-1 < \varphi < \infty$. Since we focus on the case of 1-inflation, the link $w = 1/(1 + exp[-\varphi])$ wil be used throughout the paper.

Maximum likelihood estimation of the OIZTG model

According to (1) , the likelihood function and log-likelihood function are given as

$$L(w, \mu_i) = \prod_{i=1}^{n}\left[w + (1-w)\left(\frac{1}{\mu_i}\right)\right]^{I_1(y_i)}\left[(1-w)\left(1 - \frac{1}{\mu_i}\right)^{y_i - 1}\left(\frac{1}{\mu_i}\right)\right]^{1 - I_1(y_i)}$$

and

$$logL(w, \mu_i) = \sum_{i=1}^{n}\left\{I_1(y_i) log\left[w + (1-w)\left(\frac{1}{\mu_i}\right)\right] + (1 - I_1(y_i))[log(1-w) + (y_i - 1)log(\mu_i - 1) - y_i log\mu_i]\right\}, \quad (2)$$

respectively, where $I_1(y_i) = 1$ if $y_i = 1$and $I_1(y_i) = 0$ if $y_i > 1$. The first derivatives of the log-likelihood are given by

$$\frac{\partial logL(w,\mu_i)}{\partial \beta_r} = \sum_{i=1}^{n}\left\{I_1(y_i)\left[\frac{(1-w)\left(-X_{ir}\left(\frac{1}{\mu_i}\right)\right)}{w + (1-w)\left(\frac{1}{\mu_i}\right)}\right] + (1 - I_1(y_i))\left[\frac{(y_i - 1)(X_{ir}\mu_i)}{(\mu_i - 1)} - \frac{y_i X_{ir}\mu_i}{\mu_i}\right]\right\}, \quad (3)$$

$r = 1,2, \dots, p$, and

$$\frac{\partial logL(w,\mu_i)}{\partial w} = \sum_{i=1}^{n}\left\{I_1(y_i)\left[\frac{1 - \frac{1}{\mu_i}}{w + (1-w)\left(\frac{1}{\mu_i}\right)}\right] + (1 - I_1(y_i))\left[\frac{-1}{1-w}\right]\right\}. \quad (4)$$

Due to their complexity, it is not easy to maximize the log likelihood using these derivatives. Thus, Nelder–Mead algorithm was used to maximize the log likelihood, in the *maxLik* package [3] in R.

Estimating the size of an unknown population via the OIZTG model

We use a Horvitz–Thompson estimator [4] to estimate the population size. In general, the Horvitz–Thompson estimator for the unknown population size is $\hat{N} = n/(1 - p(0, \hat{\theta}))$, where $p(0, \hat{\theta})$ is the estimated probability of a zero occurring under the untruncated count distribution, and $\hat{\theta}$ has been estimated under the zero-truncated count distribution.

Using the OIZTG model as the truncated count distribution, the proposed Horvitz–Thompson estimator for the population size is

$$\hat{N}^{OIZTG} = \frac{n}{1-\frac{1}{\hat{\mu}_i}} \quad (5)$$

where $\hat{\mu}_i$ is the MLE of the OIZTG model.

3. Comparing $\hat{N}^{OIZTG}$ with other estimators

To study the performance of proposed estimator ($\hat{N}^{OIZTG}$) and to compare with those of other well-known estimators. The Chao's lower bound estimator ($\hat{N}^{C} = n + (f_1^2/f_2)$) [5] was used as an estimator in zero-truncated case. The modified Chao estimator ($\hat{N}^{ModC} = n + (f_2^3/f_3^2)$) [6] was used as an estimator in zero- truncated and one-inflated case. The one inflated positive poisson model ($\hat{N}^{OIPP}$) and the one-inflated zero-truncated negative binomial model ($\hat{N}^{OIZTNB}$) were used as estimators in zero-truncated, one-inflated and covariate information case. In turn, each of these estimators will be defined and discussed in terms of their percentage relative bias, %Rbias, and percentage relative mean square error, %RMSE, under an OIZTG data-generating process.

In describing the generation of random numbers from an OIZTG, the assumed data-generating process for the OIZTG is elucidated.

Step 1: generate N counts according to the geometric distribution.

Step 2: remove the 0-counts from the sample, giving a new sample of size n. The number of 0s removed, f_0, is to be estimated, so that an estimate for $N = f_0 + n$ may be obtained.

Step 3: randomly alter wn counts such that $y = 1$. These altered counts represent members of the population who are instead observed only once.

The algorithm was repeated 10,000 times for each of several parameter configurations and sample sizes. In each repetition the various estimates are obtained, and average %Rbias and average %RMSE for the various estimators are shown in Table 2. The average %Rbias of $\hat{N}$ is calculated as $\sum_{l=1}^{M}\{[(\hat{\theta}_l - \theta)/\theta]/10000\} \times 100$ and the average %RMSE of $\hat{N}$ is calculated as $\sum_{l=1}^{M}\left\{\left[(\hat{\theta}_l - \theta)^2/\theta^2\right]/10000\right\} \times 100$ where θ is the true parameter, $\hat{\theta}$ is the corresponding MLE, and l indexes the Monte Carlo replication.

Not surprisingly, when the OIZTG distribution is the data-generating process, $\hat{N}^{OIZTG}$ itself has very little bias. The simulated bias of OIZTG is shown to be approximately 0-1% and all cases %Rbias disappears quickly as N increases, reflecting the consistency of the MLE. In most cases, the alternative estimators are quite biased. The Chao estimator is always too large, by 15–300% while the modified Chao estimator is approximately by 3 to -20%. The Horvitz–Thompson estimators under OIPP is always too small, by −10% to −40%. The OIZTNB
model provides estimates that are too large (0% to 160% Rbias), and can also fail due to the boundary problem in other settings (not reported).

4. Illustrative example

In this section, we consider the previous situation that estimates the number of heroin users in Chiang Mai, Thailand from 2013 to 2018.

Specification testing of the OIZTG model

Look at how to see if the OIZTG model is better than the OIZTNB or ZTG models. Both of the latter models are nested in the former. Specifically, when $\alpha \to 1$ the OIZTNB model collapses to the OIZTG model, and when $w = 0$ the OIZTG model collapses to the ZTG model. The models will be tested using the likelihood ratio test (LRT).

Note that the test of H_0: $\alpha = 1$ versus H_a: $\alpha \neq 1$ may be interpreted as a test of whether the OIZTG model or the OIZTNB model is preferred, and that the test of H_0: $w = 0$ versus H_a: $w \neq 0$ may be interpreted as a test for one-inflation in the ZTG model. It is recognized that

in both tests the null hypotheses are on the boundary of the parameter space, hence the LRT statistics are asymptotically $0.5\chi^2$ distributed.

Table 2. Average %Rbias (top) and %RMSE (bottom) of the various population size estimators over 10000 replications, using the OIZTG distribution as the data generating process.

N	$\theta = 0.1$					$\theta = 0.5$				
	$\hat{N}_C$	$\hat{N}_{ModC}$	$\hat{N}_{OIPP}$	$\hat{N}_{OIZTNB}$	$\hat{N}_{OIZTG}$	$\hat{N}_C$	$\hat{N}_{ModC}$	$\hat{N}_{OIPP}$	$\hat{N}_{OIZTNB}$	$\hat{N}_{OIZTG}$
$w = 0.1$										
500	29.60	0.75	-9.97	0.47	0.01	17.74	3.36	-43.06	19.18	0.68
	9.67	0.53	1.01	0.13	0.03	5.33	10.85	18.64	257.41	0.62
1000	28.68	-0.08	-9.99	0.33	-0.02	17.08	-0.98	-43.21	8.78	0.47
	8.65	0.21	1.01	0.06	0.01	3.97	4.15	18.72	57.75	0.31
2000	28.41	-0.56	-10.00	0.19	-0.02	16.64	-2.87	-43.30	5.39	0.29
	8.28	0.10	1.00	0.03	0.01	3.29	1.88	18.77	8.91	0.16
$w = 0.3$										
500	145.43	-0.76	-9.94	1.14	-0.17	59.76	-4.24	-43.06	142.99	0.89
	223.03	0.58	1.01	0.19	0.03	41.03	10.79	18.64	2556.56	0.79
1000	142.86	-1.74	-9.98	0.79	-0.13	58.42	-9.02	-43.28	84.63	0.26
	209.36	0.21	1.01	0.09	0.01	36.61	4.38	18.78	3286.34	0.38
2000	141.46	-2.22	-9.99	0.53	-0.09	57.75	-10.86	-43.31	44.93	0.19
	202.60	0.12	1.00	0.04	0.01	34.58	2.71	18.78	316.03	0.19
$w = 0.5$										
500	345.10	-2.08	-9.89	1.92	-0.37	114.97	-10.92	-43.21	164.02	0.05
	1260.47	0.51	1.00	0.27	0.03	144.05	11.50	18.78	2328.01	0.87
1000	338.65	-3.15	-9.94	1.39	-0.24	112.96	-15.83	-43.36	93.49	-0.22
	1176.83	0.26	1.00	0.12	0.01	133.07	5.67	18.86	625.16	0.43
2000	333.64	-3.53	-9.98	0.98	-0.18	111.38	-17.75	-43.40	49.94	-0.20
	1127.51	0.19	1.00	0.06	0.01	126.82	4.45	18.86	201.04	0.21

The result shows that the LRT statistic for $H_0: \alpha = 1$ vs. $H_a: \alpha \neq 1$ is -2.66 (p-value = 0.22), indicating that the OIZTG model was more appropriate than the OIZTNB model, and the test for $H_0: w = 0$ vs. $H_a: w \neq 0$ provides strong evidence of one-inflation (p-value = 0.00). In addition, Figure 1 supported that the OIZTG distribution fits better than the other distributions.

Fitting the various models

The data is analysed by using the various population size estimators. Table 3 reports the estimation results. OIPP is always smaller than the alternative estimators. Similarly to the simulation experiments, the values of the estimators maintain the inequality $\hat{N}_{OIZTNB} > \hat{N}_C > \hat{N}_{OIZTG} > \hat{N}_{ModC} > \hat{N}_{OIPP}$, without exception. Under OIZTG, one-inflation is 30%, θ is estimated at 0.48, and the population size is estimated to be approximately 1620 people.

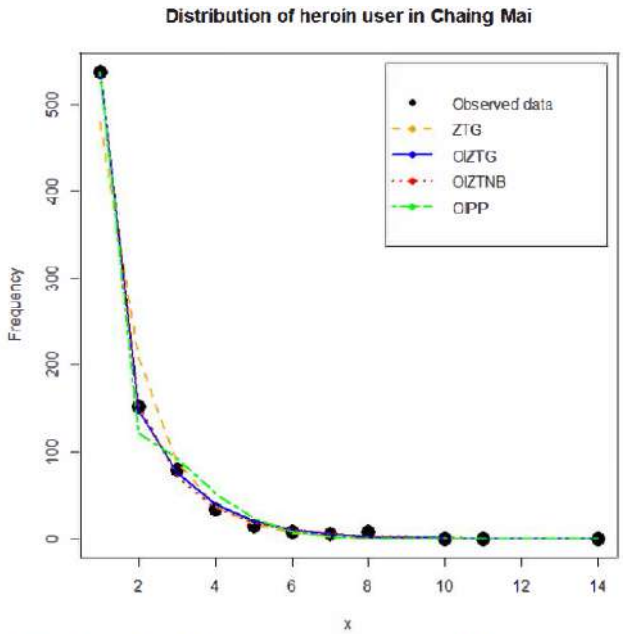


Figure 1. Frequency distribution of heroin users among the observed counts, ZTG, OIZTG, OIZTNB and OIPP distribution.

Table 3. Various population size estimates and statistics for the heroin user data

Model	$\hat{N}$	$\hat{\lambda}$	$\hat{a}$	$\hat{\theta}$	$\hat{w}$
C	2740	-	-	-	-
ModC	1392	-	-	-	-
OIPP	938	2.29	-	-	0.51
OIZTNB	4368	0.37	0.22	-	0.15
OIZTG	1620	-	-	0.48	0.30

5. Conclusion and discussion

The OIZTG model has been proposed as a new method for modeling count data that has been truncated at zero. The OIZTG model was proposed to deal with 1-inflation and observed heterogeneity in a zero-truncated geometric distribution and is used as the truncated count distribution in Horvitz–Thompson estimation of the population size.

Simulations reveal that, when the OIZTG distribution is the data-generating process, conventional estimators can be extremely biased. The Chao estimator relies on the observed frequency of 1-counts, they are extremely biased (15–300%) whereas the modified Chao estimator developed to avoid the overestimation caused by one-inflation is able to reduce bias. In addition, the Horvitz–Thompson estimator via the OIPP distribution provides too small biased. To deal with observed heterogeneity, the Horvitz–Thompson estimator via the OIZTNB model always suffers from the boundary problem in some cases, but the Horvitz–Thompson estimator via the OIZTG distribution is very small biased. As a result, simulation studies imply that if the OIZTG is the data-generating process, using other population size estimators will lead to massive bias.

The OIZTG is illustrated through case studies. One-inflation and observed heterogeneity seem likely. In these cases, the OIZTG fits the data very well and can provide markedly different results from the other population size estimators considered in the paper.

With caution, we have seen that the data-generating process for the OIZTG is not realistic. It can generate random numbers from geometric distribution with one-inflation rate such that, 20% extra-ones with a population of 500 is generated by N=400 follows a geometric distribution and the one-inflation has size 100. Moreover, we have seen that modelling after truncating inflated counts is crucial for the predictive value. Of course, the estimator needs to be modified as n contains the one-inflated part. This leads to $\hat{N} = n + (n - f_1)p(0, \hat{\theta})/(1 - p(0, \hat{\theta}) - p(1, \hat{\theta}))$ where $p(x, \hat{\theta})$ is the estimated probability of a x occurring under the untruncated count distribution, and $\hat{\theta}$ has been estimated under the 0-1-truncated count distribution. Future work is needed for more realistic situations.

We have discussed a very general framework of point estimation for unknown size of populations. Some attention has to be paid to the confidence interval for N as well as uncertainty assessment. A more in-depth investigation will be carried out in future work.

References

[1] Godwin, R. T. and Böhning, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. Journal of the Royal Statistical Society: Series C (Applied Statistics), 66, 425–448.

[2] Godwin, R. T. (2017). One-inflation and unobserved heterogeneity in population size estimation. Biometrical Journal, 59, 79–93.

[3] Henningsen, A. and Toomet, O. (2011). maxLik: a package for maximum likelihood estimation in R. Computational Statistics, 26, 443–458.

[4] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663–685.

[5] Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. Biometrics, 43, 783–791.

[6] Böhning, D., Kaskasamkul, P., and van der Heijden, P. G. M. (2018). A modification of Chao's lower bound estimator in the case of one-inflation. Metrika, 82(3), 361–384.

[7] Böhning, D. and Friedl, H. (2021). Population size estimation based upon zero-truncated, one-inflated and sparse count data. Stat Methods Appl, 30, 1197–1217.

[8] Godwin, R. T. (2019). The one-inflated positive Poisson mixture model for use in population size estimation. Biometrical Journal, 61(6), 1541-1556.

[9] Z. MOHD. D. AL-BALUSHI and M. M. ISLAM. (1987). Geometric Regression for Modelling Count Data on the Time-to-First Antenatal Care Visit. Journal of Statistics Advances in Theory and Applications, 23(1), 35-57.

Development of hybrid tofu with dietary fiber supplementation as an alternative for dietary protein

Sarod Chaiworn [1], Suphat Phongthai[2], Churdsak Jaikang[3], Jetsada Ruangsuriya [1, 4, *]
[1]Departmentof Biochemistry, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand.
[2]School of Agro-Industry, Faculty of Agro-Industry, Chiang Mai University, Chiang Mai 50200, Thailand.
[3] Department of Forensic, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand.
[4]Functional Food Research Center for Well-Being, Science and Technology Research Institute, Chiang Mai University, Chiang Mai 50200, Thailand

**Corresponding Author Email:* jetsada.ruang@cmu.ac.th

Abstract. Tofu, widely popular around the world as an alternative source of dietary protein, is high-quality protein for healthy adults; however, tofu contains lowest sulfur-containing amino acids without dietary fiber. This study aimed to develop hybrid tofu (HT) from soymilk and cow milk with dietary fiber supplementation from oyster mushroom. Optimization of HT production was conducted and produced. Then, the HT was subjected for proximate, texture, and amino acid component analyses. For 50 g dried soybeans, it was found that 1,000 mL of water was optimized for soymilk with 4-time extraction within the optimal water volume. Suitable coagulant was 0.2% citric acid while 5% cow milk was optimized, and 10% ground mushroom was also optimized. Proximate analyses revealed that the HT contained higher humidity ($p \leq 0.001$) and fibers ($p \leq 0.001$) while contained lower nitrogen ($p \leq 0.001$), protein ($p=0.021$), and fat ($p \leq 0.001$) contents compared to conventional tofu (CT). Texture analyses indicated insignificant hardness, chewiness, and gumminess, whilst cohesiveness ($p=0.003$) and springiness ($p=0.003$) were significantly higher in CT. The microstructure of HT was revealed by obvious dietary fiber distribution. Lactose was not detectable in the HT. However, amino acid contents, including essential and non-essential amino acids, revealed a significantly lower in the HT. In conclusion, we successfully produced HT, as an alternative dietary protein, having similarly characteristics to CT with lower nutritious, especially fat and protein but higher dietary fibers. The HT might be suitable for certain groups such as patients with liver and kidney problems.

Keywords: hybrid tofu; conventional tofu; dietary fibers; dietary protein; alternative protein

1. Introduction

Dietary proteins are fundamentally important for the growth and necessary repairment of the body. When the body correctly receives adequate amounts of dietary protein daily, it can remain normal health. Although there are varieties of possible protein sources available in the market, high-quality proteins are strongly recommended due to the adequate amount of essential amino acids for our bodies need in all ages. The proteins from animal origins are classified as a high-quality protein (1). Therefore, the consumption of sufficient animal protein results in being healthy. On the other hand, certain population groups consider that consumption of animal products such as meat is cruel and violates the animal welfare, so they avoid consuming meats and meat products as much as they be able to perform. Those populations are recognized as vegetarian, and flexitarian. In addition, some elderly people who are incompetent to chew meats due to tooth deterioration, as well as a decline of the digestive system, can reduce the nutritional intake of meat and protein utilization in the body. The consequence of this avoidance may lead to protein-under nutritional condition. Therefore, high-quality protein alternative must be supplemented to compensate this defective condition.

The most famous alternative protein sources are plant proteins, especially legumes such as soybeans and peas (2). However, the limit contents of certain essential amino acid of plant protein poses a risk factor for insufficient consumption of essential amino acids to assemble the body's need. Tofu, a native food of east Asia, is another alternative dietary processing of plant protein of which the soy proteins are extracted. It is considered as a high-protein diet and become world-popular for many population groups, who are less restrictive and/or restrictive meat consumption. In addition, the tofu texture is normally too soft, fragile, and inflexible which may be typically caused by a lack of dietary fiber components.

The recommendation of dietary fiber intake is in the range of 25 - 38 g / day or 14 g / 1,000 kcals of food intake (3). The researcher indicated that dietary fiber intakes are beneficial for health and help prevent the occurrences of diverticulosis. Dietary fiber is a polymer group that is resistant to human enzymatic digestion. Most maximum dietary fibers are obtained from plants and non-plants. Despite the indigestibility, dietary fiber intakes are beneficial for health and help prevent the occurrences of diverticulosis (4) and hemorrhoid (5). Moreover, it reduces the risk of obesity (6), controls blood sugar (7), and decreases the absorption of fat and cholesterol, which are risk factors for cardiovascular diseases (8). Mushrooms has been a source of dietary fiber with the belief that their fibers could potentially adjust the texture of tofu which is soft and fragile to meat-like texture. Oyster mushroom or *Pleurotus pulmonarius* has been reported to contain over 26.05% protein of its dry weight. It produces three principal essential amino acids: isoleucine (Ile), leucine (Leu), and phenylalanine (Phe) (9). Moreover, it contains a significant amount of flexible fiber up to 28% of its dry weight (10).

Cow milk is a famous source of high quality protein consumed worldwide with affordable price and wide availability (11). However, certain groups of population may confront of lactose intolerance in which fresh cow milk should be processed before consumption as in yogurt and cheese in order to lactose eradication (12).

In this research, we would like to develop and analyse a new tofu from soymilk, cow milk, and oyster mushroom so-called hybrid tofu (HT), as for an alternative source of dietary protein in comparison with the conventional tofu (CT) available in the market. We hypothesised that overall physical properties of HT will not significantly change when compared with those of CT. However, the HT contained better notional properties in certain points of view when compared with the CT. We have achieved producing the HT of which the overall physical properties were in line with those of CT, but rich in dietary fiber. However, we could not achieve enrichment of dietary protein as well as amino acid as we intended. Nevertheless, the HT is a new formula of tofu may be suitable for certain groups of the population.

2. Method

2.1. Chemicals and materials

Food-graded soybean, cow milk powder, magnesium sulfate ($MgSO_4$), calcium sulfate ($CaSO_4$), and citric acid ($C_6H_8O_7$) were purchased from bakery supplier store (Mueang, Chiang Mai, Thailand) while oyster mushroom was purchased from a local market (Hangdong, Chiang Mai, Thailand).

2.2. Tofu preparation and optimization

50 g of soybean were cleaned and soaked in deionized (DI) water at 4°C overnight. Then, the water was completely decanted, and the soaked beans were grounded with different volume of water for its optimization of soymilk extraction as well as different times of water extraction using food processor (HR2115, Philips, China). The soymilk was separated from the Okara by filtration through a linin cloth. Next, the soy milk was heated at 95°C for 5 minutes and left cool down to 80°C. Three different types, which were citric acid, magnesium sulfate and calcium sulfate, as well as concentrations of the coagulants were investigated by adding them into the warmed soymilk and let the coagulation proceed for 30 minutes to form protein curd. The curd was separated from the whey and pressed in the 14×10.5×9 cm home-made tofu mold with 6 kg pressing weight for 3 hours. The tofu was separated to determine physical properties and stored at -80°C to determine nutritional properties. However, during optimization of tofu production, centrifugation at 2000×g for 5 minutes was used for curd and whey separation and tofu pressing.

The HT was produced at the optimized conditions; however, precipitation of cow milk was identical to soymilk precipitation. Then, addition of the cow milk protein precipitate into the soymilk and coprecipitated with an optimized coagulant was performed. Oyster mushroom was cooked by steaming for 5 minutes and ground with the food processor. The mixture was squeezed to discard the water using a linin filter cloth and supplemented into the soymilk and the cow milk curds. The curd was separated from the whey and pressed in the 14×10.5×9 cm home-made tofu mold with 6 kg pressing weight for 3 hours.

The CT was produced with the similar processes and ingredients, except the coagulant which was 0.2% (w/v) magnesium sulfate. The CT contained neither cow milk nor mushroom.

2.3. Total soluble protein determination by Lowry method

5 µL of samples, which were the soymilk and the whey, were diluted with 495 µL Distilled water (DI water). 100 µL of the diluted sample was mixed with 500 µL alkaline solution containing sodium hydroxide (NaOH) (28244.295, VWR, Czech), copper (II) sulfate ($CuSO_4$) (2771.290, VWR, Italy), potassium sodium tartrate tetrahydrate ($KNaC_4H_4O_6 \cdot 4H_2O$) (27068.233, VWR, Germany), and sodium carbonate (Na_2CO_3) (27771.290, VWR, Germany) and with 50 µL Folin-Ciocalteu solution (31360.264, VWR, France). The mixture was mixed and incubated at room temperature (RT) for 10 minutes. The absorbance of 650 nm was determined with a plate reader (Synergy™, H4BioTek, USA) and the bovine serum albumin (BSA) (AD0023, Biotechnology, Canada) at different concentrations was used as a standard protein for total protein calculation in the samples (13).

2.4. Proximate analysis of tofu

Proximate analysis of tofu was performed according to the standard protocol (AOAC 2000). Nitrogen, protein (with 6.25 conversion factor), moisture, fat, and crude fiber were determined and calculated with different methods (14), this analysis was performed at the Faculty of Agroindustry, Chiang Mai University, Thailand.

2.5. Amino acid analysis of tofu

10 mg of the wet tofu were mixed with 6 M HCl, degassed, and heated at 110°C for 22 hours. Then, the mixture was dried under vacuum and the dry matter was dissolved with 0.02

M HCl. The sample was then filtered through 0.2 µm filter. Finally, 20 µL of the sample were analysed with the amino acid analyzer (L-8900, Hitachi, Japan) at the office of scientific instrument and testing (OSIT), Prince of Songkhla University, Thailand (15).

2.6. Lactose determination

To determine whether there was lactose remaining in the hybrid tofu (HT), 5 g of the tofu were re-suspended into 5 mL DI water, mixed well, and centrifuged at 6,000× g for 5 minutes. 400 µL of the supernatant were mixed 400 µL Fearon solution containing sodium hydroxide (28244.295, VWR, Czech) and methylamine hydrochloride (CH_3NH_2HCl) (65600, Fluka AG, Switzerland), incubated at 70°C for 5 minutes, and cooled down on ice for 5 minutes. The absorbance was determined at 540 nm along with the 0-200 mg/mL lactose standard (61345, Sigma, Netherland)(16).

2.7. Texture profile analysis (TPA) of tofu

Texture profile analyses of tofu including hardness, cohesiveness, springiness, gumminess, and chewiness were determined with a texture analyser (TA.XTplus, Texture Analyser | Stable Micro Systems, UK). The sample was cut into 2x2 cm and placed onto the analyser with P/50 probe with the speed and compression degree of 1 mm/s and 30%, respectively (17). This analysis was performed at the Faculty of Agroindustry, Chiang Mai University, Thailand.

2.8. Water holding capacity (WHC)

10 g of the wet tofu were placed into a 50 mL centrifugal tube which contained cotton and filter paper. The tube was then centrifuged (MPW-352R, MPW MED Instrument, Poland) at 6,000×g for 20 minutes and the sample was removed and weighted. The WHC was calculated as in the percentage (18).

2.9. Scanning electron microscope (SEM)

The tofu samples were cut into small sizes and fixed with 2.5% (w/v) glutaraldehyde (G7651-10ML, Sigma, USA) in 0.1 M phosphate buffer, pH=7.4 (10010-023, Gibco, USA). Then, the samples were washed and dehydrated with 20-100% (v/v) ethanol (20821.321, VWR, France). The dehydrated samples were placed into a critical point dryer (K85, Quorum technologies, UK) overnight. The completely dried samples were immobilized on the aluminium stub prior to gold-sputter coat with Modular coating system (Q150 Plus, Quorum technologies, UK). The electron microscopic photographs were taken with a scanning electron microscope (JSM-6610LV, JEOL, Japan)(19).

2.10. Statistical analysis

Data shown were mean ± standard deviation (SD) from the triplicate results with 3 independent experiments. SPSS 25 (licencing by Chiang Mai University) was used for statistically analyses as indiscriminating between two groups by the student *t*-test while among groups with one-way analysis of variance (ANOVA) test followed with Tukey's or Dunn's multiple comparison tests as a *pos-hoc* test. The significant level was set at $p < 0.05$ for two-tail analysis.

3. Results and Discussion

Results

3.1. Optimization of the conditions for production of the hybrid tofu (HT) containing both cow milk and mushroom in a laboratory scale

The optimal volume of the water for soymilk production from 50 g dried soybean was 1,000 mL due to the significantly minimum water used but maximal protein gained (figure 1). We further observed that four times of soymilk extraction showed the maximal protein gained

in the soymilk (figure 2). The most suitable coagulant for both soymilk and cow milk precipitation were citric acid at 0.2% (w/v). Magnesium sulfate could only precipitate soymilk while calcium sulfate could not do neither types of milk as observed the high protein contents in the whey supernatant (figure 3). Addition of cow milk powder at 5 -10% (w/v) did not disturb the capacity of the coagulant used for tofu production as observed in the protein contents in the whey supernatant (figure 4). In addition, 10% of the ground well-done oyster mushroom was optimized for hybrid tofu (HT) production as due to the degree of dispersion, texture, and coordinated properties (table 1) observed by eyes. Therefore, the optimal condition for HT production was suggested as mentioned above, 5% (w/v) of cow milk powder was used in HT production as for minimising the cost. Due to the frozen storage had a strong effect on physical properties of both tofu (20), we evaluated those properties of our tofu without frozen storage. In contrast, the storage at -80°C was required for nutritional analysis because this method could effectively. Prevent nutritional degradation from metabolic processes (21) in both HT and CT.

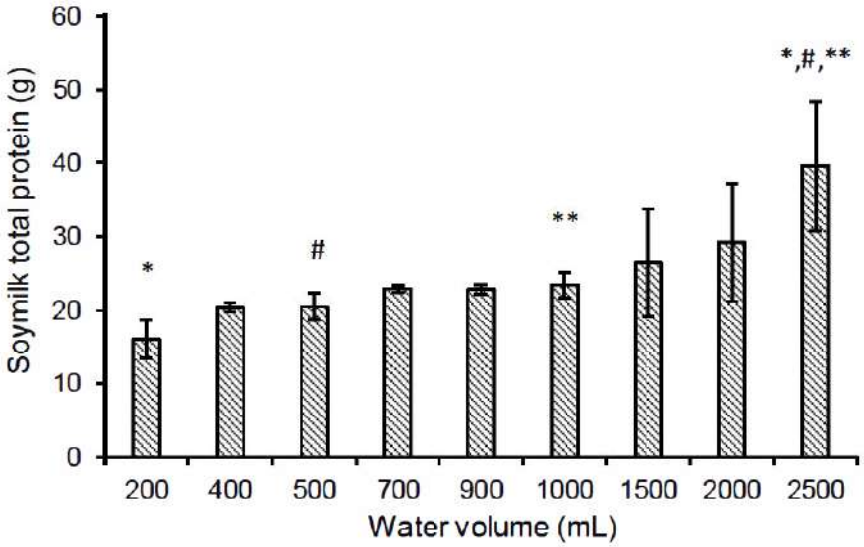


Figure 1. Protein contents in soymilk extracted with different water volumes. 50 g of the dried soybean was used to extract the soy milk with different volume of water (mL). Different designated letter indicates significant levels set at $p < 0.05$.

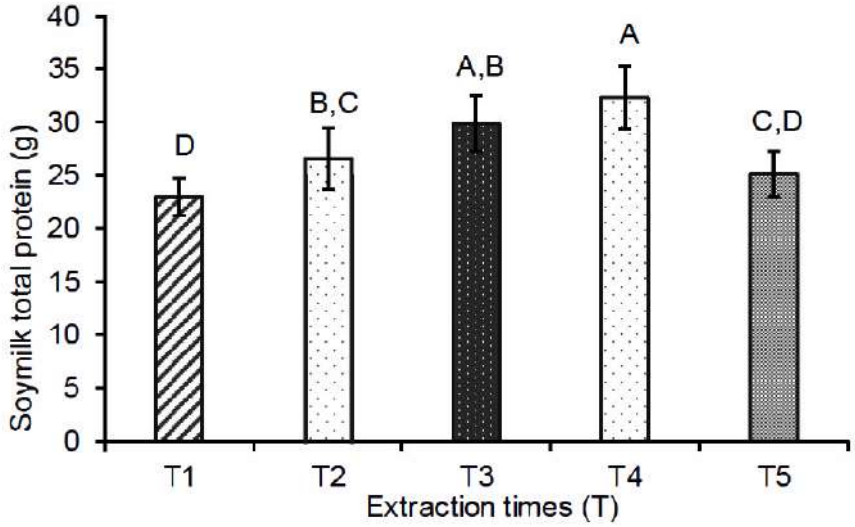


Figure 2. Protein contents in soymilk extracted with the same water volumes but different times of extraction. 50 g of the dried soybean and 1,000 mL DI water were used to extract the soy milk with times of extraction (T). Different designated letter indicates significant levels set at $p < 0.05$.

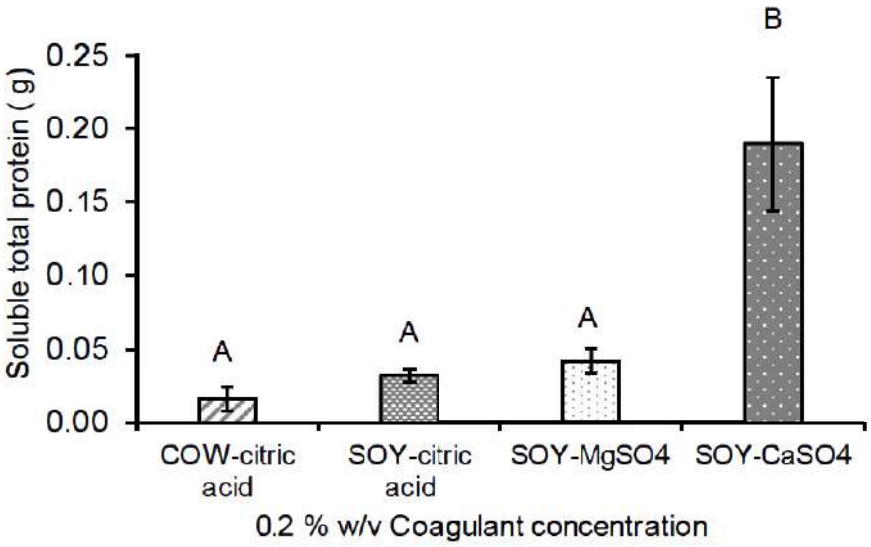


Figure 3. Soluble protein content in a whey supernatant after soymilk precipitation. Three types of coagulants, which were citric acid, magnesium sulfate, and calcium sulfate at 0.2% (w/v). Different designated letter indicates significant levels set at $p < 0.05$.

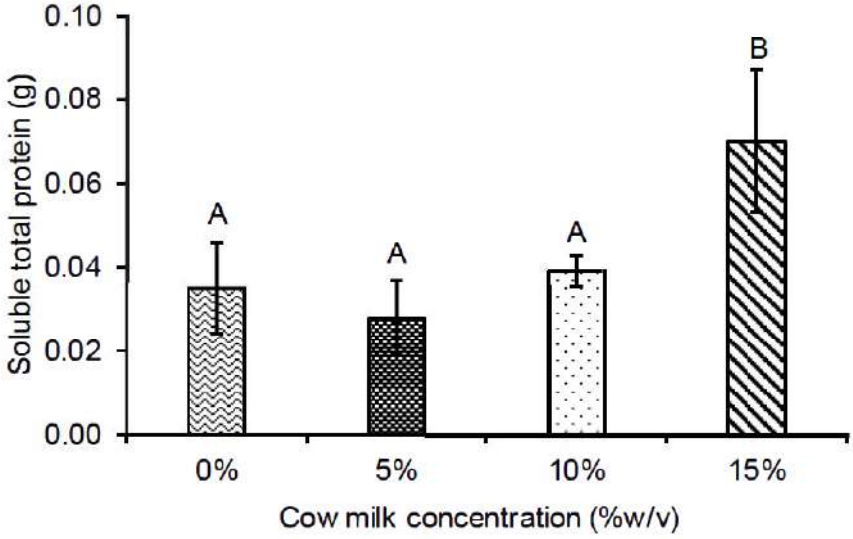


Figure 4. Soluble protein content in a whey supernatant of cow milk and soymilk mixture at different percentages (% (w/v)) after precipitation with 0.2% (w/v) citric acid. Different designated letter indicates significant levels set at $p < 0.05$.

Table 1. Hybrid tofu characteristics observed by eyes.

	Mushroom(%w/v)		
	5%	10%	15%
Dispersion	**	***	*
Texture coordinate	***	***	*

Different proportions of grounded well-done oyster mushroom (5 – 15 %w/v) were added into the tofu during precipitation. *, **, and *** indicate fair, good, and excellent, respectively.

3.2. Nutritional and physical properties in hybrid tofu

Proximate analyses revealed that the HT contained significantly higher humidity and fiber while contained significantly lower nitrogen, protein, and fat when compared with the CT (figure 5). In addition, amino acid contents in the CT were significantly higher than that of the HT in both essential as well as non-essential amino acids (table 2). Cysteine was undetectable in the HT, but very limited in the CT. Texture analyses indicated insignificant hardness, chewiness, and gumminess, whilst cohesiveness and springiness were significantly higher in CT (figure 6). The microstructure of HT was revealed by obvious dietary fiber distribution on rough surface topography (figure 7). Lactose contents were not detectable in both HT and CT (table 3).

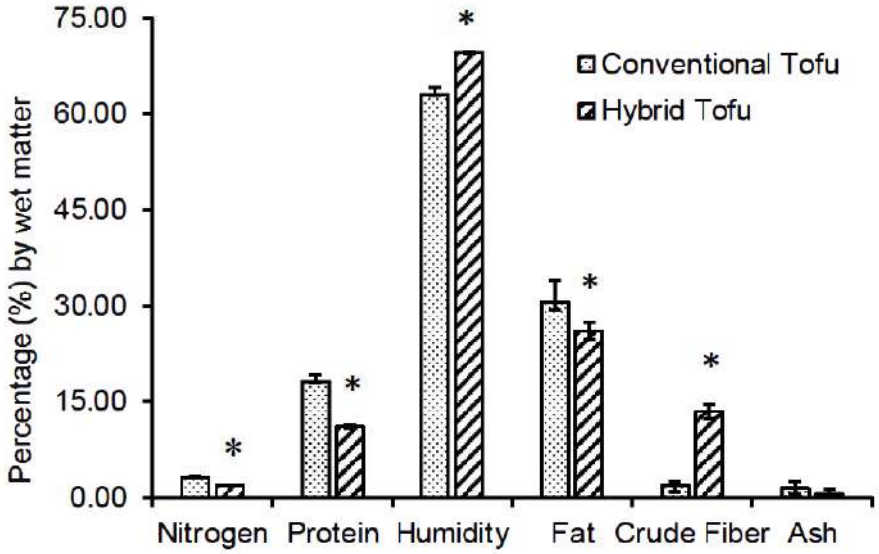


Figure 5. Proximate analyses of the hybrid tofu (HT) in comparison with conventional tofu (CT). The parameters include nitrogen, protein, humidity, fat, crude fiber, and ash between CT and HT. * indicates significant levels set at $p < 0.05$ between CT and HT.

Table 2. Amino acid of contents of the conventional tofu and hybrid tofu.

Amino acid	*mg of amino acid/100 g proteins*	
	Conventional tofu (CT)	*Hybrid tofu (HT)*
Essential amino acids		
His	1.828±0.11	0.907±0.00*
Ile	3.434±0.28	1.541±0.00*
Leu	6.203±0.00	2.902±0.09*
Lys	5.096±1.38	2.086±0.00*
Met	0.775±0.00	0.181±0.00*
Phe	4.376.0.22	1.723±0.00*
Thr	1.939±0.22	1.179±0.00*
Val	4.652±0.06	4.171±0.09*
Non-essential amino acids		
Asp	8.917±0.00	2.992±0.09*
Ser	3.877±0.17	1.723±0.00*
Glu	16.339±0.00	5.078±0.00*
Pro	1.772±0.33	0.907±0.09*
Gly	3.434±0.006	1.451±0.00*
Ala	3.046±0.28	1.814±0.00*
Cys	0.116±0.00	0.000±0.00*
Tyr	1.994±0.00	0.997±0.00*
Arg	5.926±0.006	2.267±0.00*

* indicates significant levels set at $p < 0.05$ between CT and HT.

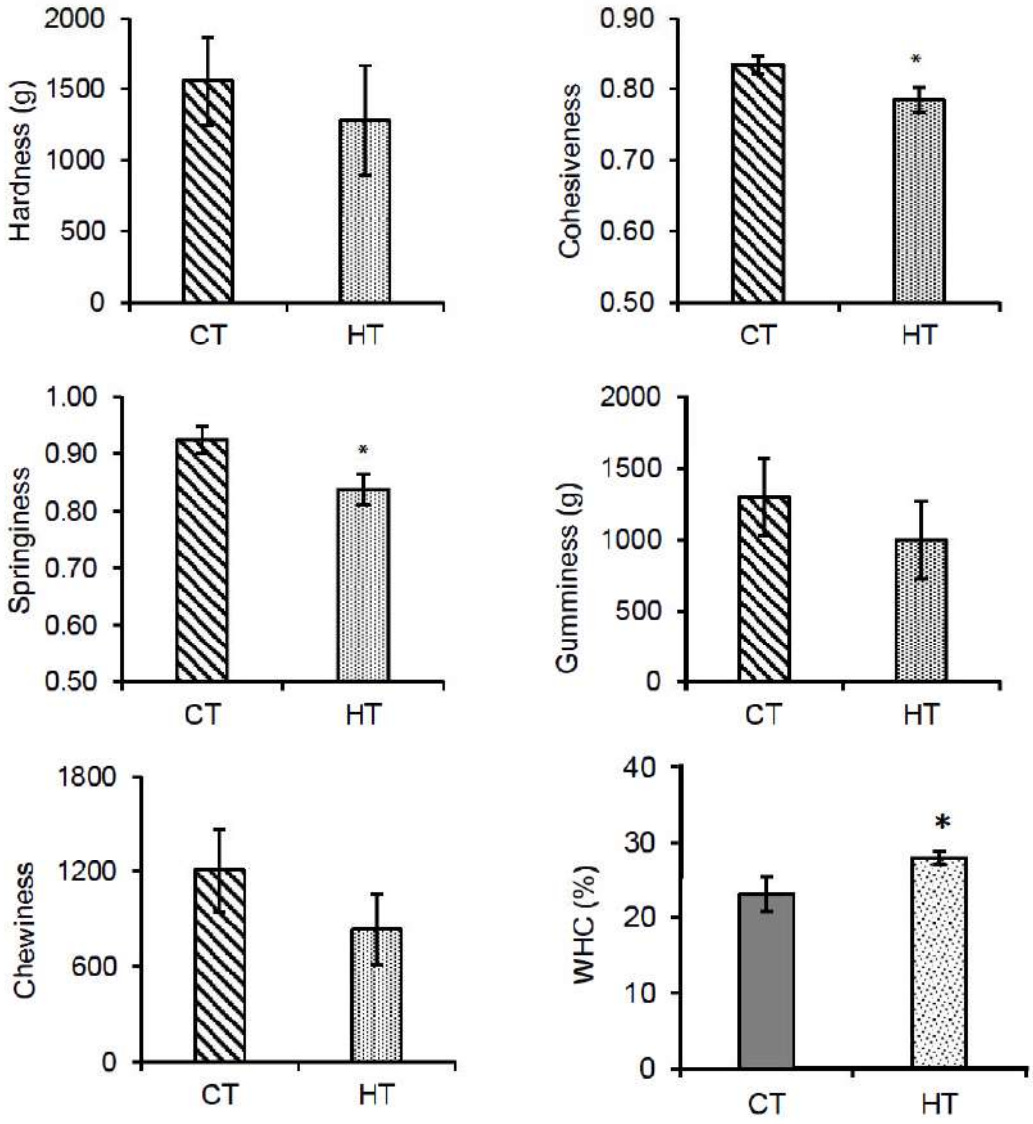


Figure 6. Texture profile of the hybrid tofu (HT) in comparison with the conventional tofu (CT). The parameters include hardness (g), cohesiveness, springiness, gumminess (J), chewiness (J) as well as, percentage of water holding capacity (%WHC) * indicates significant levels set at $p < 0.05$ for between CT and HT.

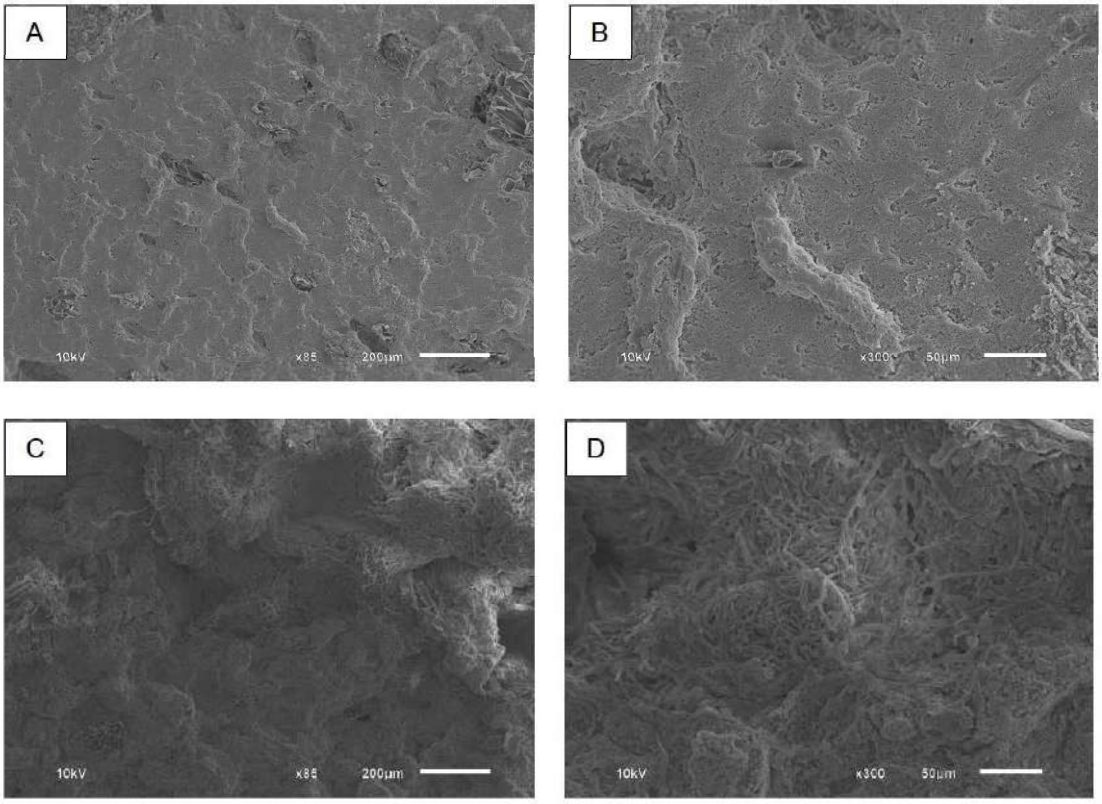


Figure 7. Electron microscopic picture of the tofu. The conventional tofu (A, B) in comparison with the hybrid tofu (C, D) were revealed at 85X (A, C) and 300X (B, D) magnification, and the scale bars represent the sizes in µm.

Discussion

We successfully optimized the production of the hybrid tofu (HT), which contained both soymilk, cow milk proteins and dietary fiber from oyster mushroom. First, we achieved the volume of water used for soymilk production at 1,000 mL per 50 g dried soybeans (1:20). There has been a great variety of the water volume used during soymilk protein extraction. The different studies report different ratios of soybean to water such as 1:5 (22), 1:7 (23), and 1:10 (24), and 1:18 (25) with successful tofu production at the final step. This implies that the volume of water used during soymilk extraction could be varied and have no significant effects on the tofu production.

It required 4 times of extractions to gain highest content of proteins in the soy milk. This finding was inconsistent to other reports where the soymilk extraction was done once from the soaked soy beans using a specific machine for soymilk production (26). However, our study followed the conventional method of extraction where simple equipment and tools should be available in basic kitchen. Huan and colleagues followed a soymilk extraction twice and the protein content in the soymilk was significantly higher than only once extraction (27).

Many coagulants have been reported for tofu production such as calcium salts including calcium chloride ($CaCl_2$), calcium sulfate ($CaSO_4$), and calcium acetate ($C_4H_6CaO_4$). Magnesium salts such as magnesium sulfate ($MgSO_4$) and tri-magnesium citrate ($C_{12}H_{10}Mg_3O_{14}$), acids such as citric acid ($C_6H_8O_7$), lactic acid ($C_3H_6O_3$), tartaric acid ($C_4H_6O_6$), malic acid ($C_4H_6O_5$), and gluconic acid ($C_6H_{12}O_7$) and enzymes such as chymotrypsin, pepsin, and papain have been referred against (28). Nevertheless, the most available tofu currently being sold in the market so-called conventional tofu (CT) is coagulated by either magnesium sulfate or calcium sulfate (29-31). This may be due to their market price affordability and the originality of the tofu production. The physical appearances of the CT are different depending

on the categories of coagulants. For example, tender and firm texture is a result from magnesium sulfate while soft and watery texture is a result from calcium sulfate (32). The suitable coagulant was reported at 0.4% (w/v) calcium sulfate for soymilk protein precipitation (29). Nevertheless, our HT was coagulated with citric acid due to inability of both magnesium sulfate and calcium sulfate to coagulate the cow milk protein. We had tested the taste of HT in volunteers, and it was discovered that there was nope sour flavor at all.

Another explanation of inability of both calcium and magnesium could not precipitate the cow milk was the presence of major phosphoprotein known as casein in cow milk (33), so ionic salt might not be suitable for its precipitation. Alternatively, acid coagulation is famous for precipitation of cow milk during cheese production. The cheese produced by 0.14% (w/v) citric acid was reported and the texture of the cheese was elastic and juicy due to hydrophobic interaction of the milk protein (34).

Due to the addition of cow milk protein into HT, the concern about lactose contamination was resisted. Our result proved that the lactose level in our HT was safe for anyone who cannot tolerate it. Regarding Fearon method sensitivity to detect lactose, our observation in the laboratory revealed that the minimum concentration of lactose detectable was 0.78 mg/mL. The undetectable results could be interpreted that there might be lactose in either the amount lower than 0.78 mg/mL or the absence. However, the amount of the lactose lower than 0.78 mg/mL was harmless to lactose intolerance subjects due to the finding that lactose consumption at 48 mg/mL/day was allowed for these individuals (35). In the other word, lactose more than 12 g/day (250 mL of fresh milk) are consumed once triggers the symptoms of lactose intolerance (36). In case of cheese, lactose is always free due to the protein precipitation by the coagulant and the complete separation of the protein and the whey (37). We added 5% (w/v) cow milk into soymilk for HT production in which 0.2% (w/v) citric acid was able to precipitate both sources of proteins. The mechanism on how citric acid can do their work was due to the principle of isoelectric precipitation (pI). When the pH of the milk solution was reached 4-6, it was matched with the protein's pI by causing total protein charge as 0 (38). Also, cow milk at 15% might be due to excess concentration protein (figure 4). In contrast, magnesium sulfate was used for tofu production from soymilk, through the process called salting out and isoelectric point theory precipitation increasing protein surface interaction for exposing hydrophobic state in protein structure and precipitation (39).Therefore, we are confident that the Fearon method potentially be applied to detect the lactose at the safer level for those lactose intolerance subjects.

We added 10% oyster mushrooms for supplementation of dietary fiber. Oyster mushroom or *Hed Nangpha* in Thai is one of the most famous edible mushrooms consumed among Thai people. It was higher contained as vitamin B, mineral and carbohydrate as well as fiber (40). Dietary fiber is improved physical properties such as texture and moisture as well as health improved properties as gut microbial (41). Like other mushroom, it is well known that oyster mushroom contains high-water contents which could help adjust the texture and the physical properties of the HT. Moreover, the affordability and availability of the oyster mushroom in the local market are always reasonable. Therefore, we chose oyster mushroom in our recipe of HT production as to provide dietary fiber. Oyster mushroom contained 11.05% of crude fiber (42, 43) including β-glucans which may ameliorate body such as improved glucose and lipid metabolism, associated with change of blood pressure and body weight, and involve in the control of hungriness sensations (44). In addition, chitin was of 24.11% in the crude fiber by its dried matter (43, 45), which was associated with immune responses (46).

Generally, the texture analyses of the HT were not different from those of the CT. Two parameters, which were cohesiveness and springiness, were significantly lower than those of the CT. It is considered acceptable as those two parameters are of the consequences to each other. Previous experiments by Huan *et al.* showed texture characteristics of the soft tofu from fermented whey as an coagulant that the hardness was 375 – 425 g, the cohesiveness was 0.35 – 0.45, the springiness was 0.85 – 0.9, the gumminess was 140 – 200 g, and the chewiness was 120 – 180 g (27). The difference of those texture parameter may cause by

Hofmeister series, by which protein-protein hydrophobic interaction in the structure of tofu was robust by cationic coagulant (47). In this study, HT and CT had similar characteristics of hardness, gumminess, and chewiness, but different from cohesiveness and springiness. These differences may due to the presence of both casein protein and mushroom fiber reducing internal bonding force as well as increasing brittleness in HT structures (27).

Nutritional compositions were low in HT because a mixture of HT contains an arrangement such as soy protein, cow milk, and fiber from mushrooms with relative infiltrated and changed size or volume by fiber. At certain points of view, HT was reduced its nutritional value especially protein and amino acid contents, but the fat content of HT seemed beneficial (figure 5). However, nutritional content in HT had no effect with alternative dietary protein consumption, which was consistent with Young and colleagues reporting that different extraction processes in soybean products contained different amino acids or nutrients (48). Both HT and CT contained relatively low amino acid contents, but not much different in term of protein content when compared with hen eggs. The protein contents of HT, CT, and hen egg were 11.03, 18.06, and 13.00%, respectively. Considering the amino acid contents of both CT and HT, all essential and non-essential amino acid components were much lower than hen egg (> 100X). Moreover, sulfur-containing amino acids including methionine and cysteine were absolutely limited in HT and CT while tryptophan and methionine were limited in hen eggs, but still met the body's need in all ages (49).

Due to our HT had no lactose or ineffective level of lactose, low fat, protein, and amino acid contents, it may be suitable for all groups who are interested in consumption it, especially the group of people who have problem with lactose consumption (37). In addition, it might be recommended to certain patients who have cirrhosis, chronic kidney disease (CKD), and end-stage renal disease (ESRD) that protein-calorie intake must be limited (50, 51). Moreover, anyone who requires caloric restriction (52), and who need low dietary protein digestion (53) can consume our HT. Although this finding showed the unexpected results in which the protein and the amino acid contents should have been significantly higher in HT due to the supplementation of cow milk, the HT is still a novel recipe for all people when produced and sold in the market as another source of dietary protein and the source of dietary fiber.

4. Conclusion

In conclusion, our hybrid tofu (HT) was produced successfully as an alternative source of dietary protein with similar characteristics of the conventional tofu as well as with lower nutritious in terms of amino acid and protein contents; however, it was beneficial over the CT in terms of dietary fibers and lower fat contents. It could be a potential dietary protein of choice suitable for certain groups whose dietary protein intake must be limited such as patients with liver and kidney diseases and whose dietary fuels must be limited such as fatness people.

Acknowledgments

The authors would like to acknowledge the support of the Department of Biochemistry, Faculty of Medicine, and Functional Food Research Center for Well-Being, Science and Technology Research Institute, Chiang Mai University, Chiang Mai 50200, Thailand. In addition, the TA/RA scholarship provided by the Graduate School, Chiang Mai University must be grateful.

References

[1] Tome D (2012) Criteria and markers for protein quality assessment - a review *British Journal of Nutrition 108*(2) 222-229.

[2] Mathai JK, Liu Y, and Stein HH (2017) Values for digestible indispensable amino acid scores (DIAAS) for some dairy and plant proteins may better describe protein quality than values calculated using the concept for protein digestibility-corrected amino acid scores (PDCAAS) *British Journal of Nutrition 117*(4) 490-499.

[3] Position of the American Dietetic Association 2008 Health implications of dietary fiber *Journal of the American Dietetic Association 108*(10) 1716-1731.

[4] Rezapour M, Ali S, and Stollman N (2018) Diverticular disease an update on pathogenesis and management *Gut and Liver 12*(2) 125-132.
[5] Higuero T, *et al*. (2016) Guidelines for the treatment of hemorrhoids (short report) *Journal of Visceral Surgery 153*(3) 213-218.
[6] Dayib M, Larson J, and Slavin J (2020) Dietary fibers reduce obesity-related disorders as mechanisms of *action Current Opinion in Clinical Nutrition and Metabolic Care 23*(6) 445-450.
[7] Müller M, Canfora EE, and Blaak EE (2018) Gastrointestinal transit time a glucose homeostasis and metabolic health as modulation by dietary fibers *Nutrients 10*(3).
[8] Jesch ED and Carr TP (2017) Food ingredients that inhibit cholesterol absorption Preventive nutrition and food *Science 22*(2) 67-80.
[9] Smiderle F, *et al*. (2012) Exopolysaccharides proteins and lipids in *Pleurotus pulmonarius* submerged culture using different carbon sources *Carbohydrate Polymers 87* 368–376.
[10] Raman J, *et al*. (2021) Cultivation and nutritional value of prominent *Pleurotus* spp. *Mycobiology 49*(1) 1-14.
[11] Smolenski G, *et al*. (2007) Characterisation of host defence proteins in milk using a proteomic approach *Journal of Proteome Research* 6(1) 207-215.
[12] Heine RG, *et al*. (2017) Lactose intolerance and gastrointestinal cow's milk allergy in infants and children common misconceptions revisited *The World Allergy Organization Journal 10*(1) 41.
[13] Ongprasert K, *et al*. (2020) Macronutrient immunoglobulin a and total antioxidant capacity profiles of human milk from 1 to 24 months a cross-sectional study in Thailand *International Breastfeeding journal 15*(1) 90.
[14] AOAC (2000). The association of official analytical chemists Gaithersburg MD USA In *Methods*.
[15] He W, Li P, and Wu G (2021) Amino acid nutrition and metabolism in chickens amino acids in nutrition and health Amino *Acids in Nutrition of Companion Zoo and Farm Animals* 109-131.
[16] Ruppersberg K, Herzog S, Kussler MW, and Parchmann I (2020) How to visualize the different lactose content of dairy roducts by Fearon's test and Woehlk test in classroom experiments and a new approach to the mechanisms and formulae of the mysterious red dyes *Chemistry Teacher International* 2(2).
[17] Yuan S and Chang S (2007) Texture profile of tofu as affected by instron parameters and sample preparation and correlations of instron hardness and springiness with sensory scores Journal of *Food Science 72*(2) 136-145.
[18] Petravic TV, *et al*. (2011) Rheological properties water-holding and oil-binding capacities of particulate β-glucans isolated from spent brewer's yeast by three different procedures *Food Technology and Biotechnology 49* 56-64.
[19] Guo Y, Hu H, Wang Q, and Liu H (2018) A novel process for peanut tofu gel as its texture microstructure and protein behavioral changes affected by processing conditions *Lebensmittel Wissenschaft and Technologie 96* 140-146.
[20] Poudyal RL, Kobayashi R, Suzuki T, and Watanabe M (2019). Effect of different freezing and storage condition on the physical properties of protein coagulum (Firm Tofu) *International Journal of Refrigeration 107* 11-19.
[21] Bishnoi S, Singhania N, and Rani S (2020) Freezing and frozen storage *In* 37-58.
[22] Khan SU, Pal MA, Wani SA, and Salahuddin M (2014) Effect of different coagulants at varying strengths on the quality of paneer made from reconstituted milk *Journal of Food Science and Technology 51*(3) 565-570.
[23] Liu ZS and Chang SK (2004) Effect of soy milk characteristics and cooking conditions on coagulant requirements for making filled tofu *Journal of Agricultural and Food Chemistry 52*(11) 3405-3411.

[24] Kao FJ, Su NW, and Lee MH (2003) Effect of calcium sulfate concentration in soymilk on the microstructure of firm tofu and the protein constitutions in tofu whey *Journal of Agricultural and Food Chemistry 51*(21) 6211-6216.
[25] Mullin WJ, *et al.* (2001) An interlaboratory test of a procedure to assess soybean quality for soymilk and tofu production *Food Research International 34* 669-677.
[26] Cao FH, *et al.* (2017) Effects of organic acid coagulants on the physical properties of and chemical interactions in tofu *Lebensmittel Wissenschaft and Technologie - Food Science and Technology 85* 58-65.
[27] Huang Z, *et al.* (2022) Evaluating the effect of different processing methods on fermented soybean whey-based tofu quality nutrition and flavour *Lebensmittel Wissenschaft and Technologie 158* 113-139.
[28] Zhang Q, *et al.* (2018) Research progress in tofu processing from raw materials to processing conditions *Critical Reviews in Food Science and Nutrition 58*(9) 1448-1467.
[29] Kao FJ, Su NW, and Lee MH (2003) Effect of calcium sulfate concentration in soymilk on the microstructure of firm tofu and the protein constitutions in tofu whey *Journal of Agricultural and Food Chemistry 51*(21) 6211-6216.
[30] Mullin WJ, *et al.* (2001) An interlaboratory test of a procedure to assess soybean quality for soymilk and tofu production *Food Research International 34* 669-677.
[31] Prabhakaran MP, Perera CO, and Valiyaveettil S (2006) Effect of different coagulants on the isoflavone levels and physical properties of prepared firm tofu *Food Chemistry 99*(3) 492-499.
[32] Jayasena V, Tah WY, and Abbas S (2014) Effect of coagulant type and concentration on the yield and quality of soy-lupin tofu *Quality Assurance and Safety of Crops and Foods 6* 159-166.
[33] Kennedy BT (2011) Caseins *Handbook of Food Proteins* 13-29.
[34] Obatolu VA (2008) Effect of different coagulants on yield and quality of tofu from soymilk *European Food Research and Technology 226*(3) 467-472.
[35] Shaukat A, *et al.* (2010) Systematic review as effective management strategies for lactose intolerance *Annals of Internal Medicine 152*(12) 797-803.
[36] [Internet] I O (2006) Living with lactose intolerance
[37] Mullin WJ and Emmons DB (1997) Determination of organic acids and sugars in cheese milk and whey by high performance liquid chromatography *Food Research International 30*(2) 147-151.
[38] Liu HH (2011) Effect of microwave heating on the viscoelastic property and microstructure of soy protein isolate gel *Journal of Texture Studies 42* 2011-2042.
[39] Guan X, *et al.* (2021) Changes of soybean protein during tofu processing *Foods 10*(7)
[40] Ueawiwatsakul S, Mungcharoen T, and Tongpool R (2014) Life cycle assessment of sajor-caju mushroom (Pleurotus sajor-caju) from different sizes of farms in Thailand *International Journal of Environmental Science and Development 5* 435-439.
[41] Makki K, Deehan EC, Walter J, and Bäckhed F (2018) The impact of dietary fiber on gut microbiota in host health and disease *Cell Host and Microbe 23*(6) 705-715.
[42] Agriculture USDO (2018) *Mushrooms, oyster, raw.*
[43] Ogundele G, Salawu S, Abdulraheem IA, and Bamidele O (2017). Nutritional composition of oyster mushroom (*Pleurotus ostreatus*) grown on softwood (*Daniella oliveri*) sawdust and hardwood (*Anogeissus leiocarpus*) sawdust *British Journal of Applied Science and Technology 20* 1-7.
[44] Dicks L and Ellinger S (2020) Effect of the Intake of oyster mushrooms (*Pleurotus ostreatus*) on cardiometabolic parameters a systematic review of clinical trials *Nutrients 12*(4) 1134.
[45] Kabir M, *et al.* (2020).Production of chitosan from oyster mushroom for α-amylase immobilization *Bangladesh Journal of Botany 49* 593-599.
[46] Elieh D, Sharma L, and Dela CS (2018) Chitin and its effects on inflammatory and immune responses *Clinical reviews in Allergy and Immunology 54*(2) 213-223.

[47] Hyde AM, *et al.* (2017) General principles and strategies for salting-out informed by the hofmeister series *Organic Process Research and Development 21*(9) 1355-1370.
[48] Young VR (1991) Soy protein in relation to human protein and amino acid nutrition *Journal of the American Dietetic Association 91*(7) 828-835.
[49] Munks B, Robinson A, Beach EF, and Williams HH (1945) Amino acids in the production of chicken egg and muscle *Poultry Science 24*(5) 459-464.
[50] Eghtesad S, Poustchi H, and Malekzadeh R (2013) Malnutrition in liver cirrhosis:the influence of protein and sodium *Middle East Journal of Digestive Diseases 5*(2) 65-75.
[51] Zha Y and Qian Q (2017) Protein nutrition and malnutrition in CKD and ESRD *Nutrients 9*(3).
[52] Dirks AJ and Leeuwenburgh C (2006) Caloric restriction in humans as potential pitfalls and health concerns *Mechanisms of Ageing and Development 127*(1) 1-7.
[53] Delimaris I (2013) Adverse effects associated with protein intake above the recommended dietary allowance for adults *International Scholarly Research Notices Nutrition 2013* 126929-126929.

Effect of Various Extraction Solvents on Efficiency of Natural Pigment Based Dye-Sensitized Solar Cell

Ponnambalam Sabarikirishwaran[1,2], Pattranan Junluthin[4], Yuwalee Unpaprom[2,3] and Rameshprabu Ramaraj[1,2*]

[1]School of Renewable Energy, Maejo University, Chiang Mai, 50290, Thailand
[2]Sustainable Resources and Sustainable Engineering Research Lab, Maejo University, Chiang Mai, 50290, Thailand
[3]Program in Biotechnology, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand
[4]Biomedical Science, Faculty of Science (Salaya Campus), Mahidol University, Nakhon Pathom 73170, Thailand

**Corresponding Author Email: rrameshprabu@gmail.com*

Abstract. Natural dye has been utilized as a substitute for ruthenium-based dyes in Dye-Sensitized Solar Cell (DSSC). Natural dye was successfully extracted from *Ficus benjamina* using ethanol, methanol and double distilled water as a solvent. The effect of solvents on the efficiency of prepared DSSCs has been studied extensively. Results revealed that pigment extracted with methanol has demonstrated the highest efficiency(η) of 0.198%, open-circuit voltage (V_{OC}) 61.6 mV and short circuit current (I_{SC}) 0.424 mA for *F. benjamina*. It has been observed that methanol has enhanced the efficiency of DSSC by 15.67% compared to double distilled water and 11.05% compared to ethanol for *F. benjamina*. Further, UV spectrophotometer is used to analyze the absorption characteristics of prepared natural dyes.

Keywords: DSSC, Natural Dye, Solvents, *Ficus benjamina*, Biosolar

1 Introduction

Every growing energy demand at a global scale on one side of the equation and global warming on the other side of the equation does not balance very well. Renewable energy sources can bring the balance we are looking for. They must meet certain criteria, such as low costs of production and high biodegradability and sustainability. These renewable sources can replace high carbon dioxide emissions production methods like natural coal, gas and oil [1–4]. Solar cells technology has demonstrated the harvesting of solar energy with significantly good efficiencies [5, 6], and the technology has constantly been evolving by making iterative improvements to the power conversion efficiency of the cell. Solar Photovoltaics (PV) technology is a well-established technology, but factors like the production cost complexity of manufacturing make it difficult to use PV cells for various applications [7, 8]. Therefore, several PV designs were explored to overcome these challenges. Dye-Sensitized Solar Cells (DSSC) has attracted substantial attention over the

past decade for their straightforward manufacturing method, high cost to efficiency ratio, and ability to deform their shape [9–11]. However, DSSCs struggle to maintain stability and photoconversion efficiencies are lower than that of the crystalline silicon-based PV cells [12–15]. Over the years, several techniques were adopted to improvise the efficiency of DSSCs, noticeably interested in developing photoanode capable of photocatalytic dye degradation [16–18]. DSSCs use dye molecules as sensitizers to harvest light energy and convert it into electrical energy. The dye molecules are often loaded on wide bandgap semiconductor material such as titanium dioxide (TiO_2) or zinc oxide (ZnO) nanostructures.

Transparent Conductive Oxide (TCO), such as Indium tin oxide or Fluorine doped tin oxide (FTO), is used as a substrate for electrodes. Platinum has been widely used as a counter electrode for DSSCs due to its very good catalytic ability. This is important because the reaction rate at the counter electrode has to be twice as fast as that of the working electrode. Iodine-based redox couple (I^-/I_3^-) is used as an electrolyte. When DSSC is illuminated with sunlight, the dye molecules readily absorb the light and excite electrons from Highest Occupied Molecular Orbital (HOMO) to Lowest Unoccupied Molecular Orbital (LUMO) shown in Figure 1. The excited electron jumps to the conduction band of the TiO_2 working electrode and flows through the external circuit to reach the counter electrode. In the meantime, electrolytes lend one electron to dye molecules to retore them and undergo reduction. Whereas electron from the counter electrode restores the electrolyte to keep the reaction going [19–21]. Copper complexes based redox electrolytes has been reported to achieve the highest efficiency of 13.1% under 100 mW/cm^2 conditions and the efficiency goes even higher up to 32% for photon flux of 1000 lux using indoor light [22].

This study investigates the effects of solvents such as ethanol, methanol, and double-distilled water (dd-water) as an extraction solvent for natural dyes extracted from *Ficus benjamina* leaves. Pigment concentrations of as prepared natural dye were estimated and studies in correlation with the photoelectric performance of the prepared DSSCs. In addition, UV-Vis Spectrophotometer is employed to study the absorbance characteristics of the prepared dye solution.

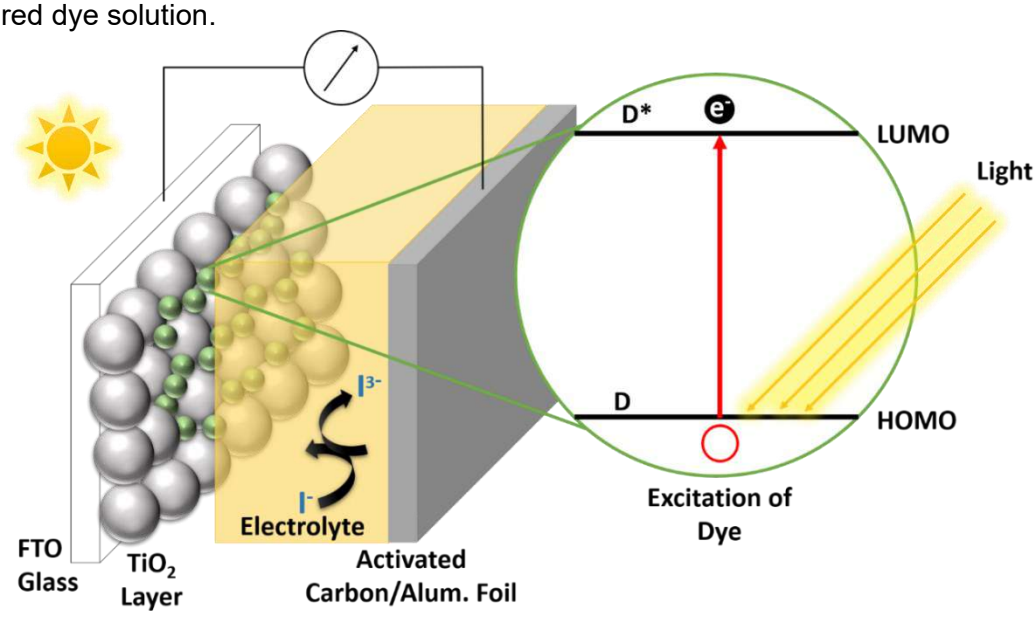


Figure 1 Schematic of Dye-Sensitized Solar Cell.

2 Materials and Methods

Chemicals such as ethanol (99.99%), acetonitrile (99.99%), potassium iodide (99.99%), ethylene glycol (99.99%), surfactant (Tween 20), acetic acid (99.99%), and iodine and

titanium dioxide (99.99%) were purchased from Union Science. The Fluorine-doped Tin Oxide (FTO glass) with a resistance of 10 ohms/sq. meter was purchased from Hangzhou, Zhejiang, China.

2.1 Dye Extraction

Natural dye extract was prepared using *F. benjamina* leaves, commonly known as Weeping Fig. The leaves were collected and washed to remove any visible contaminants and then dried in the dark to avoid chlorophyll being exposed to sunlight. The leaves are then weighted and blended with ethanol. The blended solution is kept in a dark and dry place for a few minutes until the ethanol reacts with the leaves. Later this mixture is vacuum filtered with filter paper using a rotatory pump. The filtrate containing pigment is then stored in a dark and cool place to avoid the disintegration of chlorophyll. This extraction method has been adopted by sumanta et al. [23]

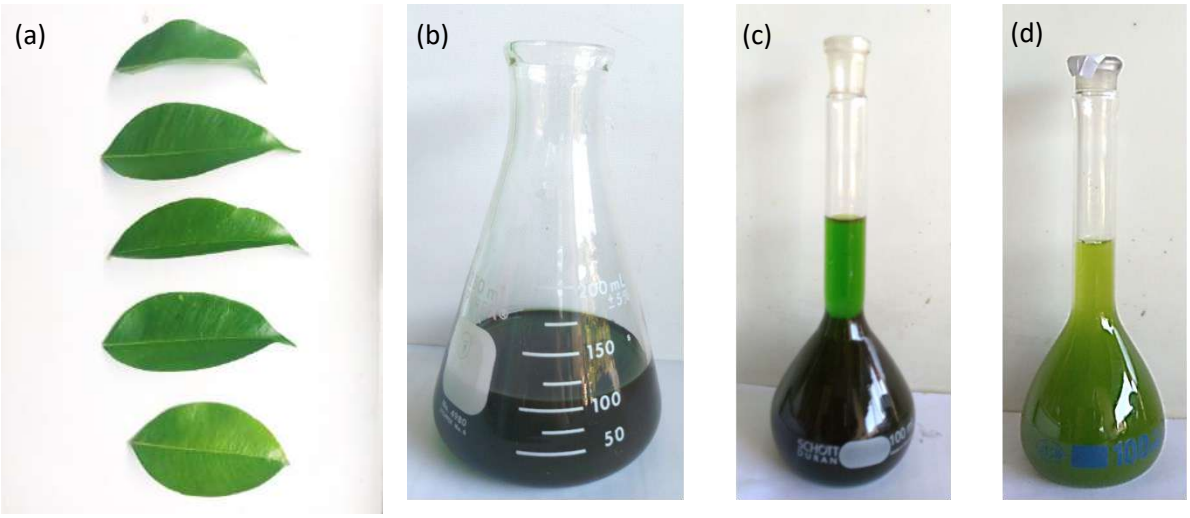


Figure 2 (a) Leaves of *F. benjamina* (weeping fig), Dye extract prepared using (b) Ethanol, (c) Methanol (d) double distilled Water (dd-water).

2.2 Photoanode Preparation

Mesoporous TiO_2 thin film photoanode was prepared by grinding the TiO_2 powder with a magnetic stirrer until it became fine powder and then acetic acid and surfactant were added and mixed thoroughly. Then the prepared TiO_2 paste is deposited into the FTO glass using Doctor's Blade method and sintered to form the mesoporous layer.

2.3 Dye Loading

The photosensitizer was absorbed onto the photoanode by carefully applying the prepared dye solution dropwise and dried in the dark for a few hours. Moreover, this process is repeated two more times to produce better adsorption of dye molecules onto the photoanode.

2.4 Preparation of Counter Electrode

The counter electrode was prepared by mixing activated carbon powder with ethanol, and the paste was then deposited onto aluminum foil using Doctor's Blade method and then dried at high temperature.

2.5 Electrolyte

I^-/I^{3-} redox couple used as a liquid electrolyte and prepared using a simple one-step method adopted from Gu et al. [24] for small-scale DSSC preparation. Among other liquid electrolytes, iodine-based redox couple electrolyte has been reported to perform better than others and have been studied widely by researchers [24].

2.6 Fabrication of DSSC

DSSC fabrication method is shown in the Figure 3. Cell assembly starts by attaching copper contacts to FTO glass deposited with TiO_2 photoanode. Then polyethylene membrane soaked in iodide electrolyte is sandwiched between a photoanode and a counter electrode made of aluminum foil coated with a thin layer of activated carbon.

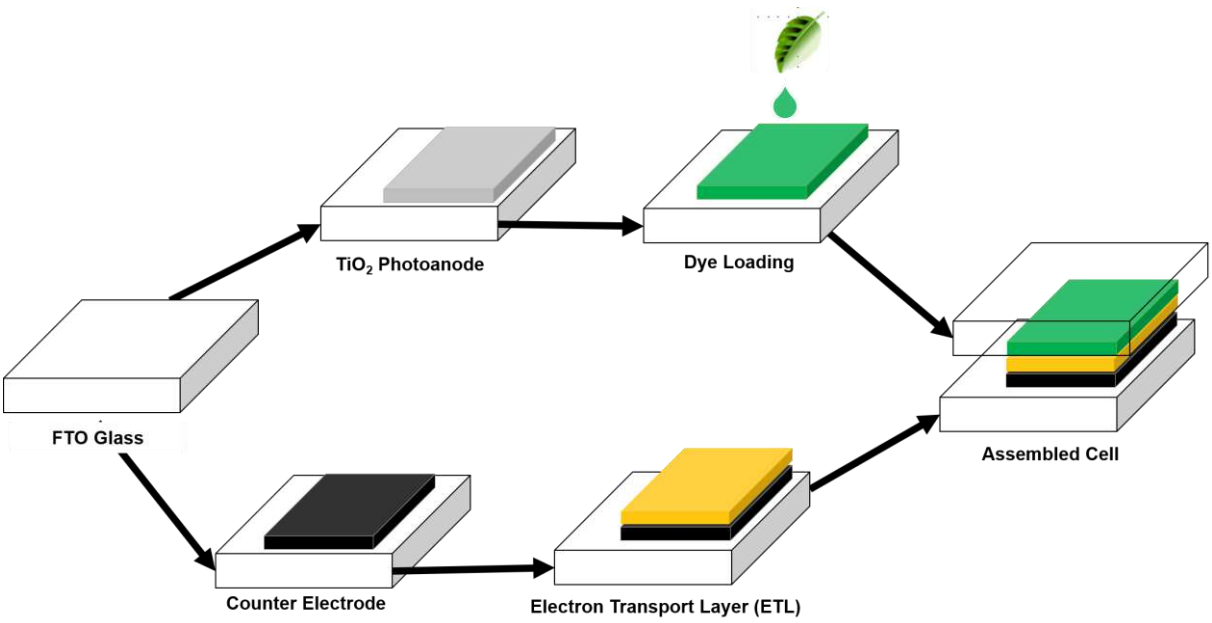


Figure 3 Schematic of DSSC Cell Assembly.

2.7 Photoelectric Characterization of DSSC

The prepared cells' photoelectric conversion efficiency (PCE) was measured using the solar simulator. The DSSCs with an active area of 3 cm^2 were placed under the simulator with a power of 190 mW cm^{-2}. MCP41010 microcontroller was programmed to increase resistance in a step-wise manner while the voltage and current response of the cell was recorded. The PCE was determined using the equation (1):

$$\eta = \frac{J_{SC} \times V_{OC} \times FF}{P_{in}} \quad (1)$$

where J_{SC}, V_{OC}, FF, and P_{in} are the Short-Circuit Current Density, Open-Circuit voltage, Fill Factor, and Incident Power. Then, the ratio of maximum power produced by the

cell to the theoretically maximum power for the given cell is given by Fill Factor (FF) and calculated using equation (2):

$$FF = \frac{P_m}{(I_{SC} \times V_{OC})} = \frac{I_m \times V_m}{(I_{SC} \times V_{OC})} \quad (2)$$

2.8 Statistical Analysis

The experiment was conducted with three replicates. Hence, the data were presented as mean ± standard deviation (SD).

3 Results and Discussion

3.1 Photovoltaic Performance of Cell

The PV efficiency of DSSCs can be calculated using the voltage-current (I-V) line. The parameters for the output PV of DSSCs include short-circuit current (Isc) and the open circuit potential (Voc) as well as Fill factor (FF) along with the power conversion rate (e). The photovoltaic efficiency of the as-prepared cells proved the presence of F. benjamina dye is a key component. The PV performance of DSSCs is estimated from the current–voltage (I–V) curve. The output PV parameters of DSSCs are short-circuit current (Isc), open-circuit potential (Voc), fill factor (FF), and power conversion efficiency (η). The photovoltaic performance of as-prepared cells showed *that F. benjamina dye-based DSSCs could attain* steady-state operation in converting solar energy to electrical energy. *Table 1* captures critical photovoltaic parameters of DSSCs and the results indicate methanol solvent for dye extraction has the highest power conversion efficiency (η) of 0.198% for *F. benjamina* based dye extract.

Table 1 DSSC Performance Comparison

Dye	Solvent	V_{oc} (V)	J_{sc} (mA/cm^2)	FF (%)	η (%)	Reference
F. benjamina	Ethanol	0.2511	0.0089	14.98	0.016	**This Work**
	Methanol	0.0616	0.0447	31.72	0.198	
	dd-H_2O	0.6026	0.0060	12.88	0.012	
Verdant-turmeric	Ethanol	0.540	1.02	60	0.33	Hossain et al. [25]
	Methanol	0.555	0.970	65	0.31	
Acanthus sennii chiov.	Ethanol + 1% HCl	0.507	0.491	60.4	0.150	Ayalew et al. [26]
	Distilled water + 1% HCl	0.475	0.351	60.6	0.101	
Lyceum Shawii	Ethanol	0.580	0.420	42	0.100	latif et al. [27]
Doum pericarp	Ethanol	0.37	0.005	63	0.012	Mohammed et al. [28]
	Distilled water	0.50	0.010	66	0.033	
Olive grain	Ethanol	0.550	0.580	38	0.120	Tamira et al. [29]

3.2 UV-Vis Characterization

UV-Vis is a different low-cost, rapid, and easy method of characterization that is typically used to Nano-Materials (NMs) research. Certain Magnetic Nano-Materials (MNMs) possess optical properties which are sensitive to size, shape of the agglomeration, state of agglomeration concentration, shape and the refractive index close to the surface of the NM which makes UV-Vis spectroscopy an important method to study these materials. The absorption characteristics of natural dye extracted using different solvents are studied using a UV-Vis photo-spectrometer and the results are shown in Figure 4. Absorbance spectra show peaks in 400 - 550 nm and 600 - 750 nm. This absorbance range indicates the signature of chlorophyll pigment [30], [31]. Further, the line shape of absorbance peaks for chlorophyll are in close conjunction with the literature [32, 33]. The spectra depict maximum absorbance λ_{max} at 440, 470 and 480 nm for dd-water, methanol and ethanol, respectively; this shows lower concentrations of chlorophyll in dd-water. On the other hand, ethanol and methanol showed similar absorbance characteristics, whereas ethanol with strong absorbance at 480 nm.

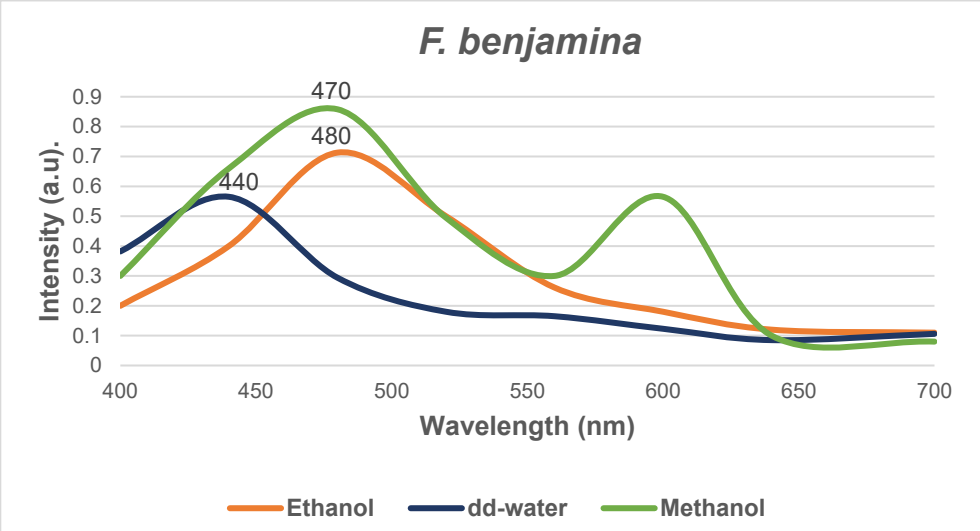


Figure 4 UV-Vis absorbtion spectra of dye extracted from *F. benjamina* using ethanol, methanol and dd-water as a solvent

Figure 5 captures Mortar and piston, plant extract and dye solution used for grinding plant leaves with solvent, which is later used to estimate pigments in the solution. The solvent is added to the raw plant extract solution to make it up to 100ml. The solution in the cuvette represents the sample used for UV characterization.

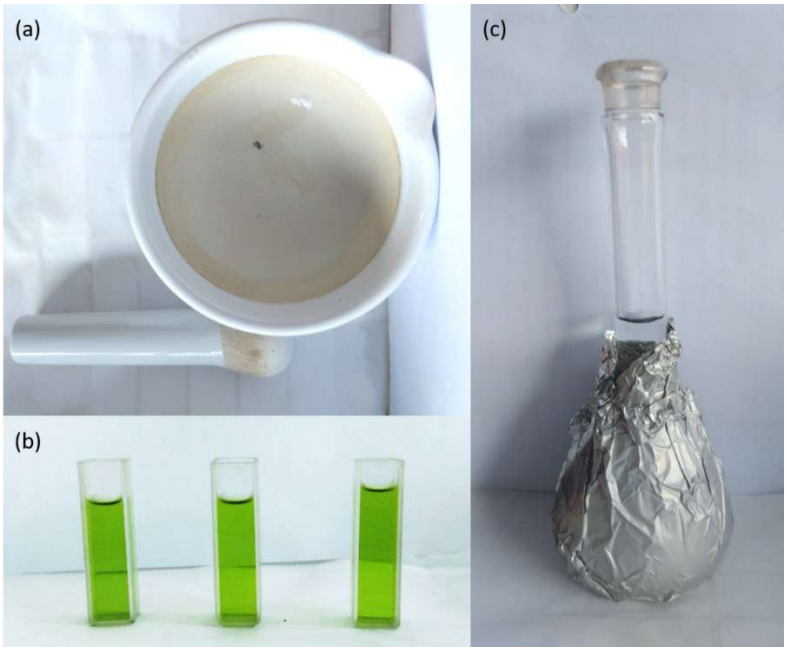


Figure 5 (a) Mortar and Piston (b) Plant extract on cuvette (c) Prepared Dye solution

4 Conclusion

The photochemical performance of DSSCs with *F. benjamina* based natural dye extract has been studied extensively. The effects of three different extraction solvents on the cell's performance have been carefully analyzed. The pigment analysis of the extracted dye revealed that the carotenoid to chlorophyll concentration ratio plays a crucial role in the light harvestability of the natural dye. Photochemical studies have revealed that methanol has shown high performance compared to other solvents with an efficiency of 0.198% and Current Density of 0.0447 mA/cm^2, Maximum Power Density of 0.0113 μW/cm^2, Fill Factor (FF) of 31.72% and open-circuit voltage of 0.0616 V. Methanol has acted as anti-aggregation agent for dye molecules which enhanced the overall efficiency of the DSSC. In summary, methanol as a solvent has performed better than ethanol and ethanol has performed better than dd-water.

Acknowledgements The authors kindly acknowledged the School of Renewable Energy, Energy Research Center, Maejo University, Chiang Mai, Thailand and the Program in Biotechnology for the research fund and facilities to accomplish this experimental study.

References

1. Ramaraj R, Junluthin P, Dussadee N, Unpaprom Y (2022) Potential evaluation of biogas production through the exploitation of naturally growing freshwater macroalgae *Spirogyra varians*. Environ Dev Sustain 1-12.
2. Junluthin P, Pimpimol T, Whangchai N (2021) Efficient conversion of night-blooming giant water lily into bioethanol and biogas. Maejo Int J Energ Environ Comm 3 (2):38–44.
3. Bautista K, Unpaprom Y, Junluthin P. et al. Ethanol production from corn stalk juice by *Saccharomyces cerevisiae* immobilized yeast using a green method. Biomass Conv Bioref. (2022).
4. Trejo M, Bhuyar P, Unpaprom Y, et al (2021) Advancement of fermentable sugars from fresh elephant ear plant weed for efficient bioethanol production. Environ Dev Sustain 1–11.

5. Outhred H, Retnanestri M (2015) Insights from the experience with solar photovoltaic systems in Australia and Indonesia. Energy Procedia 65:121–130.
6. Razykov TM, Ferekides CS, Morel D, et al (2011) Solar photovoltaic electricity: Current status and future prospects. Sol Energy 85:1580–1608.
7. Polman A, Knight M, Garnett EC, et al (2016) Photovoltaic materials: Present efficiencies and future challenges. Science (80-) 352:. https://doi.org/10.1126/SCIENCE.AAD4424/SUPPL_FILE/POLMAN.SM.PDF
8. Singh R, Alapatt GF, Lakhtakia A (2013) Making solar cells a reality in every home: Opportunities and challenges for photovoltaic device design. IEEE J Electron Devices Soc 1:129–144. https://doi.org/10.1109/JEDS.2013.2280887
9. Mejica GFC, Unpaprom Y, Ramaraj R (2021) Fabrication and performance evaluation of dye-sensitized solar cell integrated with natural dye from Strobilanthes cusia under different counter-electrode materials. Appl Nanosci. https://doi.org/10.1007/s13204-021-01853-0
10. Khammee P, Unpaprom Y, Whangchai K, Ramaraj R (2020) Comparative studies of the longan leaf pigment extraction as a photosensitizer for dye-sensitized solar cells' purpose. Biomass Convers Biorefinery. https://doi.org/10.1007/s13399-020-01060-x
11. Khammee P, Unpaprom Y, Thurakitseree T, et al (2021) Natural dyes extracted from Inthanin bok leaves as light-harvesting units for dye-sensitized solar cells. Appl Nanosci. https://doi.org/10.1007/s13204-021-01769-9
12. Nazeeruddin MK, Baranoff E, Grätzel M (2011) Dye-sensitized solar cells: A brief overview. Sol Energy 85:1172–1178. https://doi.org/10.1016/J.SOLENER.2011.01.018
13. Lee CP, Lin RYY, Lin LY, et al (2015) Recent progress in organic sensitizers for dye-sensitized solar cells. RSC Adv 5:23810–23825. https://doi.org/10.1039/C4RA16493H
14. Maiaugree W, Lowpa S, Towannang M, et al (2015) A dye sensitized solar cell using natural counter electrode and natural dye derived from mangosteen peel waste. Sci Reports 2015 51 5:1–12. https://doi.org/10.1038/srep15230
15. Wöhrle D, Meissner D (1991) Organic Solar Cells. Adv Mater 3:129–138. https://doi.org/10.1002/ADMA.19910030303
16. Momeni MM (2016) Dye-sensitized solar cell and photocatalytic performance of nanocomposite photocatalyst prepared by electrochemical anodization. Bull Mater Sci 39:1389–1395. https://doi.org/10.1007/S12034-016-1280-1/TABLES/2
17. Kanimozhi S, Prabu KM, Thambidurai S, Suresh S (2021) Dye-sensitized solar cell performance and photocatalytic activity enhancement using binary zinc oxide-copper oxide nanocomposites prepared via co-precipitation route. Ceram Int 47:30234–30246. https://doi.org/10.1016/J.CERAMINT.2021.07.203
18. Preeyanghaa M, Vinesh V, Sabarikirishwaran P, et al (2022) Investigating the role of ultrasound in improving the photocatalytic ability of CQD decorated boron-doped g-C3N4 for tetracycline degradation and first-principles study of nitrogen-vacancy formation. Carbon N Y 192:405–417. https://doi.org/10.1016/J.CARBON.2022.03.011
19. Shaheen SE, Brabec CJ, Sariciftci NS, et al (2001) 2.5% efficient organic plastic solar cells. Appl Phys Lett 78:841. https://doi.org/10.1063/1.1345834
20. Grätzel M (2004) Conversion of sunlight to electric power by nanocrystalline dye-sensitized solar cells. J Photochem Photobiol A Chem 164:3–14. https://doi.org/10.1016/J.JPHOTOCHEM.2004.02.023
21. Goncalves LM, De Zea Bermudez V, Ribeiro HA, Mendes AM (2008) Dye-sensitized solar cells: A safe bet for the future. Energy Environ Sci 1:655–667. https://doi.org/10.1039/B807236A
22. Cao Y, Liu Y, Zakeeruddin SM, et al (2018) Direct contact of selective charge extraction layers enables high-efficiency molecular photovoltaics. Joule 2:1108–1117. https://doi.org/10.1016/j.joule.2018.03.017

23. Aher RD, Kumar BS, Sudalai A (2014) Spectrophotometric analysis of chlorophylls and carotenoids from commonly grown fern species by using various extracting solvents. Synlett 25:97–101. https://doi.org/10.1055/s-0033-1340072
24. Gu P, Yang D, Zhu X, et al (2017) Influence of electrolyte proportion on the performance of dye-sensitized solar cells. AIP Adv 7:. https://doi.org/10.1063/1.5000564
25. Esgin H, Caglar Y, Caglar M (2022) Photovoltaic performance and physical characterization of Cu doped ZnO nanopowders as photoanode for DSSC. J Alloys Compd 890:161848. https://doi.org/10.1016/j.jallcom.2021.161848
26. Ayalew WA, Ayele DW (2016) Dye-sensitized solar cells using natural dye as light-harvesting materials extracted from Acanthus sennii chiovenda flower and Euphorbia cotinifolia leaf. J Sci Adv Mater Devices 1:488–494. https://doi.org/10.1016/j.jsamd.2016.10.003
27. Abdel-Latif MS, Abuiriban MB, El-Agez TM, Taya SA (2015) Dye-sensitized solar cells using dyes extracted from flowers, leaves, parks, and roots of three trees. Int J Renew Energy Res 5:294–298
28. IK M, Uthman ISAH K (2015) The effect on extracting solvents using natural dye extracts from hyphaene thebaica for dye-sensitized solar cells. J Mater Sci Eng 05: https://doi.org/10.4172/2169-0022.1000208
29. Tamirat AG, Su WN, Dubale AA, et al (2015) Efficient photoelectrochemical water splitting using three dimensional urchin-like hematite nanostructure modified with reduced graphene oxide. J Power Sources 287:119–128. https://doi.org/10.1016/j.jpowsour.2015.04.042
30. Alhamed M, Issa A, Doubal A (2012) Studying of natural dyes properties as photo-sensitizer for dye sensitized solar cells (DSSC). J Electron Devices 16:1370–1383
31. Lim A, Haji Manaf N, Tennakoon K, et al (2015) Higher performance of DSSC with dyes from *Cladophora* sp. As mixed cosensitizer through synergistic effect. J Biophys. https://doi.org/10.1155/2015/510467
32. Kim HJ, Bin YT, Karthick SN, et al (2013) Natural dye extracted from rhododendron species flowers as a photosensitizer in dye sensitized solar cell. Int J Electrochem Sci 8:6734–6743
33. Chang H, Lo YJ (2010) Pomegranate leaves and mulberry fruit as natural sensitizers for dye-sensitized solar cells. Sol Energy 84:1833–1837. https://doi.org/10.1016/j.solener.2010.07.009

Travel time impact in multi-modal mode: a case study of the new railway line Ban Phai to Nakhon Phanom

Khanuengnij Prakhammin[1*], Nawapon Nakharutai [2], Manad Khamkong [2]
[1]PhD Degree Program in Applied Statistics, Faculty of Scienes, Chiang Mai University
[2] Department of statistics, Chiang Mai University, Chiang Mai and Thailand

**Corresponding Author Email: Khanuengnij_pr@cmu.ac.th*

Abstract. Setting a new rail network is one of important projects that that have been developed to support public transportation in Thailand. A new railway line (Ban Phai-Nakhon Phanom) which is a double-track train will be finished in 2027 [1]. However, there is no study to research the effects of this rail line. This study examined the impact of travel time on the Ban Phai-Nakhon Phanom railway line. Comparing with taking any combination of public transportations, we showed that combining a conventional rail, which is a single-track train, from Bangkok can reduce the travel time to any destination along Ban Phai-Nakhon Phanom rail line. Therefore, it might be worthy to upgrade the single-track railway in Bangkok to be a double-track in order to further decreasing travel time. Exploiting the travel time on Ban Phai-Nakhon Phanom rail line, we studied the impact of travel time between provinces above and below the railway. To do so, we considered are different travel options (single-track trains, double-track trains, cars and buses) and a combination of these options and compare their travel times. The results showed that traveling with the new double-track train spent less or equal to the travel time of driving a car except the case of Khonkaen to Yasothon due to the train station is far away from a bus station that travellers need to transfer. On the other hand, comparing with taking a bus, which is a traditional public transport in this area, we found that traveling with the new double-track train extremely save the travel time. Thus, it might be suitable to have a public transportation linking between train and bus stations or to expand the double-track railway toward the cities along the main line. Therefore, the new double-track railway has advantage on reducing travel time, but it can be further beneficial if it can be corrected to other travel modes such as a bus station.

Keyword: travel time; double-track; railway; multi-modal travel

1. Introduction

Infrastructure is the physical structure necessary for communities and people to use and facilitate the public.There are many projects to develop transportation, energy, and digital infrastructures. Developing transportation will be an important foundation for building the country's competitiveness [2]. At present, Thailand's transportation system has planed to develop infrastructure throughout the country for the benefit of linking different regions together and to serve as the transportation hub of the Asian and other regions of the world in

the near future. Some of regions have already started to implement according to the development plan [2].

Railway system development is another important policy of the government that has assigned the State Railway of Thailand to act urgently to support transportation policies due to the fact that the rail system is a low-cost mode of transportation. In addition, the development of the current long distance railway network also supports the development of the area by linking special economic zones and promoting tourism and the Thai economy towards a better future[2]. The survey in 2018 by [1] found that the proportion of routes of the current long-distance railway network is divided into 91% single track, 6% double track, and 3% triple track. By 2024, it is expected to see that the current ratio of long distance railway network will be adjusted to 33% single track 65% double-track and 2% triple-track [1]. Changing single-track to double-track, trains can run in a opposite direction without having to wait for each other to pass. Consequently, travel time can be significantly reduced. The question is what is an impact of reducing travel time from this double-track.

Travel time, as a critical performance measure of transportation efficiency and service quality for travelers [3, 4], plays an important role in transportation systems and vehicle systems [5, 6]. The measurements of travel time are not consistently employed due to data problems, compulsory reporting standards, and other challenges [7]. There are several studies that have focused on the study of travel time. For example, Bates et al., [8], Carrion and Levinson, [9], Small, [10] studied travel time reliability. In their works, time reliability is usually related to the day-to-day variation of travel time for a certain link, route, or network. Salonen and Toivonen [11] presented a comparison of different travel time measures that covers both travel times using private vehicles and public transportation. Additionally, Büchel and Corman [12] Proposed the model on the travel time of road-based public transport that can help directly improve current passenger service and operating efficiency

In Thailand, the government decied to build a new railway line from Ban Phai to Nakhon Phanom) in 2013 and is expected to be completed in 2027.Meanwhile, the railway from Bangkok to Ban Phai has been adjusted to be a double-track railway line but not all services have been fully operaterd [13].In this study, we will investigate the travel time impact on upgrading single-track trains to be double-track trains from Bangkok to Nakhon Phanom passing Ban Phai route Moreover, we will study the impact of travel time on different travel models between provinces. above and below the railway. To do so, we compared different public transport options including single-track trains, the new double-track trains, cars and buses with a mixed use of these options. We consider the provinces where the new rail line will pass, and the provinces that the new rail line will not pass but still benefit from the new railway line. The scope of this study will be divided into content as follows.

(1) Comparison of the travel time of a round trip from Bangkok to Nakon Phanom by a with single-track trains, double-track trains, cars and buses with transffering at Ban Phai.
(2) Comparison of travel times for provinces that are not served on new railways but these province ares above or below the railways by using multi-modal travel which is defined as a mode of travel that combines double-track trains with cars, buses, and the conventional train (if applicable). Then,we compare the travel time between several options on multi-modal travel.

The rest of this study is organized as follows. Section 2 provides the methodology including the defining of defining the path of studied route, the multi-modal mode and studied route, data, and studying the travel time model. Section 3 and Section 4 present results and discussion. Section 5 concludes the study and outlines future work.

2. Methodology

2.1 Defining the path of studied route

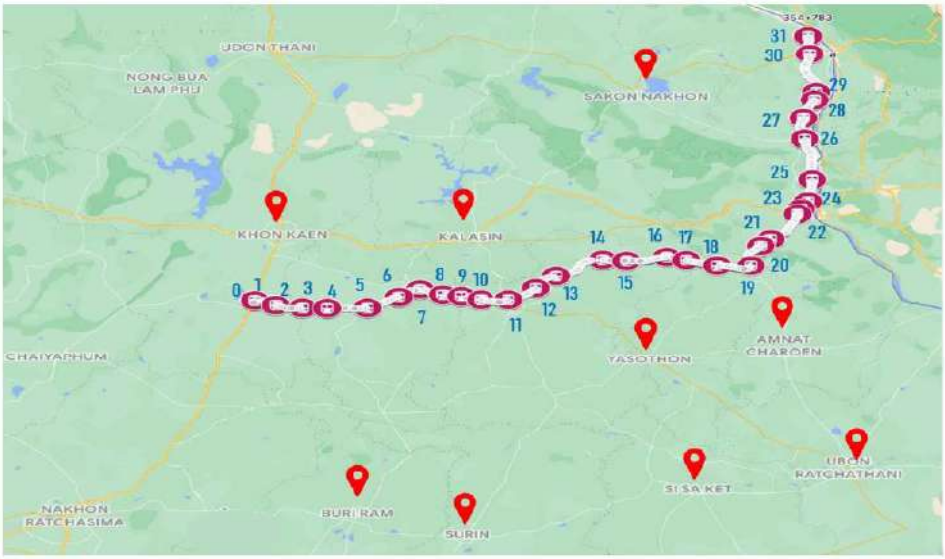
In this study, we focus the double-track new railway line from Ban Phai to Nakhon Phanom. Figure 1 shows the New Railway Line (blue line) and Northeastern Railway Line (yellow line) which is a single-track railway. The new railway has 31 stations which are 1 junction 18 stations and 12 Railway Halt [14]. and detail of station as shown in Figure 2.

Figure 1. Northeast Railway Line and New Railway Line in 2027.
Source: Johomaps Networks. (2011). [Rail map of Thailand].
http://www.johomaps.com/as/thailand/thailandrail.html

When the Ban Phai to Nakhon Phanom railway is completed, a total of four trains will be opened, namely:

(i) Express train takes about 3 hours and 30 minutes from Ban Phai to Third Thai-Lao Friendship Bridge (Nakhon Phanom) [14].
(ii) Rapid train1 takes about 4 hours and 30 minutesfrom Ban Phai to Third Thai-Lao Friendship Bridge (Nakhon Phanom) [14].
(iii) Rapid train2 takes time from Ban Phai to Roi Et is approximately 1 hours and 20 minutes [14].
(iv) Local train is train that stop at every station and stopping place. Travel time from Ban Phai to Third Thai-Lao Friendship Bridge (Nakhon Phanom) is about 5 hours and 10 minutes [14].

We choose the transit route according to the station where the express train stops. There are a total of 8 stations, namely Ban Phai, Maha Sarakham, Roi Et, Loeng Nok Tha, Mukdahan, That Phanom, Nakhon Phanom and Friendship Bridge 3 (Nakhon Phanom) [14]. There is a bus of the starting province passing through to reach the destination province.



Stations

0. Ban Phai	7.Maha Sarakham	14.AmphoePhon Thong	21.Ban Pong Daeng*	28.Na Ton*
1.Nong Waeng Rai	8.Khwao*	15.Moei Wadi*	22.Mukdahan	29.Ban Klang*
2.Phu Lek	9.Si Somdet*	16.Nong Phok	23.Ban Dan Kham*	30.Nakhon Phanom
3.Na Pro*	10.Si Kaeo*	17.Khok Sawang*	24.Second Thai-Lao Frienship Bridge	31.Third Thai-Lao Frienship Bridge
4.Kut Rung	11.Roi Et	18.Hong Saeng*	25.Wan Yai	
5.Borabue	12.Chiang Khwan	19.Loeng Nok Tha	26.That Phanom	
6.Nong No*	13.Pho Chai	20.Nikhom Kham Soi	27.Renu Nakhon	

* Railway Halt.

Figure 2. The new rail line route from Banphai to Nakhon Phanom and province around a new railway line zone.
Source: Google Map. (2022). [The new rail line route of Northeast from Banphai to Nakhon Phanom]. https://www.google.co.th/maps/.

Traveling for a round trip from Bangkok to Banphai there are four trains, but we chose to study trains that travel similarly to driving a car and taking a bus that takes the least time from Bangkok to Ban Phai [15] as shown in Figure 3(a), which is the Bangkok Express Train Nong Khai No. 77 and Express Train No. 78 for a return trip [15] in figure 3(b).

(a) (b)

Figure 3. (a) Route from Bangkkok to Khon Kaen by cars and bus.
Source: Google Map. (2022). [The line route from Bangkok to Khonkaen (Banphai)]. https://www.google.co.th/maps/.
(b) Northeastern Railway Map.
Source: State Railway of Thailand. (2017). [The rail line route of Northeast, Thailand]. https://www.facebook.com/pr.railway/

2.2 Defining multi-modal mode and studied route

We would like to investigate whether one can benefit from this new railway if they live in those provinces. To do so, we defined multiple modes of transport so that those provinces can take advantage of the new railway. From the origin you can travel by cars, buses or single-track trains (if applicable) and then transfer to a double-track train to reach the nearest city in order to change for other mode until arriving at the destination as shown in Figure 4.

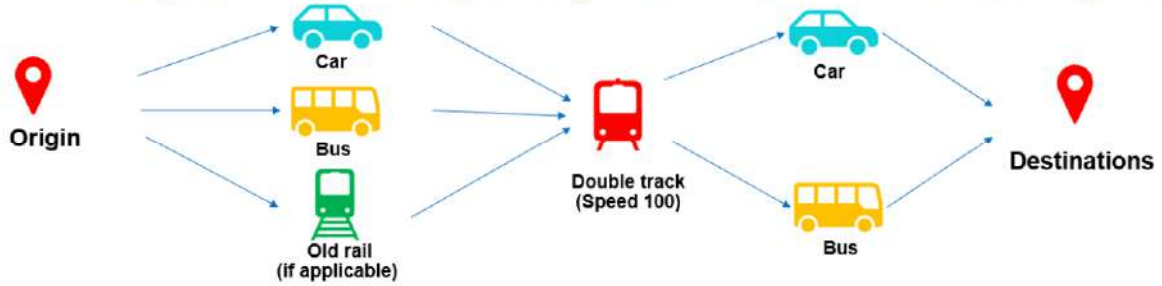


Figure 4. Transit of multi-modal travel.

There are a lot of several combination routes between provinces above and below this railway system. In this study, we select some of theseis routes to explicitly study:specificity study at follow;

(a) Route 1 starts at Khon Kaen, take a car to Yasothon, Buriram, Surinthorn, Amnat Charoen, Sisaket and Ubon Ratchathani.
(b) Route 2 starts from Ubon Ratchathani to Khon Kaen, Kalasin and Sakon Nakhon.

In fact, there is the old rail from Ubon Ratchathani to Kalasin and Sakon Nakhon this route did not take advantage of the new railway so that for we do not include this route in our study.

2.3 Data

Table 1 shows different travel time of different mode from various sources. Train travel times are obtained from the timetable of trains number 77 and number 78 . Car travel times are calculated by Google Maps from origin to destination. Bus travel time is obtained from the bus schedule on the website of the Department of Transport, Khon Kaen Province . Finally, the travel time of a double-track train is calculated as the distance divided by the speed . Note that we select the travel time of bus routes that follows the considered train route and exclude the routes that are totally different the the considered train.

Table 1 Variables in Travel Time information.

Variables	Travel Time Information
1. Sigle-track trains, rail (no.77), and rail(no.78) Old rail	Northeastern Railway Timetable from the State Railway of Thailand [15].
2. Cars	Total travel time from Google map website [16].
3. Buses	Bus schedule from Khon Kaen Provincial Transport Office [17] and the bus website [18, 19, 20, 21].
4. Double-track*	Travel time from the distance when setting different speeds [2, 16, 17].

*Double-track trains have speeds at 100, 110 and 120 kilometers per hour.

In this study, we have the limitation of the travel time mode as follows. We cannot obtain exact waiting time when the trains access or egress at the train station and the transit between the bus station and the train station. Therefore, we exclude the waiting time and the transit time of these cases and we only consider the actual travel time spending on the travel mode.

2.4 To study travel time model

Huang et al. [22] defined the objective function to optimize the train timetable under the condition that the number and type of trains are determined. The model below is to minimize the total traveling time of all trains:

$$\min z = \sum_{k \in K} (TA_m^k - TD_1^k), \qquad (1)$$

Where $TA_m^k - TD_1^k$ is the traveling time of train k between the initial station 1 and terminal station m.

TA and TD are traveling time of arrival time and departure time.

We use equation (1) to find the total travel time of different transportation vehicles and compare travel times along from Bangkok to Nakhon Phanom.

Based on Llorca et al. [23], we proposed a model to compare the traditional travel time (car or bus) with a multi-modal travel of transport along the new railway line.

$$tt'_{o-d} = \min(tt^{m}_{o-d}, tt^{multi\,modal}_{o-d}) \qquad (2)$$

Where tt'_{o-d} is the travel time in the double track scenario between o and d.

tt_{o-d}^{m}	the travel time of the travel mode m (car or bus) in the base scenario between o and d.
o and d	are origin and destination, respectively.

$$tt_{o-d}^{multimodal} = tt_{o-i}^{m} + tt_{i-j}^{m} + tt_{j-d}^{m} \quad (3)$$

Where	$tt_{o-d}^{multimodal}$	is the travel time of the multi-modal travel from o to d.
	tt_{o-i}^{m}	is the travel time of the travel mode m from o to i.
	tt_{i-j}^{m}	is the travel time of the travel mode m from i to j.
	tt_{j-d}^{m}	is the travel time of the travel mode m from j to d.
	m	is travel vehicle there are a car, a bus and a conventional train (if applicable).
	i and j	are starting and ending point of that transit travel mode, respectively.

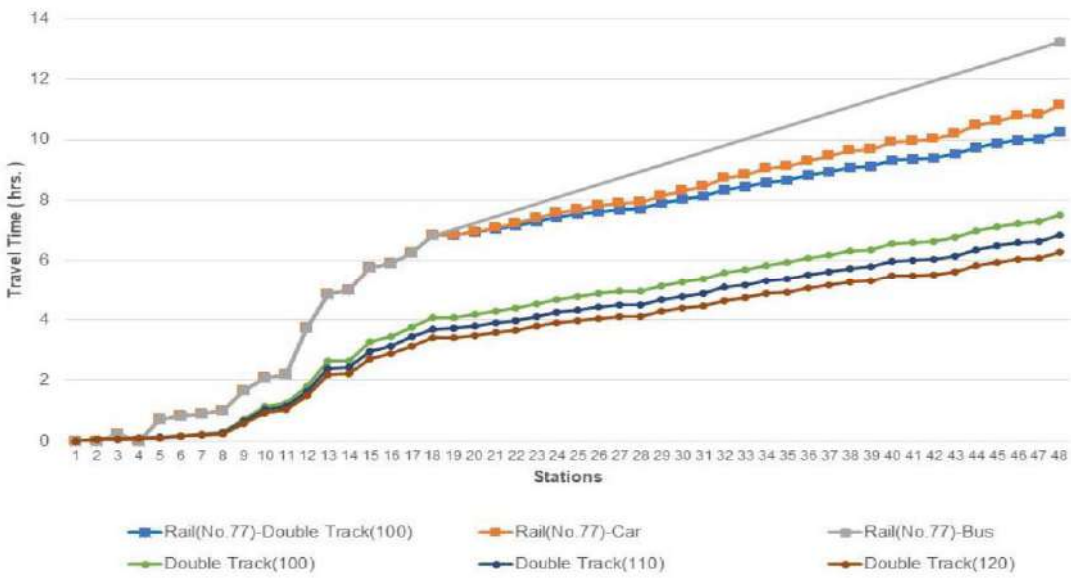
We calculate the shortest travel time of multi-modal travel (cars, buses, single- track trains and double-track trains) in equation (2) by using Microsoft Excel 2019 on the travel time.

3. Results and discussion

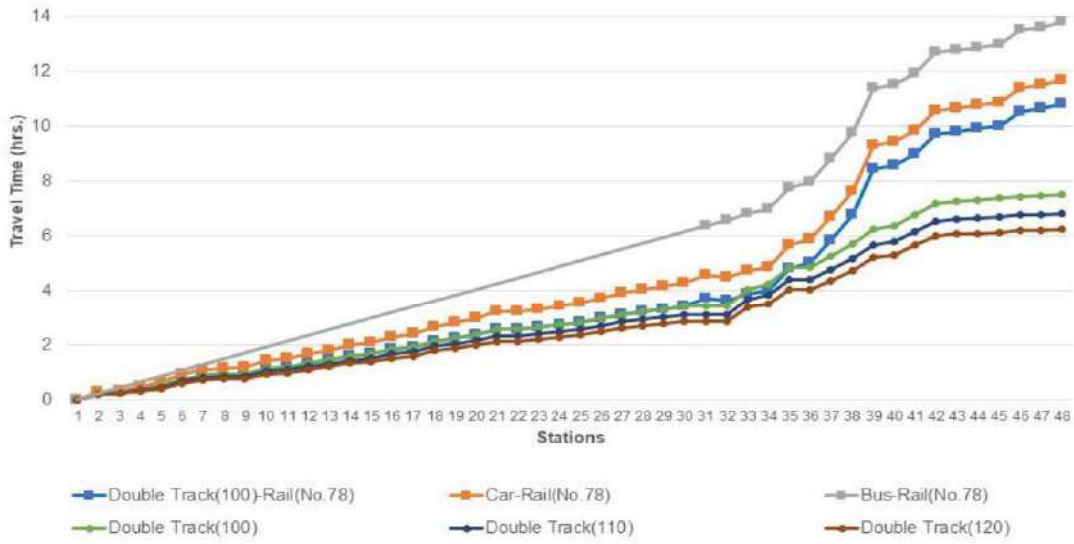
Results

3.1 Comparison the travel time of a round trip from Bangkok to Nakon Phanom.

Consider the travel time of a round trip from Bangkok to Nakon Phanom in Figure 5 (a) and (b), we found that the travel time on double-track train is the lowest time among possible combination of travel modes. In addition, the double-track train can reduce the travel time from Bangkok to any destination along Ban Phai-Nakhon Phanom railway line. Therefore, the new railway line will benefit to people who live around the train stations since they can travel faster to any destination along Ban Phai-Nakhon Phanom railway line.



(a)



(b)

Figure 5. (a) Travel times from Bangkok to Nakhon Phanom by using conventional train (No.77) and with mixed either Double-track trains, Cars and Buses compared to travel times of the same route by Double-track trains with different speed in the bracket. See Table A in the appendix for the list of the railway stations.
(b) Travel times from Nakhon Phanom to Bangkok by using conventional train (No.78) with either double-track trains, cars and buses compared to travel times of the same route by double-track trains with different speed in the bracket. See Table B in the appendix for the list of the railway stations.

3.2 Comparison of the multi-modal travel with cars and buses.

In this section, we will investigate the impact of the new railway on the travel time for people who live the province above or below the new railway in order to see whether they can reduce the travel time by using the double-track trains or not. To do so, we suppose that we start at Khonkaen and then travel to other six provinces (Yasothon, Burirum, Surin, Amnat charoen, Sisaket, and Ubon ratchathani) below the new railway. The possible options to travel are as follows.

(i) We can travel by a car from Khonkaen to other six provinces.
(ii) We can drive a car from Khonkaen to the nearest station of the bouble-track railway and take the bouble-track train to the nearest station of the destination before driving to the destination by a car.
(iii) We can do the similar strategy in (ii) except taking a single-track train Khonkaen to the nearest station of the bouble-track railway.
(iv) We can do the similar strategy in (ii) except taking a bus at the station to the destination.

We calculate the travel time of these possible options and present the results in figure 6. According to the result, traveling with the new double-track train can spend less or equal time to other options.

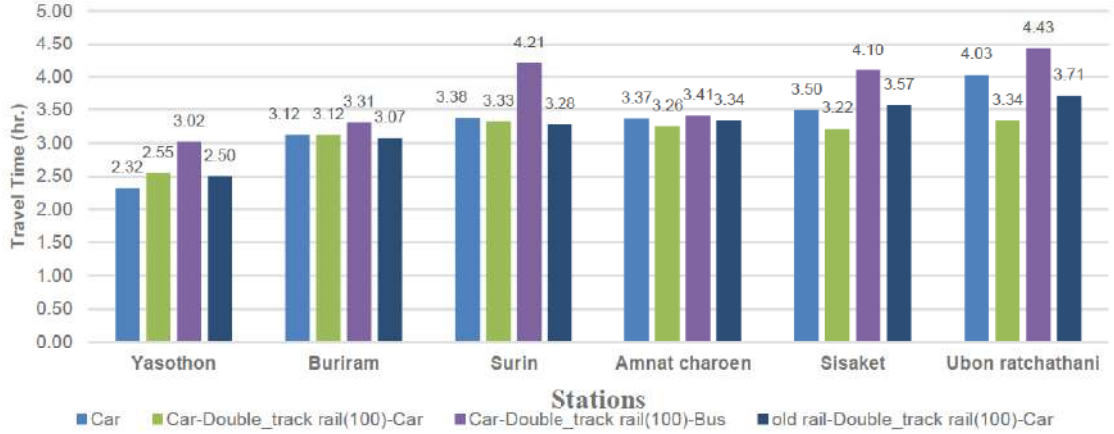


Figure 6. Travel time from Khon Kaen to six provinces on different travel options with cars.

On the other hand, suppose that we start at Khonkaen and then travel to other six provinces Yasothon, Burirum, Surin, Amnat charoen, Sisaket, and Ubon ratchathani) with other different four options, specifically,

(i) We can take a bus from Khonkaen to other six provinces.
(ii) We can take a bus from Khonkaen to the nearest station of the bouble-track railway and take the bouble-track train to the nearest station of the destination before taking another bus.
(iii) We can do the similar strategy in (ii) except taking a single-track train Khonkaen to the nearest station of the bouble-track railway.
(iv) We can do the similar strategy in (ii) except driving a car from the train station to the destination.

The travel times of these travel options are shown in figure 7. We found that traveling with the new double-track train extremely reduce the travel time except in the case of Khonkaen to Yasothon, where the train station is far away from city center so that travellers need time to transfer.

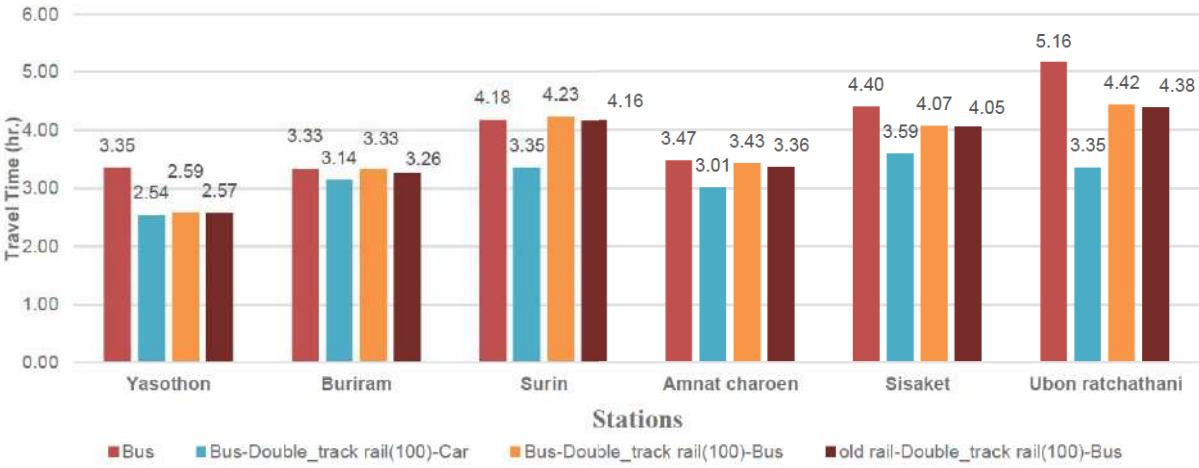


Figure 7. Travel time from Khon Kaen to six provinces on different travel options with buses.

In addition, we suppose that we start at Ubon Rachathani and then travel to other three provinces (Kalasin, Sakon Nakhon, and Ubon Ratchathani) above the new railway. The possible options to travel are as follows.

(i) We can take a car or a bus from Ubon Ratchathani to other three provinces.
(ii) We can take a car or a bus from Ubon Ratchathani to the nearest station of the bouble-track railway and take the bouble-track train to the nearest station of the destination before driving to the destination by a car or taking another bus.

Figure 8 presents the travel time of these travel options. We found that traveling with multi-modal travel spent time less or equal to the travel time of driving a car and traveling by a bus.

(a) (b)

Figure 8. Travel time from Khon Kaen to three provinces.
(a) on different travel options with cars.
(b) on different travel options with buses.

Discussion

The section describes the impact of travel time on a new railway line from Ban Phai to Nakhon Phanom by using mode travel there are single-track trains, double-track trains, cars and buses. A comparison study of the travel time of a round trip from Bangkok to Nakon Phanom by transferring at Ban Phai showed that the travel time on double-track train is the lowest travel time among possible combination of travel modes. This means that the double-track train can reduce the travel time from Bangkok to any destination along Ban Phai-Nakhon Phanom railway line. Therefore, the new railway line will have the impact on travel time since passengers can travel faster from Bangkok to any destination along Ban Phai-Nakhon Phanom railway line.

In addition, a comparison study of the multi-modal travel with cars and buses found that traveling with the new double-track train extremely reduce the travel time except in the case of Khonkaen to Yasothon, where the train station is far away from city center so that passengers need time to transfer. This result showed that the double-track train can reduce the travel time to nearby cities along the Ban Phai-Nakhon Phanom railway line.

4. Conclusion

In this paper, we studied the impact of travel time that can be reduced on the new double-track railway line from Ban Phai to Nakhon Phanom. We found that combining the new double-track railway with the single-track train can reduce the travel time from Bangkok to any destination along the Ban Phai-Nakhon Phanom railway line. Therefore, it might be worthy to change the single-track railway in Bangkok to Ban Phai to be a double-track railway in order to further decrease travel time. Traveling in a multimodal, we found that the provinces above and below the new railway would take less time. Moreover, we found that the bus will take the most travel time. Maybe it might be suitable to have a public transportation link between train and bus stations or to expand the double-track railway toward the cities along the main line.

In the future, we can extend our work to consider other factors, for example, the distance, waiting time, the number of public transport that have to take, and the cost of travel. Statistical methods such as linear programming model or multinomal logit moel can be used in these future studies.

Acknowledgments

The work described in the paper was support by the National Science and Technology Development Agency (NSTDA).

References

[1] State Railway of Thailand. 2020. The double-track railway. *SRT Journal*,2020(6), pp. 2-6.

[2] Ministry of Transport. 2014. Strategy for the development of Thai transport infrastructure 2015-2022 and transport action plan, urgent phase 2015. Bangkok: Ministry of Transport.

[3] Erhardt, G. D., Freedman, J., Stryker, A., Fujioka, H., & Anderson, R. 2007. Ohio long Distance travel model. *Transportation Research Record: Journal of the Transportation Research Board No.2003*, pp 130–138.

[4] Chen, C., A. Skabardonis, and P. Varaiya. 2003.Travel-Time Reliability as a Measure of Service. *Transportation Research Record No.1885*, pp. 74–79.

[5] Bates, J., Ashley, D. and Hyman, G.1987. The Nested Incremental Logit Model: Theory and Application to Modal Choice-Transportation Planning Method. *Proceeding of Seminar C Held at The PTRC Summer Annual Meeting*, BATH, England.

[6] Bates, J. 1988. Econometric Issues in Stated Preference Analysis. *Journal of Transport Economics and Policy*, 22(1), pp. 56-69.

[7] Tang, L., Ghader, S., Mishra, S., Zhang, L. 2015. Impact of travel reliability on mode choice: empirical analysis with observed travel time variation data in Washington. *Transportation Research Board 94th Annual Meeting,* Washington DC, United States.

[8] Bates, J., Polak, J., Jones, P., Cook, A. 2001. The valuation of reliability for personal travel. *Transport. Res. Part E: Logist*,37(2), pp.191–229.

[9] Carrion, C., Levinson, D. 2013. Valuation of travel time reliability from a GPS-based experimental design. *Transport Research Part C*,35(2013), pp.305–323.

[10] Small, K.A. 1999. Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation. Transportation Research Board.

[11] Salonen, M., Toivonen, T. 2013. Modelling travel time in urban networks: comparable measures for private car and public transport. *Journal of Transport Geography*,31, pp.143-153.

[12] Büchel B. and Corman F. 2020. Review on Statistical Modeling of Travel Time Variability for Road-Based Public Transport. *Frontiers Built Environ*, 6(70), pp.1-14.

[13] Prachachat.net Online. 2021. Expropriation of the double trackway "Ban Phai-Nakhon Phanom", a 66 billion baht train to open the country connecting the AEC. https://www.prachachat.net/(Retrieved 10 November 2021).

[14] Thailand Infracstructure. 2020. Railway construction project Ban Phai, Maha Sarakham, Roi Et, Mukdahan, and Nakhon Phanom line. Facebook. https://www.facebook.com/ (Retrieved 15 November 2021).

[15] State Railway of Thailand. 2020. *SRT Timetabl*. Retrieved 20 November 2021 from https://www.railway.co.th/Station/StationList

[16] Google. n.d. Google Maps directions. Retried form https://www.google.co.th/maps/
[17] Provincial Land Transport Office Of Khonkaen.2022. bus lane. https://www.dlt.go.th/site/khonkaen/ (Retrieved 10 December 2021).
[18] Mukdahan City Hall. 2022. Double-track railway construction project, Ban Phai, Maha Sarakham, Roi Et, Mukdahan, Nakhon Phanom. https://www2.nakhonchaiair.com/ (Retrieved 16 September 2021)
[19] Nakhonchai Air. 2022. Bus routes from Bangkok to Nakhon Phanom. from https://www2.nakhonchaiair.com/(Retrieved 25 November 2021).
[20] SAHAMIT UBON. 2022. Bus schedule. http://www.sahamit.net/ (Retrieved 28 November 2021)
[21] Tickets online. 2021. Book bus tickets online. https://xn--n3cc8act3d5k.com/ (Retrieved 10 November 2021)
[22] Huang, Z. et al. 2021. Optimizing Train Timetable Based on Departure Time Preference of Passengers for High-Speed Rails. *Journal of Advanced Transportation*, pp.1-20.
[23] Llorca, C., Ji, J., Molloy, J., & Moeckel, R. 2018. The usage of location based big data and trip planning services for the estimation of a long-distance travel demand model; predicting the impacts of a new high-speed rail corridor. *Transportation Economics*,72, pp. 27- 36.
[24 Li, Z., Hensher, D.A., Rose, J.M. 2010. Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. *Transportation Research Part E*, 46 (3), pp.384–403.
[25] Li, Z., Hensher, D. 2011. Prospect theoretic contributions in understanding traveller behavior: a review and some comments. *Transport Review*, 31 (1), pp. 97–115.
[26] Li, H., Tu, H., Hensher, D. 2016. Integrating the mean-variance and scheduling approaches to allow for schedule delay and trip time variability under uncertainty. *Transportation Research Part A*. 89, pp.151–163.
[27] Castillo,E., Conejo, AJ., Castillo,C., Mínguez,R. 2007.Closed formulas in local sensitivity analysis for some classes of linear and non-linear problems. 15(2), pp.355–371.
[28] Kumar, A. 1980. Use of Incremental Form of Logit Models in Demand Analysis. *Transportation Research Record*, 775, pp.21-27.
[29] Johomaps Networks. 2011. Rail Map of the world. http://www.johomaps.com/as/thailand/thailandrail.html (Retrieved 25 October 2021).
[30] Koppelman, F. 1983. Predicting Transit Ridership in Response to Transit Service Changes. *Journal of Transportation Engineering*,109, pp.548-564.
[31] Preston, J. 2015. Demand forecasting for new local rail stations and services. *Journal of Transport Economics and Policy*, 25(2), pp. 183-202.
[32] Rich, J., Mabit,S.L. 2012. A long-distance travel demand for Europe. *Journal of Transport and in fracture Research*,12(1),pp.1-20.
[33] Ties Brands et al. 2014. Modelling Public Transport Route Choice, with Multiple Access and Egress modes. *Transportation Research Procedia*, 1, pp. 12 – 23.
[34] Zhao, L., & Shen, L. 2019. The impacts of rail transit on future urban land use development: A case study in Wuhan, China. *Transport Policy*, 81, pp.396-405.

Appendix

Table A Name train stations from Bangkok (Train No.77) to Nakhon Phanom.

No.	Stations	No.	Stations	No.	Stations
0	Bangkok	16	Muang Phon	32	Moei Wadi
1	Sam Sen	17	Ban Phai	33	Nong Phok
2	Bang Sue Junction	18	Nong Waeng Rai	34	Khok Sawang
3	Nikhom Rotphai Khomo Sip-et	19	Phu Lek	35	Hong Saeng
4	Bang Khen	20	Na Pro	36	Loeng Nok Tha
5	Lak Si	21	Kut Rung	37	Nikhom Kham Soi
6	Don Muang	22	Borabue	39	Ban Pong Daeng
7	Rangsit	23	Nong No	40	Mukdahan
8	Ayutthaya	24	Maha Sarakham	41	Ban Dan Kham
9	Saraburi	25	Khwao	42	Second Thai-Lao Frienship Bridge
10	Kaeng Khoi Junction	26	Si Somdet	43	Wan Yai
11	Pak Chong	27	Si Kaeo	44	That Phanom
12	Nakhon Ratchasima	28	Roi Et	45	Renu Nakhon
13	Thanon Chira Junction	29	Chiang Khwan	46	Na Ton
14	Muang Khong	30	Pho Chai	47	Ban Klang
15	Bua Yai Junction	31	Amphoe Phon Thong	48	Nakhon Phanom

Table B Name train stations from Nakhon Phanom to Bangkok(Train No.78).

No.	Stations	No.	Stations	No.	Stations
0	Nakhon Phanom	16	Amphoe Phon Thong	32	Bua Yai Junction
1	Ban Klang	17	Pho Chai	33	Muang Khong
2	Na Ton	18	Chiang Khwan	34	Thanon Chira Junction
3	Renu Nakhon	19	Roi Et	35	Nakhon Ratchasima
4	That Phanom	20	Si Kaeo	36	Pak Chong
5	Wan Yai	21	Si Somdet	37	Kaeng Khoi Junction
6	Second Thai-Lao Frienship Bridge	22	Khwao	39	Saraburi
7	Ban Dan Kham	23	Maha Sarakham	40	Ayutthaya
8	Mukdahan	24	Nong No	41	Rangsit
9	Ban Pong Daeng	25	Borabue	42	Don Muang
10	Nikhom Kham Soi	26	Kut Rung	43	Lak Si
11	Loeng Nok Tha	27	Na Pro	44	Bang Khen
12	Hong Saeng	28	Phu Lek	45	Nikhom Rotphai Khomo Sip-et
13	Khok Sawang	29	Nong Waeng Rai	46	Bang Sue Junction
14	Nong Phok	30	Ban Phai	47	Sam Sen
15	Moei Wadi	31	Muang Phon	48	Bangkok

Question Generation from Thai Wiki paragraphs by using Transformer

Napat Srichantapong[1], Wachirawut Thamviset[1,*]
[1] Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand

**Corresponding Author Email: twachi@kku.ac.th*

Abstract. Question Generation (QG) is one of the tasks in Natural Language Processing (NLP) that allows computers to understand the text and human language, which is important in education. The purpose of QG is to generate interrogative sentences similar to humans. However, there are very few researches related to generating Thai language questions. In this study, we proposed a method for generating a Thai *wh*-question (what, when, where, who, and which) with a multilingual variant of T5, a model in a Transformer architecture whose efficiency is state-of-the-art on many multilingual benchmarks covering 101 languages, including Thai. We used the iApp Wiki QA dataset for training, validation, and test. We generated questions using Thai paragraphs from the iApp Wiki QA dataset and evaluated the performance of the questions with precision and recall. Finally, we evaluated the diversity of generated questions in two methods: 1) normal input paragraphs and 2) splitting a sentence into sub-sentences. Both methods have the number of questions equal to 5788. We evaluated both methods on Self-BLEU4 and the percentage of repetition of questions. As a result, Method 1 obtained Self-BLEU4 of 19.01 and repetition of 42.03%, and Method 2 obtained Self-BLEU4 of 17.43 and repetition of 20.97%.

Keywords: Question Generation; Natural language processing; Natural language understanding; Thai language processing

Introduction

Questions are the beginnings of the human cognitive process. If the human is asked a question, an effective method of finding the answer will be analyzed. Furthermore, question generation is a strategy that encourages learning to improve thinking skills, interpretation, reflection, and the organization of human learning processes. Moreover, it leads to the development of academic knowledge. Types of questions that help develop understanding, such as problem questions and analytical questions. The question should design effectively, which requires an expert to create questions. However, the creation of human questions has limitations on the quality of the diversity of questions and the time burden of design thinking.

In the last decade, there has been the development of generating automatic questions with artificial intelligence to make it similar to humans. It reduces the burden of designing questions and helps develop self-learning. It can apply to tasks such as intelligent learning systems that allow humans to create questions and answers to assess their knowledge [1], and QG is one [2] of the tasks of Natural Language Processing (NLP).

However, there are very few researches related to generating questions in Thai languages, and there are differences between Thai and English languages. For example, there are no spaces to divide words and full stop (.). Therefore, generating Thai sentences is much more complicated than in English.

In this study, Our goal is to introduce a method for generating a Thai *wh*-question (what, when, where, who, and which) model with a multilingual variant of T5 (mT5) [3].

In evaluation, we would like to measure the performance of the model in generating a diversity of questions using Self-BLEU4 [4] and the percentage of repetition of questions, which we generated questions using Thai paragraphs from the iApp Wiki QA dataset [5] and compare the generated questions from our model and target questions from the iApp Wiki QA test dataset with precision and recall. After that, we selected the best score model to generate questions in two methods: 1) normal input paragraphs, and 2) splitting a paragraph into sub-sentences. Then, we took questions from both methods to compare to find the best way to generate a diversity of questions. Finally, we conclude the experiment results and recommend improvements to our model.

Related work

1.1. Model architecture

Ashish Vaswani et al. [6] mentioned the problem with the Seq2Seq model, which processed word by word. As a result, it could cause some data to be lost. Therefore, they introduced the transformer architecture to solve this problem, improving computational efficiency and accuracy in the Seq2Seq model.

Therefore, we selected the multilingual variant of T5 (mT5) [3], a model in a Transformer architecture whose efficiency is state-of-the-art on many multilingual benchmarks covering 101 languages, including Thai. Furthermore, it has been proposed to improve computational efficiency and accuracy in Sequence to sequence models (Seq2Seq).

1.2. Thai question generation

Chonlathorn Kwankajornkiet et al. [7] introduce automatic multiple-choice fill-in-the-blank QG in Thai text by separating the keywords from sentences and adding spaces in place of the keywords as question phrases. Keywords were used as correct answers and to create distractor. Distractors used the same category of word selections as the 3 choice keywords from WordNet. Therefore, the question phrases, answers, and distractors were composed as fill-in-the-blank questions

1.3. English question generation

R.Mitkov et al. [8] presented generating multiple-choice test items. Questions were generated from electronic documents. The subject gave that answer and distractor to find "coordinate" from WordNet. WordNet is the lexical database of semantics where hypernyms and coordinates are related words in the same category as other words. The word with the best 3 distractor score was subsequently selected to create distractors. They combined sentence, answer, and distractor as question phrases.

Susmita Gangopadhyay et al. [1] introduced an automatic question generation system with a focus generator module (FGM) based on a deep learning model. The model selects the best focus sentences to generate practical questions and answers. Exam-styled questions generated from automatic question generation are mostly "wh"("what," "who," and "where") or reading comprehension types. Questions were generated to be similar to humans.

Luis Enrico Lopez et al. [9] used transformer architecture for automated question generation tasks, which describe the sequence-to-sequence model (Seq2Seq) as having a complex architecture and method. The transformer has been proven to simplify methods and achieve faster efficiency than Seq2Seq model. They used a generative pre-trained transformer 2 (GPT-2). They fine-tuned a model with SQuAD and generated questions as output from the input context.

Methodology

The method consists of 2 main steps: 1) Model creation and 2) Evaluation, as shown in Figure 1.

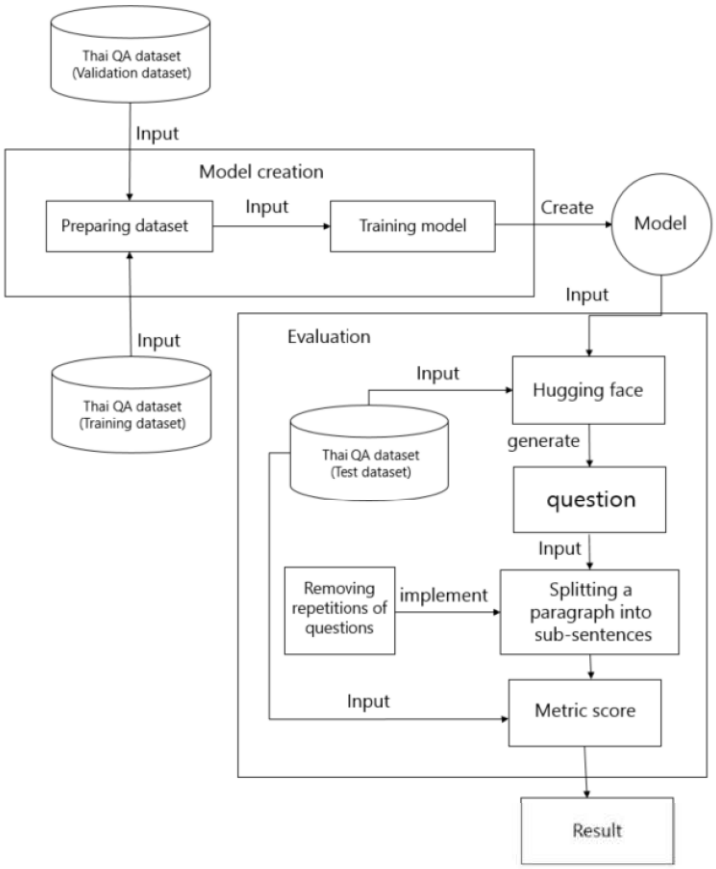


Figure 1. Illustration of the proposed overall QG.

1.4. Preparing dataset

We trained a model for learning to analyze paragraphs and texts on iApp Wiki QA datasets [5].

iApp Wiki QA dataset is the Stanford question answering dataset from Thai Wikipedia, published by iApp Technology Company Limited. The dataset consists of 9154 datasets:

- Training dataset: 7290 datasets consist of 5761 questions and 1529 paragraphs, as shown in Table 1.
- Test dataset: 933 datasets consist of 739 questions and 192 paragraphs, as shown in Table 2.
- Validation dataset: 931 datasets consist of 742 questions and 191 paragraphs, as shown in Table 3.

Each dataset was formatted as one question per line (OQPL) [9], as shown in Table 4. However, the dataset used paragraphs from Wikipedia, which contained special characters. Therefore, we used PyThaiNLP [10] to clean the data :

- To remove special character: \n, (), {}, and [].
- To remove duplicate letters or duplicate words: "เต็มไปหมดเลยอะอะอะอะ (It is full)" to "เต็มไปหมดเลยอะ"

After that, we input the cleaned data in order, as shown in Figure 2.

Table 1. A total of 7290 training datasets, 1529 paragraphs, each paragraph containing multiple questions, and each interrogative sentence was unique.

Number of questions	Number of pargraphs	Total of data for training
1	111	111
2	236	472
3	259	777
4	390	1560
5	493	2465
6	13	78
7	6	42
8	2	16
9	2	18
10	5	50
11	2	22
12	3	36
14	2	28
15	1	15
17	2	34
18	1	18
19	1	19
Total	1529	5761

Table 2. A total of 933 test datasets, 739 paragraphs, each paragraph containing multiple questions, and each interrogative sentence was unique.

Number of questions	Number of pargraphs	Total of data for Test
1	10	10
2	27	54
3	35	105
4	56	224
5	57	285
6	1	6
7	1	7
8	1	8
9	1	9
10	2	20
11	1	11
Total	192	739

Table 3. A total of 931 validation datasets, 742 paragraphs, each paragraph containing multiple questions, and each interrogative sentence was unique.

Number of questions	Number of pargraphs	Total of question for validation
1	16	16
2	28	56
3	34	102
4	39	156
5	65	325
7	1	7
8	1	8
9	3	27
10	2	20
11	1	11
14	1	14
Total	191	742

Table 4. One question per line (OQPL).

Paragraphs	Target Questions
Paragraph 1: พัทธ์ธีรา ศรุติพงศ์โภคิน(เกิด 3 ธันวาคม พ.ศ.2533) หรือชื่อเล่นว่า อร เป็นนักแสดงหญิงชาว ไทย สำเร็จมัธยมศึกษาจาก Catholic Cathedral College ประเทศนิวซีแลนด์และปริญญาตรี จาก Raffles International College สาขา Business Marketing เข้าสู่วงการตั้งแต่ อายุ 6 ขวบ จากการแสดงละครเวทีกับ ครูชล ประคัลภ์จันทร์เรือง จากนักเล่นโฆษณาในวัยเด็ก 2- 3 ชิ้น และยังเคยแสดงช่วงละครสั้น ในรายการ ซุปเปอร์จิ๋ว ประมาณปี2542 ปัจจุบันเป็นทั้ง นักแสดง , พิธีกร และ วีเจ อยู่ที่คลื่นเก็ท 102.5 Bangkok International Hits Music Station และยังเป็นพิธีกรให้กับ ช่อง ทรูมิวสิก (Patthira Sarutiphongpokin (born in 3rd December 1990), and nicknamed Orn, is a Thai actress. She graduated high school from Catholic Cathedral College, New Zealand, and has a bachelor's degree from Raffles International College in Business Marketing. She entered the industry at the age of 6 years from theatrical perfor-mances with Teacher Chonprakun Chanruang, then took part in 2-3 commercials in childhood and also acted in short dramas. Like in the program Super Jiw back in 1999. She is currently both an actor, MC and VJ at Kluen Get 102.5 Bangkok International Hits Music Station and is also an MC for True Music)	Q1: พัทธ์ธีราศรุติพงศ์โภคิน เกิดวันที่เท่าไร (What date was Patthira Sarutiphongpokin born ?) Q2: พัทธ์ธีราศรุติพงศ์โภคิน มีชื่อเล่นว่าอะไร (What is Patthira Sarutiphongpokin's nickname?) Q3: พัทธ์ธีรา ศรุติพงศ์โภคิน ทำอาชีพอะไร (What does Patthira Sarutiphongpokin do?) Q4: พัทธ์ธีรา ศรุติพงศ์โภคิน จบการศึกษาจากประเทศอะไร (What country did Patthira Sarutiphongpokin graduate from?)
Paragraph 2: ท่านผู้หญิงสุภรภ์เพ็ญ หลวงเทพนิมิตหรือชื่อเดิม สุภร เพ็ญ หลวงเทพ (สกุลเดิม ณ พัทลุง) รองราช เลขานุการในพระองค์สมเด็จพระนางเจ้าสิริกิติ์ พระบรมราชินีนาถ สำนักราชเลขาธิการและ ผู้รับผิดชอบโรงฝึกศิลปาชีพฯ สวนจิตรลดา ท่านผู้หญิงสุภรภ์เพ็ญ สมรสกับพันเอกศิรชณ์ รงค์หลวงเทพ นิมิต (Lady Supanraphen LuangThepmimit, formerly Suphornpen Luang Thep (formerly Na Phatthalung), Deputy Royal Secretary of Her Majesty Queen Sirikit Her Majesty the Queen works at the Office of the Royal Secretariat as the person in charge of the Chitralada Suan Chitralada Training Center. Lady Supornphen, is married to Colonel Sirnarong Luang Thepnimit.)	Q1: สุภรภ์เพ็ญ หลวงเทพนิมิต ชื่อเดิมว่าอะไร (What is the former name of Suparaphen Luang Thepnimit?) Q2: สุภรภ์เพ็ญ หลวงเทพนิมิต มีตำแหน่งอะไร (What is the position of Suparaphen Luang Thepnimit?) Q3: สุภรภ์เพ็ญ หลวงเทพนิมิต เป็นรองเลขานุการ ของใคร (Who is the deputy secretary of Suparaphen Luang Thepnimit?) Q4: สุภรภ์เพ็ญ หลวงเทพนิมิต เป็นรองเลขานุการ ของสำนักอะไร (Supornphen Luang Thepnimit is the deputy secretary of what office?)

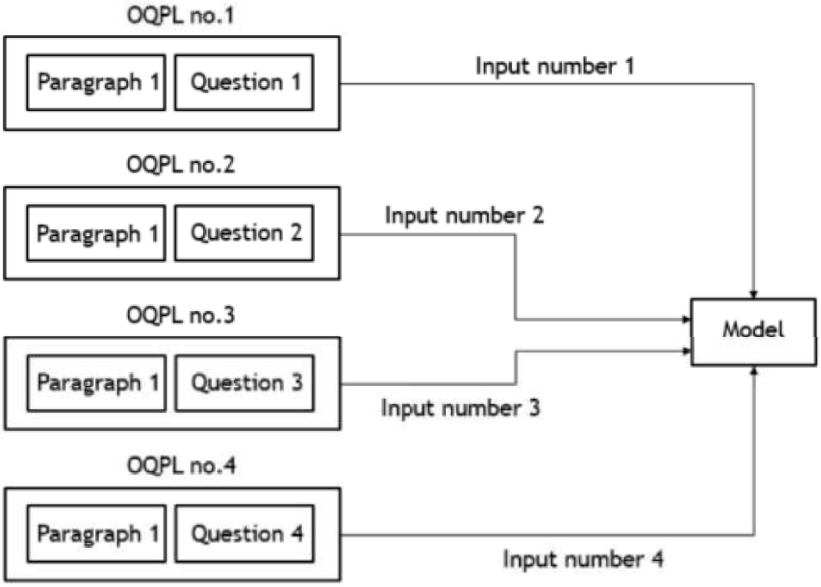


Figure 2. Input one question per line (OQPL) in order to model

1.5. Training model

For training, we selected a massively multilingual pre-trained text-to-text transformer-base (mT5-base) [3] which is a transformer neural network model type. We trained a model to generate wh-question in Thai from the input dataset. We used google cloud graphics processing units (Google cloud GPU) from google colab [11] for training.

We trained a model with 7290 training datasets. However, there was a problem with insufficient GPU memory. We adjusted the hyperparameter and reduced the amount of training dataset. Which GPU was able to work with 2500 training datasets and did not reduce the validation dataset.

We optimized the mode by training 10 epochs, and each epoch had a batch size of 8. We used the Adam optimizer [12] and lr_find() function to get the best learning rate with 3 x 10-4. We also set up mT5-base's default maximum sequence length of 512.

We used the base models because of hardware limitations since They could not work with larger and more complex models, as shown in Table 5.

Table 5. The number of parameters of each a massively multilingual pre-trained text-to-text transformer

Types of a massively multilingual pre-trained text-to-text transformer	Parameters
mT5-small	300 million
mT5-base	580 million
mT5-large	1.2 billion
mT5-xl	3.7 billion
mT5-xxl	13 billion

Evaluation and Result

Similar to the work of Luis Enrico Lopez et al. [9], we performed automatic evaluation metrics:

- BLEU metric [13] depends on precision only, which evaluates n-gram from the translation.
- ROUGE_L [14] depends on recall only.
- METEOR [15] depends on precision only, which evaluate all word from the translation hypothesis.

We also used the Self-BLEU4 [4] metric to evaluate the diversity of the generated questions. The lower Self-BLEU, the more diversity is shown, and we created a removing repetition of questions algorithm for counting the percentage of repetition of questions.

1.6. Removing repetitions of questions

We needed to remove repetitions of questions (RRoQ). We used the removing repetitions of questions algorithm by taking the question that the model generated as input, as shown in Algorithm 1.

Algorithm 1: Removing repetitions of questions (RRoQ)

```
1: Input: Q
   // Q = the number of questions
   // P = previous data
   // C = current data
2. Output: output[]: an array of unique question
3: initialize P = "", C = "", output = []
4: for index in range(Q): // looping equals the number of questions
5:     C = Q[index]
6:     if C not equal to P:
7:         output.add(C)
8:     P = C
9: end for
```

We took a question to calculate the percentage of repetition of questions by using the equation:

The percentage of repetition of questions = ((Total of Generated Question - RRoQ) x 100) / (Total of Generated Question)

1.7. Optimizing QG model

We generated questions by inputting paragraphs into our model. We implemented our model with Hugging Face [16], a machine for implementing the mT5 model with parameters to increase the efficiency of generating questions for our model.

We adjusted Hugging Face's parameters. We used nucleus sampling [17] to select the sum of probability is >= top_p, which should be between 0 < top_p <= 1. Therefore, we needed to find the best value of top_p to evaluate a diversity of questions. We start from 0.1 at the end of round 1, increasing it to 0.1 every round until 1.0.

We also set up the hugging face's parameter with a max length of 30 and maximum lengths of question sentences. We took paragraphs and questions from the iApp Wiki QA test

dataset, which used paragraphs as input and target questions as references to evaluate metrics. We generated the number of questions equal to the number of Target questions in the paragraphs. Each top-p generated the same number of questions with 739 questions. We took them to evaluate a metric's score, as shown in Table 6 and Table 7.

Table 6. Each metric scores of top-p from 0.1 to 1.0.

Top-p	BLEU_1	BLEU_2	BLEU_3	BLEU_4	ROUGE_L	METEOR
0.1	66.87	52.94	43.60	36.85	41.74	37.25
0.2	66.29	52.48	43.28	36.54	41.82	37.48
0.3	67.16	53.45	44.23	37.47	41.86	38.70
0.4	67.42	53.57	44.46	37.85	42.07	40.48
0.5	65.95	51.64	42.30	35.32	42.79	41.92
0.6	67.20	52.73	43.50	36.39	42.97	44.91
0.7	65.11	50.69	41.79	34.89	42.90	45.93
0.8	65.44	50.81	41.65	34.43	43.15	45.93
0.9	64.50	48.96	39.55	32.25	42.08	46.14
1.0	62.38	45.21	35.43	28.17	41.38	44.76

Table 7. Evaluating Self-BLEU4 and Repetition

Top-p	Self-BLEU4	Repetition %
0.1	20.07	73.61
0.2	19.11	68.60
0.3	19.08	61.97
0.4	19.05	51.15
0.5	18.90	41.27
0.6	18.60	31.93
0.7	18.30	21.78
0.8	18.12	16.77
0.9	18.28	17.50
1.0	18.30	17.77

As a result, we got an excellent BLEU score metric in the top-p range under 0.7. However, we got the highest ROUGE_L and METEOR in the 0.8 range. To ensure that the model could generate a diversity of questions. We also evaluated with Self-BLEU4 and Repetition.

Because a score above 0.8 makes the question sentence too rigid, the sum of the sentences must be greater than or equal to that range, causing no diversity. After that, we got the best top-p of 0.8. We tested to generate 10 questions, as shown in Table 8

Table 8. Generating questions by inputting paragraphs and using the number of question generation of 10.

paragraphs	Generated Questions
ถนนลำพูนไชย เป็นถนนสา สั้นๆ ในย่านเยาวราช พื้นที่แขวงตลาดน้อยและแขวงสัมพันธวงศ์ เขตสัมพันธวงศ์ กรุงเทพมหานคร มีจุดเริ่มต้นที่ถนนพระราม ที่ 4 ไปทางทิศตะวันตกเฉียงใต้ ผ่านทางแยกลำพูนไชยที่เป็นจุดตัดกับถนนเจริญกรุง และไปสิ้นสุดที่ถนนเยาวราชบริเวณใกล้กับวงเวียนโอเดียน ถนนลำพูนไชยเป็นที่ตั้งของสถาบันกวดวิชาดาว้องก์ และเมื่อปี พ.ศ. 2493 ยังเคยเป็นที่ตั้งของสำนักงานใหญ่ธนาคารกรุงศรีอยุธยา (Lamphun Chai Road is a short road in Yaowarat, Talat Noi and Samphan-thawong sub-districts, Samphanthawong District, Bangkok. The road starts at Rama IV Road, going southwest through the Lamphun Chai intersection at the intersection with Charoen Krung Road and going to end at Yaowarat Road, near the Odean Circle. The Dawong Tutoring Institute is located at Lamphun Chai Road and in the year 1950, the Bank of Ayudhya head office was also located at the road)	Q1: ถนนลำพูนไชย เป็นถนนสายสั้นๆในย่านอะไร Q2: ถนนลำพูนไชยมีจุดเริ่มต้นของถนนอะไร Q3: ถนนลำพูนไชย มีจุดเริ่มต้นที่ถนนอะไร Q4: ถนนลำพูนไชย มีจุดเริ่มต้นถนนพระรามที่ 4 ไปทางทิศตะวันตกเฉียงใต้ ผ่านแยกอะไร Q5: ถนนลำพูนไชย มีจุดเริ่มต้นอะไรบ้าง Q6: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร Q7: ถนนลำพูนไชย มีจุดเริ่มต้นของสำนักงานใหญ่ธนาคารอะไร Q8: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร Q9: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร Q10: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร (Q1: Lamphun Chai Road is a short road in what area? Q2: What road does Lamphun Chai road start from? Q3: What road does Lamphun Chai start at? Q4: Lamphun Chai Road has the starting point of Rama IV Road going southwest, through what junction? Q5: What is the starting point of Lamphunchai Road? Q6: Lamphun Chai Road is a short road in what area? Q7: What is the starting point for Lamphun Chai Road? Q8: Lamphun Chai Road is a short road in what area? Q9: Lamphun Chai Road is a short road in what area? Q10: Lamphun Chai Road is a short road in what area?)

1.8. Splitting a paragraph into sub-sentences

Our model generated a question by selecting a keyword. For example, from Table8, The model was interested in the word 'ถนนลำพูนไชย (Lamphunchai Road).' Because when we input paragraphs into our model. The model always pays attention to the first word. Therefore, we used the splitting a paragraph into a sub-sentences method to allow the model to select more words of interest. We used a Sentence tokenizer (ST) from PyThaiNLP to split the paragraphs into sub-sentences.

We used ST to split the paragraphs before putting them into our model. Then, we took the sub-sentence from ST that input to the model to generate questions. Finally, we generated the number of questions using the number of target questions in paragraphs, as seen in Table 9.

Table 9. It generated questions by splitting the paragraph into sub-sentences using a sentence tokenizer. Each sub-sentences needed to generate equal to the number of target questions in paragraphs. For example, The below paragraph has 3 target questions.

Paragraphs	Using ST	Questions
ถนนลำพูนไชย เป็นถนนสา สั้นๆ ในย่านเยาวราช พื้นที่แขวงตลาดน้อยและแขวงสัมพันธวงศ์ เขตสัมพันธวงศ์ กรุงเทพมหานคร มีจุดเริ่มต้นที่ถนนพระรามที่ 4 ไปทางทิศตะวันตกเฉียงใต้ ผ่านทางแยกลำพูนไชยที่เป็นจุดตัดกับถนนเจริญกรุง และไปสิ้นสุดที่ถนนเยาวราชบริเวณใกล้กับวงเวียนโอเดียน ถนนลำพูนไชยเป็นที่ตั้งของสถาบันกวดวิชาดาว้องก์ และเมื่อปี พ.ศ. 2493 ยังเคยเป็นที่ตั้งของสำนักงานใหญ่ธนาคารกรุงศรีอยุธยา (Lamphun Chai Road is a short road in Yaowarat, Talat Noi and Samphan-thawong sub-districts, Samphanthawong District, Bangkok. The road starts at Rama IV Road, going southwest through the Lamphun Chai intersection at the intersection with Charoen Krung Road and going to end at Yaowarat Road, near the Odean Circle. The Dawong Tutoring Institute is located at Lamphun Chai Road and in the year 1950, the Bank of Ayudhya head office was also located at the road)	Sentence 1: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านเยาวราช พื้นที่แขวงตลาดน้อยและแขวงสัมพันธวงศ์ เขตสัมพันธวงศ์ กรุงเทพมหานคร (Lamphun Chai Road is a short road in Yaowarat, Talat Noi and Samphan-thawong sub-districts, Samphanthawong District, Bangkok)	Q1: ถนนลำพูนไชย เป็นถนนสายสั้นๆในย่านอะไร Q2: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร Q3: ถนนลำพูนไชย เป็นถนนสายสั้นๆ ในย่านอะไร (Q1: Lamphun Chai Road is a short road in what area? Q2: Lamphun Chai Road is a short road in what area? Q3: Lamphun Chai Road is a short road in what area?)
	Sentence 2: มีจุดเริ่มต้นที่ถนนพระรามที่ 4 ไปทางทิศตะวันตกเฉียงใต้ ผ่านทางแยกลำพูนไชยที่เป็นจุดตัดกับถนนเจริญกรุง และไปสิ้นสุดที่ถนนเยาวราชบริเวณใกล้กับวงเวียนโอเดียน ถนนลำพูนไชยเป็นที่ตั้งของสถาบันกวดวิชาดาว้องก์ (The road starts at Rama IV Road, going southwest through the Lamphun Chai intersection at the intersection with Charoen Krung Road and going to end at Yaowarat Road, near the Odean Circle. The Dawong Tutoring Institute is located at Lamphun Chai Road)	Q1: ถนนลำพูนไชย มีจุดเริ่มต้นถนนพระรามที่ 4 ไปทางทิศตะวันตกเฉียงใต้ ผ่านทางแยกอะไร Q2: ถนนลำพูนไชยอยู่ทิศไหน Q3: ถนนลำพูนไชยเป็นที่ตั้งของสถาบันกวดวิชาดาว้องก์ และไปสิ้นสุดที่ถนนอะไร (Q1: Lamphun Chai Road has the starting point of Rama IV Road going southwest, through what junction? Q2: Which direction is Lamphun Chai Road? Q3: Dawong Tutoring Institute is located at Lamphun Chai Road . And where does the road end?)
	Sentence 3: และเมื่อปี พ.ศ. 2493 ยังเคยเป็นที่ตั้งของสำนักงานใหญ่ธนาคารกรุงศรีอยุธยา (and in the year 1950, the Bank of Ayudhya head office was also located at the road)	Q1: สำนักงานใหญ่ธนาคารกรุงศรีอยุธยาคืออะไร Q2: สำนักงานใหญ่ธนาคารกรุงศรีอยุธยายังเคยเป็นที่ตั้งของสำนักงานอะไร Q3: สำนักงานใหญ่ธนาคารกรุงศรีอยุธยาปัจจุบันเคยเป็นที่ตั้งของสำนักงานอะไร (Q1: Where is Bank of Ayudhya Headquarters? Q2: What was the location of Bank of Ayudhya Headquarters? Q3: What is the current location of Bank of Ayudhya Headquarters?)

We evaluated the performance of our model by generating the number of diverse questions in 2 ways: 1) normal input paragraphs (paragraphs not via splitting paragraphs into sub-sentences), and 2) splitting paragraphs into sub-sentences. We used a test dataset from the iApp Wiki QA dataset with 192 paragraphs as input. We needed to generate the same number of generated questions from both methods by having method 2 generate questions first and take the number of questions to method 1, as shown in Figure 3.

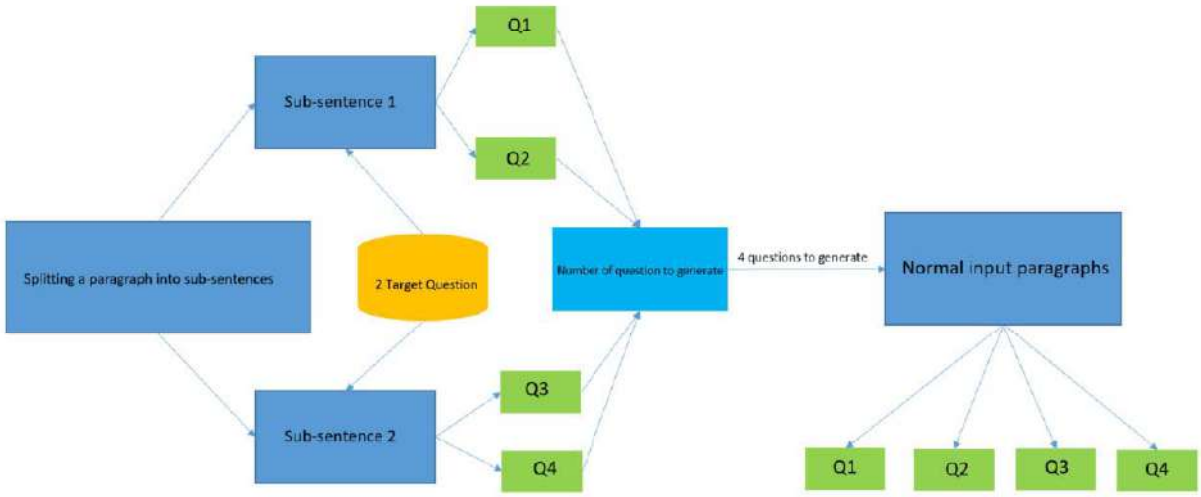


Figure 3. Generating the same number of question

We obtained 5788 generated questions from both methods and need to pass them through the removing repetitions of questions algorithm to the number of unique questions from both methods, as shown in Table 10.

Table 10. Comparing the number of questions from 2 ways: 1) normal input paragraph, and 2) splitting a paragraph into sub-sentences

Methods	Number of paragraphs	Number of generated questions	Final outputs (Using removing repetitions of questions)
Normal input paragraphs	192	5788	3355
Splitting a paragraph into sub-sentences	192	5788	4574

We took 2 methods to evaluate the metric score, as shown in Table 11 and Table 12.

Table 11. Evaluating BLEU, ROUGE_L, and METEOR score from 2 methods.

Methods	BLEU_1	BLEU_2	BLEU_3	BLEU_4	ROUGE_L	METEOR
Normal input paragraphs	64.20	49.33	39.90	32.66	40.62	37.25
Splitting a paragraph into sub-sentences	40.64	29.87	22.94	17.55	42.62	45.57

Table 12 . Evaluating Self-BLEU4 and Repetition score from 2 methods.

Methods	Self-BLEU4	Repetition %
Normal input paragraphs	19.01	42.03
Splitting a paragraph into sub-sentences	17.43	20.97

The results showed that splitting a sentence into sub-sentences (method 2) got meteor 45.57 and ROUGE_L 42.62 more than a normal input paragraph (method 1). However, method 1 got more BLEU than method 2 because method 2 produced more diversity of questions, resulting in a lower n-grams score. If method 2 had a high BLEU score, there would be no diversity of questions. To ensure, we evaluate both methods on Self-BLEU4 and Repetition. It is proven that method 2 could produce more diversity of questions, with Self-BLEU4 of 17.43. The lower Self-BLEU, the more diversity is shown.

Conculsion and Recommendations

This paper proposed a method for generating a Thai *wh*-question model. We selected mT5-base, which is a type of transformer architecture for generating a model. Then, we focused on generating the number of diverse questions in Thai with our model.

Furthermore, we used metrics scores to evaluate our model, such as BLEU, ROUGE_L, and METEOR. We also used Self-BLEU4 and the percentage of repetition of questions to evaluate a model that can generate a diversity of questions.

Our evaluation was tested using 2 ways: 1) normal input paragraph (method 1) and 2) splitting a paragraph into sub-sentences (method 2). Both methods produced the number of 5788 generated questions. Therefore, we evaluated questions from both methods.

As a result, method 2 got METEOR 45.57 and ROUGE_L 42.62 more than method 1, including Self-BLEU4 17.43 and Repetition 20.97. Although, method 1 got more BLEU than method 2. This means method 2 could generate more diversity of questions, resulting in a lower n-grams score. On the other hand, If method 2 had a high BLEU score, there would be no diversity of questions.

We conclude that the model worked well for generating a diversity of questions from paragraph input. However, using short paragraphs or sentences could produce more diversity of questions.

this paper didn't compare other baselines. Because there are very few researches related to generating questions in Thai languages, For future work, we would like to recommend 2 things:

1) To compare mT5 with WangchanBerta [18], We selected the mT5 model with higher outcome scores than previous multilingual models. However, there are disadvantages in that the mode is not designed for the Thai language specifically. Therefore, we did more research and found a model. WangchanBERTa is said to be a pre-train model designed for Thai work directly. And work effectively with the Thai languaHowever, therefore has never been a comparison between the mT5 and WangchanBERTa. Consequently, we pay attention to and compare the two models for question generation.

2) Experiment with larger, more complex models and more data sets. Because of limited resources, we cannot create a fully efficient model. Therefore, it must be proven next time.

Acknowledgments
This study was supported by Department of Computer Science, College of Computing, Khon Kaen University.

References

[1] Gangopadhyay S, S.M. R. Focused Questions and Answer Generation by Key Content Selection. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM) [Online] Delhi, India: IEEE; 2020 [cited 2022 Mar 25]. p. 45-53. Available from: https://ieeexplore.ieee.org/document/9232485
[2] Rus, V., Cai, Z. and Graesser, A., 2008. Question generation: Example of a multi-year evaluation campaign. Proc WS on the QGSTEC.
[3] Xue L, Constant N, Roberts A, Kale M, Al-Rfou R, Siddhant A, Barua A, Raffel C. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. arXiv:201011934 [cs] [Online] 2021 [cited 2022 Mar 21]; . Available from: https://arxiv.org/abs/2010.11934
[4] Zhu, Y., Lu, S., Zheng, L., Guo, J., Zhang, W., Wang, J. and Yu, Y., 2018, June. Texygen: A benchmarking platform for text generation models. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 1097-1100).
[5] Open Thai Wikipedia QA Dataset made by iApp Technology [Online]. [cited 2022 Mar 25]. Available from: https://github.com/iapp-technology/iapp-wiki-qa-dataset
[6] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I. Attention Is All You Need. arXiv:170603762 [cs] [Online] 2017 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/1706.03762
[7] Kwankajornkiet C, Suchat A, Punyabukkana P. Multiple-Choice Question Generation from Thai Text. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) [Online] Khon Kaen, Thailand: IEEE; 2016 [cited 2022 Mar 25]. p. 1-6. Available from: https://ieeexplore.ieee.org/document/7748891
[8] Mitkov R, Ha L, Karamanis N. A computer-aided environment for generating multiple-choice test items 2006; 12(2): 177-194
[9] Lowphansirikul L, Polpanumas C, Jantrakulchai N, Nutanong S. WangchanBERTa: Pretraining transformer-based Thai Language Models. arXiv:210109635 [cs] [Online] 2021 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/2101.09635
[10] pythainlp.tokenize — PyThaiNLP v2.3.0 documentation [Online]. [cited 2022 Mar 21]. Available from: https://pythainlp.github.io/docs/2.3/api/tokenize.html
[11] Program Carneiro T, NoBrega R, Nepomuceno T, Bian G, Albuquerque V, Filho P. Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications 2018; 6: 61677-61685.
[12] Kingma D, Ba J. Adam: A Method for Stochastic Optimization. arXiv:14126980 [cs] [Online] 2017 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/1412.6980
[13] Papineni, K., Roukos, S., Ward, T. and Zhu, W.J., 2002, July. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).
[14] Lin, C.Y., 2004, July. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).
[15] Denkowski, M. and Lavie, A., 2014, June. Meteor universal: Language specific translation evaluation for any target language. In Proceedings of the ninth workshop on statistical machine translation (pp. 376-380).
[16] Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P. Rault, T.. Louf R, Funtowicz M, Brew J. Huggingface's transformers: State-of-the-art natural language processing. arXiv:191003771 [cs] [Online] 2019 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/1910.03771

[17] Holtzman A, Buys J, Forbes M, Choi Y. THE CURIOUS CASE OF NEURAL TEXT DeGENERATION. arVix:190409751 [cs] [Online] 2020 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/1904.09751

[18] Lowphansirikul L, Polpanumas C, Jantrakulchai N, Nutanong S. WangchanBERTa: Pretraining transformer-based Thai Language Models. arXiv:210109635 [cs] [Online] 2021 [cited 2022 Mar 25]; . Available from: https://arxiv.org/abs/2101.09635

Smart Application for Finding Parking and Car Owners

Wiyada Yawai[1,*]
[1]Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University

**Corresponding Author Email: wiyada.y@nrru.ac.th*

Abstract. Currently, Nakhon Ratchasima Rajabhat University hosts a large number of cars each day, but on-campus parking is limited, making it time-consuming to locate and sometimes drivers are required to park in restricted areas. This causes problems in the parking lot with car blocking the entrance and preventing movement. This research developed an application to solve problems encountered in Nakhon Ratchasima Rajabhat University. There are real users of the parking spaces in all 30 buildings in Nakhon Ratchasima Rajabhat University. This application can be used for both cars and motorcycles. In a usage test with 44 participants (including administrators, faculty members, student staff, etc.) the application received an average satisfaction score of 4.06 out of 5.

Keywords: mobile application; smart application; car owners; parking application.

Introduction

Many cars use the parking lot at Nakhon Ratchasima Rajabhat University each day because parking on campus is limited, it can take a long time to find a parking space. Sometimes drivers have to park in a prohibited area. This causes a problem in which entrances and exits of parking lots are blocked, and cars are unable to enter. This application is specifically designed to solve this problem.

Mobile Application has not yet been developed to solve such problems in this specific agency. Therefore, the developer has developed this application by determining the parking space in the building and collecting parking data in each building, the number of cars parked in the building and an area that can park in each building, which this application can also use for both cars and motorcycle. We developed the application with the following objectives:

- To solve the problem of finding a parking space.
- To help in searching for car owners.

Previous Research

In the past, many vehicle parking applications were developed for serving many purposes. For instance, the application "Find My Car eLibera Parking System" (Roojai.com, 2022) that helps drivers find a place to park. This system uses GPS and Google to navigate to the parked car. It is also possible to share the user's location with the user's friends or share photos of the parking spot and send SMS to send emergency messages. This application is based on Android 2.2 and above which is used in general locations.

Thailand Highway Traffic (Roojai.com, 2022) (Bureau of Road Maintenance Administration, 2022) is an application that displays traffic information about the highways in Thailand by showing a current picture from CCTV cameras installed along the highways with the user's vehicle speed data (km/h). The flow rate of the user's vehicle (veh/h) is recorded on a graph for the past 24 hours and displayed in color. The user can search for service points such as restrooms and drinking water, along with a map showing the travel time. The application is available for both Android and IOS. OBD, Car Doctor by PNN Soft (Roojai.com. 2022), is an application that helps to check cars and shows various details about the engine's status by working through the Bluetooth system when there are defects. Car Doctor also recommends solutions for Waze users (Roojai.com. 2022) as an application that reports traffic using current technology. By crowdsourcing with real-time traffic reports, along with traffic reporting in map format and chat rooms for users using Waze in the same area, the user can chat and share real-time traffic reports. The HUDWAY-GPS Navigation HUD (Roojai.com. 2022) is an easy-to-use 3D navigation app. The user can choose a route to help calculate the distance speed. Worapon Pongphet has developed an application to help check parking buildings and vacant spaces in parking buildings within the University of Business Administration. It is used with Android operating system, GPS and Google map. This research provides building guidance and parking. by working with the use of sensor circuits and microcontrollers (Worapon Pongphet, 2012).

Montasinee Homwan et al. developed "A Car Park Searching System by WAP Service via Mobile Phone". This research developed a system to search for parking via WAP (Wireless Application Protocol) service. Shopping centers that use optical, electronic sensors to detect cars and send data through the ET-PC 8255 card into the database system in the server. The screen is displayed in Text Mode by the service users of the parking search system. The vehicle must enter the specified WAP site URL to show whether the parking location is free or not. This research is only a simulation of the parking search system (Montasinee Homwan, Komsan SiangWong and Suthat Chumphonkulwong, 2006).

Ittipol Homhoun et al. attempted to find free parking spaces and parking spaces close to areas required by visitors. To save time in finding a parking space and reducing traffic problems within Naresuan University. Data were collected from this research, including building names and parking locations. Then, the data were managed by using the QGIS program, which created a database. Then, a set of instructions was written to create a Web Map Interface in JavaScript, PHP, and HTML. The location system was connected to OpenLayers, Geoserver, and the Database Management System. To show the location of the parking place in the database. The position results were calculated every time the user searched according to specified conditions. jQuery mobile was used to develop a system to update the parking information in the smartphone system. Based on the results of the experiment, by allowing a user to choose conditions according to his or her needs, a system can support the decision of the carriage location within the dog. Colleges can help make choices more easily (Ittipol Homhoun, Chaiwat Kanfak and Pongchai Meebun, 2015).

Research Theory and Methods

This section discusses the basic principle of car license plates in Thailand, the theory of Euclidean distance, tools and application design, and detecting license plates from images which are detailed as follows.

1.1. License plates in Thailand

License plates in Thailand are divided into car license plates and motorcycle license plates, each of which details each type as follows.

1.1.1. Car license plates

At present, car license plates in Thailand are made of aluminum in 13.5 x 6 inche size. The sign surface looks reflective. Letters and numbers are embossed from the surface of the

license plate, causing the back of the license plate to be indented according to the license plate number, and the Department of Transport logo is also embossed on the bottom right corner as shown in figure 1.

	License plate for a private car with no more than 7 seats.
	License plate for a private car with more than 7 seats.
	Pickup truck license plate.
	Personal tricycle license plate.
	Inter-provincial taxi license plate.
	License plates for a trailer, a road roller, a tractor and an agricultural vehicle.
	License plates for a small truck and a two-row seat pickup truck.
	Tricycle license plate.
	Taxi license plate.
	License plates for a business service vehicle, a sightseeing vehicle and a rental car.

Figure 1. Car license plate in Thailand (Including License Plates of Each Type, 2022).

1.1.2. Motorcycle license plate

The font style of the new motorcycle license plates, as shown in figure 2, is the same as that of the license plates that replaced the old category beginning in 2012.

Figure 2. Motorcycle license plate in Thailand (Thai Car License Plates good, 2022).

1.2. Euclidean distance

Figure 3 (Michael Greenacre, 2008) is the Pythagoras' theorem applied to distances in two-dimensional space, by given a vector $x = [x_1, x_2]$ is the sum of the squares of its coordinates (see triangle OPA in figure 3 or triangle OPB – $|OP|^2$ denotes the squared length of x, that is, the distance between points O and P.) The squared distance between two vectors $x = [x_1, x_2]$ and $y = [y_1, y_2]$ is the sum of squared differences in their coordinates. (figure 3 $|PQ|^2$ denotes

the squared distance between points P and Q). To denote the distance between vectors x and y, we use the notation $d^2_{x,y}$ so that this last result can be written as equations (1):

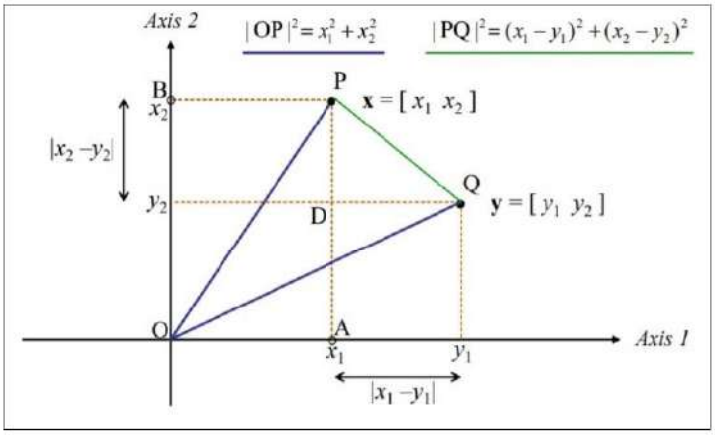


Figure 3. Pythagoras' theorem applied to distances in two-dimensional space.

$$d^2_{x,y} = (x_1 - y_1)^2 + (x_2 - y_2)^2 \tag{1}$$

That is, the distance itself, d_{x,y}, is the square root as in Equation 4.

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{2}$$

1.3. Tools and application design

1.4. Tools for Development

- MIT App Inventor 2.0 was used for development (MIT App Inventor, 2021).
- Data was stored on the Airtable Cloud platform (Airtable, 2021).
- The operating system is based on Windows 10.

1.5. System Requirements

- Android operating system.
- Internet accessible.
- Requires 6.1 MB of installation space.

1.6. License plate detection

In MIT App Inventor (Artificial Intelligence with MIT App Inventor, 2022), many artificial intelligence (AI) and Machine Learning (ML) for Image Classification use the website https://classifier.appinventor.mit.edu/ as shown in figure 4. To train the model of the license plate image data, hyperparameters were customized as shown in figure 5. Then, it was trained with the image data of car and motorcycle license plates. 50 images were used for each license plate, including 50 images of 10 car license plates, totalling 500 images, and 50 images of 10 motorcycle license plates, totalling 500 images. This included 1,000 images taken to train the model. In the model training process, learning rate 0.0001, Adam optimizer, 30 epochs and training data fraction 4 was chosen, as shown in figure 5, because this option gives the most accurate results. The results of the car license plate identification test are

shown in figure 6 and the results of the motorcycle license plate identification test are shown in figure 7.

Figure 4. Training Page by Artificial Intelligence with MIT App Inventor.

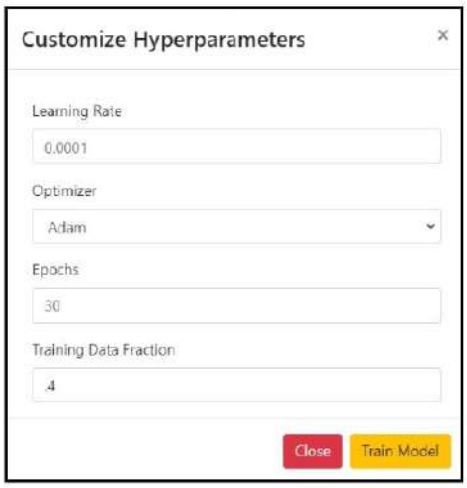


Figure 5. Customize Hyperparmeters.

Figure 6. Shows the results of the car license plate identification test.

Figure 7. Shows the results of the motorcycle license plate identification test.

1.7. Application workflow

The design of the application functionality, as shown in figure 8, starts with the users entering the username and password and then entering the system to find parking spaces according to function numbers 1 and 2, after that users will get all information about application functions as shown in function number 3. When users want to find a parking space, they can use function number 4, which will display all registered parking building and how many motorcycle and car parking spaces are empty. When users want to select a parking lot, they can press on the name of parking lot and then press the parking button. And when users want to get out of the car, press the leaving button in order to confirm their parking locations. In the event that the user forgets where their cars are parked, they can use function as shown in function number 5 and can calculate how many meters away from their car. If there is another car blocking the exit or parking in a prohibited parking area, users can use the funtion number 6 to check who the car owner is. In case they forget their parking location, they can find their car by filling out a license plate number or take a picture of the license plate to find the owner's name and phone number to call the car owner.

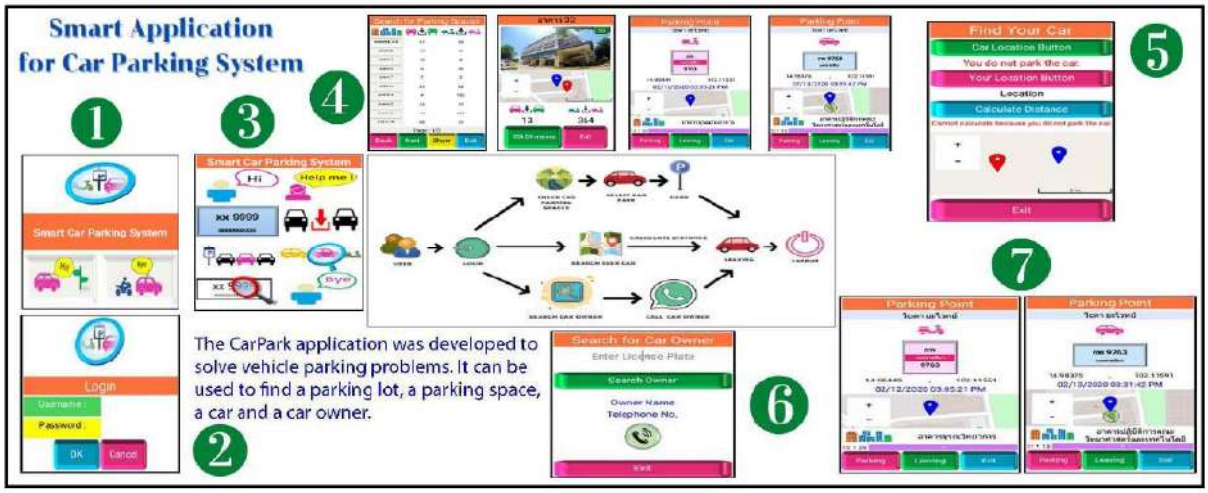


Figure 8. Show application workflow.

1.8. Data collections

Data collection uses airtables (*Airtable, 2021)*, which are stored in the cloud system. It is divided into 3 tables as follows:

1.8.1. Collecting data for the parking area in each building consisting of the following fields:

- Building code
- Building name
- Initial latitude
- Last latitude
- Initial longitude
- Last latitude

1.8.2. Set vehicle information according to the details as follows

- License plate
- Type of car
- Car color
- Car manufacturer
- Province code
- Year of manufacture of the car
- Category car

1.8.3. Specify vehicle owner information including fields as follows:

- Owner code
- Name of the car owner
- Organization name
- Phone number

1.9. The statistics used in the research

The statistics used in the research were the percentage, mean and standard deviation. The statistics were compared (dependent t-test) (Chakri Thamman and Manit Asanok, 2018) by using the results and the following evaluation criteria.

- The mean range; 4.51-5.00, indicates the highest level.
- The mean range; 3.51-4.50, indicates a high level.
- The mean range; 2.51-3.50, indicates the average level.
- The mean range; 1.51-2.50, indicates a low level.
- The mean range; 1.01-1.50, indicates the lowest level.

Results and discussion

This research was conducted by test users using the actual application at 30 buildings within Nakhon Ratchasima Rajabhat University. Using both cars and motorcycles, the total number of user assessors was 44 including executives, lecturers, students and external personnel which can be classified by gender, as shown in figure 9, divided by the type of users, as shown in figure 10. The average application satisfaction rating was 4.06, which is a high level. By referring to the (Chakri Thamman and Manit Asanok. 2018) evaluation criteria, details are shown in Table 1.

1.10. User evaluation results

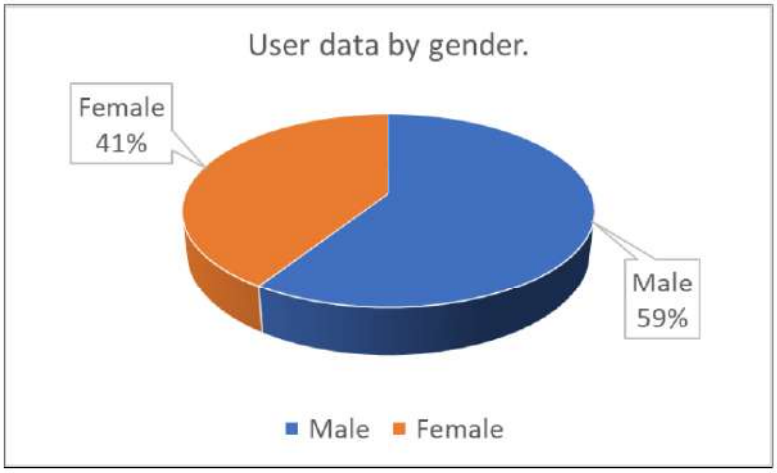


Figure 9. User data by gender.

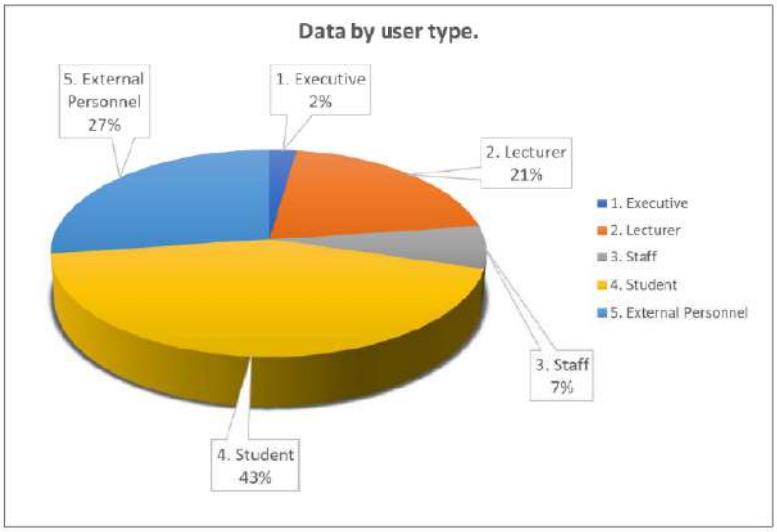


Figure 10. User data by user type.

Table 1. Evaluation results from users of application

Topic	$\bar{X}$	*SD*
1. The beauty, modernity and interesting nature of the application.	3.65	1.24
2. Easy-to-use menu.	3.89	1.17
3. It is easy and convenient to access information quickly.	3.93	1.07
4. The language used is correct. Able to convey meaning and understand correctly.	4.05	1.01
5. The accuracy of the information.	4.09	0.86
6. Find a park space.	4.30	0.79
7. Find the location of a parked car.	4.27	0.90
8. Find the owner of the car.	4.32	0.80
Average	4.06	0.98

1.11. Detection Results

The training images were 50 images of each license plate, divided into 10 motorcycles, a total of 500 images, and a total of 500 images of 10 cars.

1.11.1. Motorcycle license plate detection results

The total of 468 images of motorcycle license plates were tested, 439 correct and 29 were wrong, as shown in the figure 11 and table 2 shows a comparison of the motorcycle license plate detection results and show the results of motorcycle license plate detection in figure 12.

Table 2. Comparison of the motorcycle license plate detection results

Image No.	Motorcycle license Plate	Correct	Not Correct	Total	Accuracy (%)
1	1กภ 1321	33	2	35	94.29
2	3กษ 1234	38	3	41	92.68
3	กกก 999	40	0	40	100.00
4	กพษ 652	43	8	51	84.31
5	กลค 810	50	1	51	98.04
6	กษบ 44	33	7	40	82.50
7	ทร 8989	54	2	56	96.43
8	ปตฉ 813	34	3	37	91.89
9	วจท 508	56	3	59	94.92
10	วรพ 777	58	0	58	100.00
Average accuracy (%)					93.51

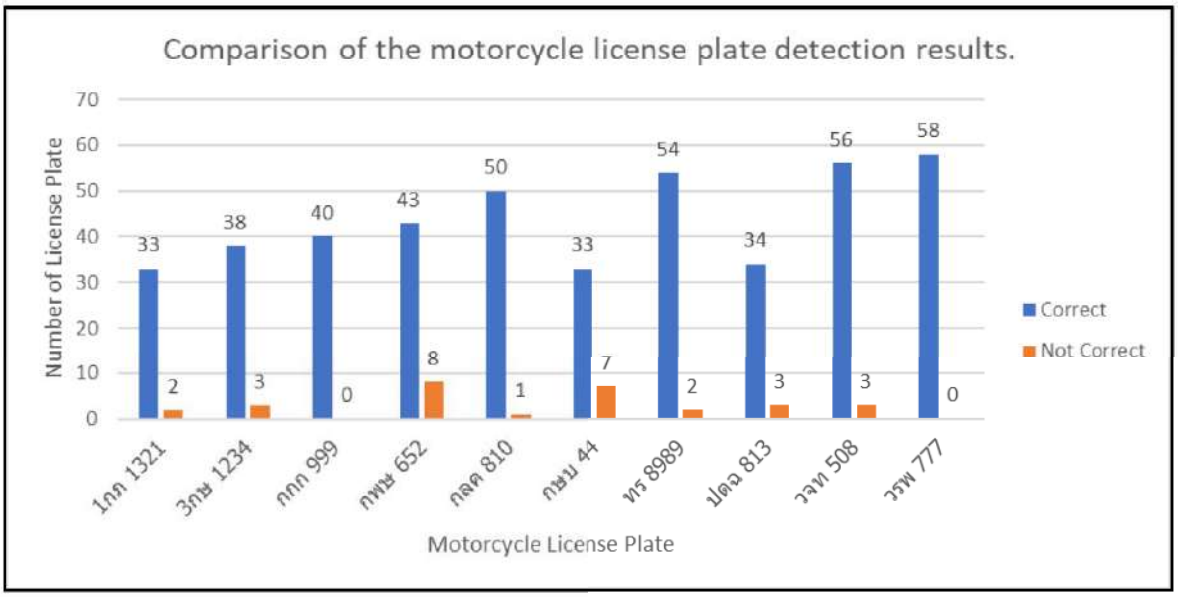


Figure 11. Comparison of the motorcycle license plate detection results.

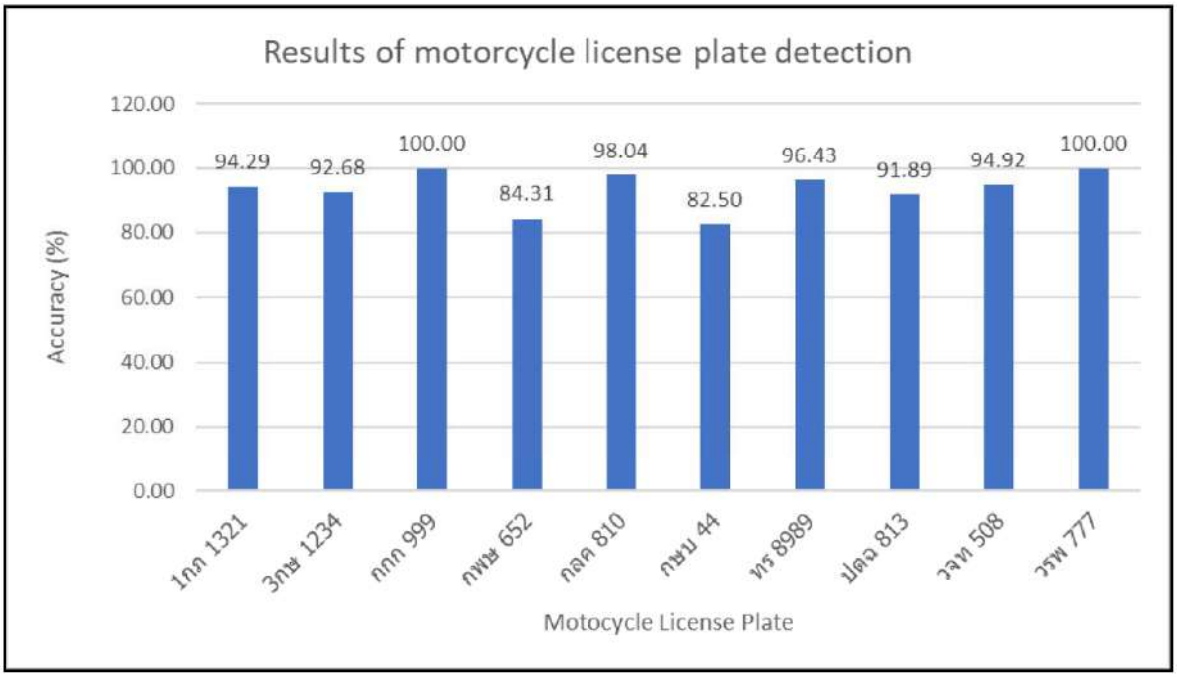


Figure 12. Results of motorcycle license plate detection.

1.8.1. Car license plate detection results

The total of 491 images of license plates were tested, 464 were correct and 27 were wrong, as shown in the figure 13 and table 3 shows a comparison of the car license plate detection result and show the results of car license plate detection in figure 14.

Table 3. Comparison of the car license plate detection results

Image No.	Car license Plate	Correct	Not Correct	Total	Accuracy (%)
1	1ขง 70	49	1	50	98.00
2	2กช 321	43	4	47	91.49
3	2ขม 46	47	6	53	88.68
4	5ฝ 4095	47	7	54	87.04
5	9กค 9535	53	3	56	94.64
6	กก 888	52	3	55	94.55
7	กท 9989	36	2	38	94.74
8	กย 9999	56	0	56	100.00
9	ฉว 5756	40	1	41	97.56
10	พร 9444	41	0	41	100.00
	Average accuracy (%)				94.67

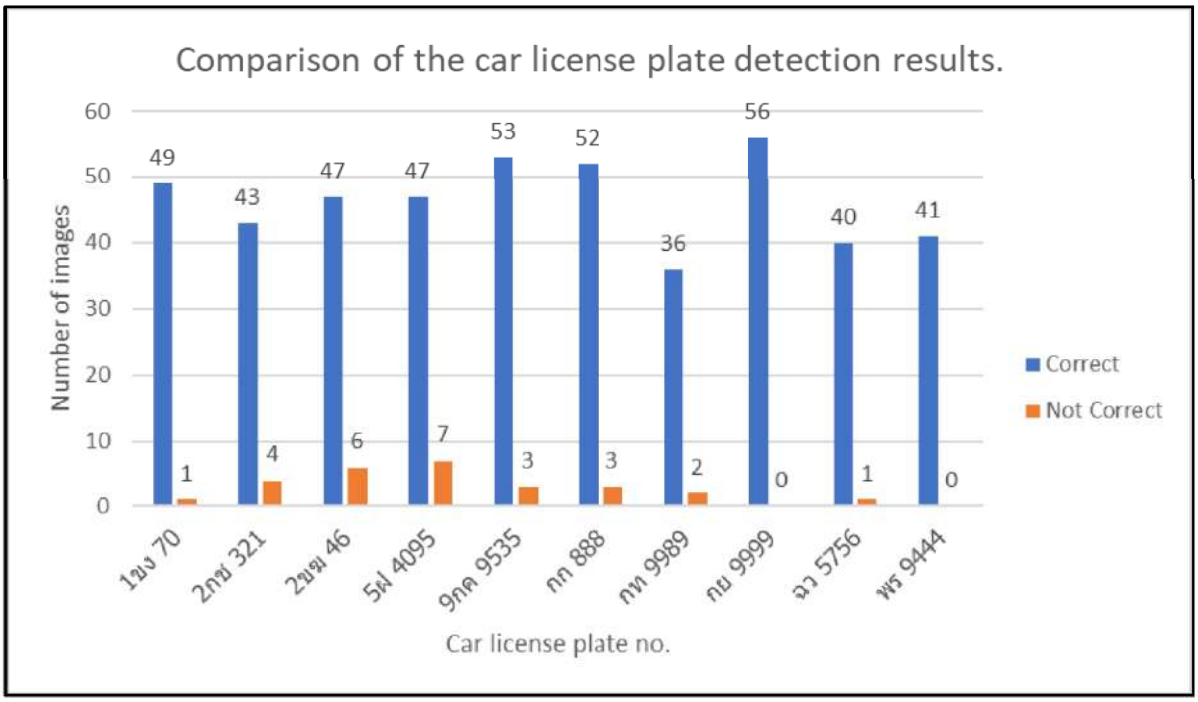


Figure 13. Comparison of the car license plate detection results.

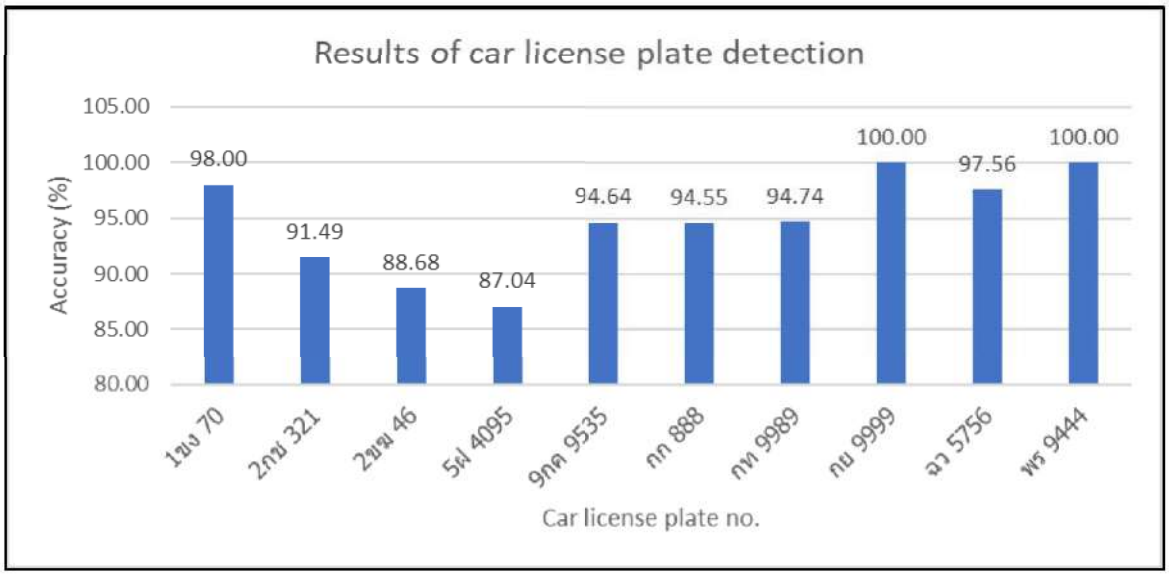


Figure 14. Results of car license plate detection.

1.8.2. Comparison of the motorcycle and car license plate detection results

The results of the photo-license plate detection test compared the results of the license plate detection by cars and motorcycles for each license plate. The average test result was that the car had an average accuracy of 94.67%, which was better than the motorcycle test result. *The* average accuracy is 93.51 as shown in table 4 and figure 15.

Table 4. Comparison of the motorcycle and car license plate detection results

Image No.	Motorcycle license plate detection accuracy (%)	Car license plate detection accuracy (%)
1	94.29	98.00
2	92.68	91.49
3	100.00	88.68
4	84.31	87.04
5	98.04	94.64
6	82.50	94.55
7	96.43	94.74
8	91.89	100.00
9	94.92	97.56
10	100.00	100.00
Detection accuracy (%)	93.51	94.67

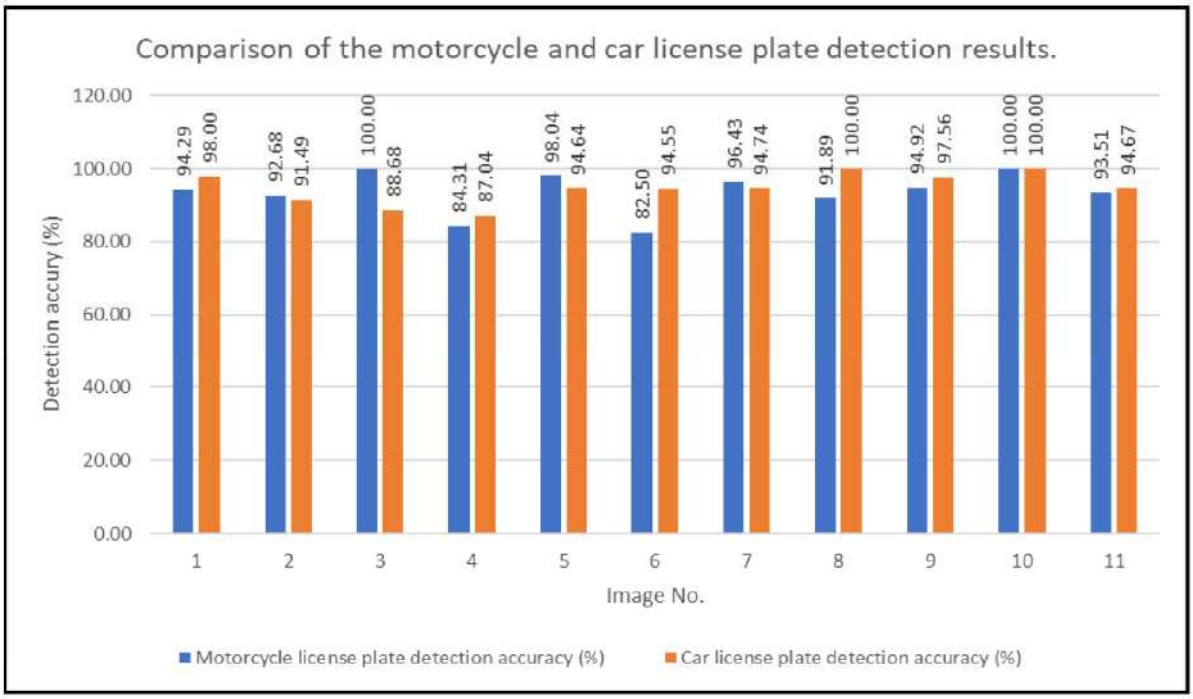


Figure 15. Comparison of the car license plate detection results.

Conclusion

This application can help to find parking spaces and car owners in case of parking in a prohibited area in Nakhon Ratchasima Rajabhat University. It can also search for license plates from photos with an average accuracy of 93% or more. The trial was conducted with various individuals, including executives, staff, students and external personnel. The satisfaction assessment result was 4.06, which was at a high level.

Discussion

From testing results, the pros and cons and further development topics can be concluded as shown in table 5.

Table 5. Conclusion of the pros, cons and further development topics of the parking system.

Pros	Cons	Further development topics
1. GPS detection works well for locating outdoor parking lots.	1. The GPS signal detection for locating indoor parking lots is likely to be inaccurate.	1. The system should use the Internet of Things (IOT) to find indoor parking lot coordinates.
2. The system is available in English.	2. The system cannot serve well for those do not understand English.	2. The system should add more languages, such as Thai, Chinese and Japanese, etc.
3. The system can be used for parking in all 30 buildings and can be parked inside the building, outside the building	3. The parking system cannot be used by other organizations, by which users configure and add parking lot data to the system by themselves.	3. The application should be developed the parking lot data configurations for users to add any parking lot data of any organization by themselves, such as schools, universities and private organizations.

Pros	Cons	Further development topics
and outside the university.		
4. There is an English audio system in each operation menu.	4. The system is not suitable for visually impaired people because it cannot operate by voice.	4. The system should add options for the visually impaired, such as voice commands and giving directions by voice.

Acknowledgments
This work was supported by a research grant from the Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University, Thailand.

References
[1] Roojai com 5 Must-have Apps for Car Lovers! Definitely good [Internet] 2017 [cited 2019 April 6] Available at://www roojai com/article/car-parts-and-car-accessories/5-app-for-car/
[2] Bureau of Road Maintenance Administration Department of Highways Thailand Highway Traffic [Internet] 2022 [cited 2022 March 4] Available at: http://bmm doh go th/website/index php/2-uncategorised/190-thailand-highway-traffic#
[3] Woraphon Pongphet 2012 Parking Building and Available Space Application using Android Operating System and Google Map (*Veridian E-Journal Science Group* Vol 5 no 3) p 492-507
[4] Montasinee Homwan, Komsan SiangWong and Suthat Chumphonkulwong 2006 A Parking Lot Searching System by WAP Service via Mobile Phone (*Bangkok University Academic Review Journal* Vol 5 no 2)
[5] Ittipol Homhoun, Chaiwat Kanfak and Pongchai Meebun 2015 Real-time parking decision support system in Naresuan University Master's thesis, Naresuan University
[6] Thai Car License Plates good [Internet] 2022 [cited 2022 March 4] Available at: https://driving information in th/license-plates html
[7] Vehicle registration plates of Thailand [Internet] 2022 [cited 2022 March 4] Available at: https://www wikiwand com/en/Vehicle_registration_plates_of_Thailand
[8] Including License Plates of Each Type [Internet] 2022 [cited 2022 March 4] Available at: https://www asiadirect co th/adbnewsdetail php?newId=62
[9] Jay Devore 2010 *Probability and Statistics for Engineering and the Sciences* Eighth Edition California Polytechnic State University, San Luis Obispo
[10] Michael Greenacre 2008 Correspondence Analysis and Related Methods [Internet] 2008 [cited 2008 April 26] Available at: http://www econ upf edu/~michael/stanford/maeb4 pdf
[11] MIT App Inventor [Internet] 2021 [cited 2021 Jan 15] Available at: https://appinventor mit edu/
[12] Airtable [Internet] 2021 [cited 2021 Jan 15] Available at: https://airtable com/
[13] Chakri Thamman and Manit Asanok (2018) Study results of information system components to promote research and academic services Faculty of Information Technology Rajabhat Maha Sarakham University (*Journal of Information Technology and Innovation Management Vol* 5 no 1) p 122-132
[14] Artificial Intelligence with MIT App Inventor 2022 [cited 2022 April 15] Available at: https://appinventor.mit.edu/explore/ai-with-mit-app-inventor

The reduction of unpleasant odor in pseudo ceramic by using natural filler

Tithinun Rattanaplome[1*], Napat Chantaramee[2,3]
[1]Rubber and polymer technology, Faculty of Engineering and Agro-industry, Maejo university 50290, Thailand
[2]Faculty of Science, Maejo University,Thailand
[3]Composite Materials and 3D Prototyping Research Unit, the Office of Agricultural Research and Extension Maejo University, T. Nongharn, A. Sunsai,, Chiangmai, Thailand

**Corresponding Author Email: tithinun@mju.ac.th*

Abstract. This research is aimed to reduce the level of unpleasant odor in pseudo ceramic by using natural filler, such as defatted rice bran and ground coffee. The natural fillers were used as odor reducing fillers and the content was varied at 0, 10, 20 and 30wt%. Firstly, the suitable content of odor reducing filler by using sensory method was determined. It was showed that the addition of natural adsorbent filler caused a decrease in the level of unpleasant odor. Especially, the pseudo ceramic filled with 20wt% of defatted rice bran and the pseudo filled with 30wt% of ground coffee gave the low level of unpleasant odor. Secondly, the odor chemical compound by using gas chromatography-mass spectrometry technique, water adsorption, flexural strength and abrasion resistance were evaluated. It was showed that only ground coffee, not defatted rice bran, caused a reduction in the average of relative peak area of N.N diethylformamide generated from clay-compound latex composite. The present of natural odor-adsorbent filler gave an increase in the percentage of water adsorption, but a decrease in flexural strength and abrasion resistance of pseudo ceramic. Pseudo ceramic filled with 30wt% of ground coffee gave the lower percentage of water adsorption (23.80%) and lower abrasive volume (242mm^3) with higher flexural strength than those of pseudo ceramic filled with 20wt% of defatted rice bran.

Keywords: Pseudo ceramic; Natural filler; Natural rubber latex; Ground coffee

1. Introduction

Green technology for sustainable development has been major concern by many researchers. It can involve in diversified areas such as bio-fuel, renewable energy, solid waste management and eco-material production. For example, the clay-based material as earthenware ceramic for construction has been consumed high energy during firing at temperature above 800°C. Thus, eco-materials for construction have been developed. Lee et.al. [1] investigated red clay composite reinforced with polymeric eco-binder, such as polyvinyl alcohol (PVA), polyvinyl pyrrolidone (PVP), carboxymethyl cellulose (CMC). Hinchiranan, N. [2] studied the production

of soil brick composite from rice husk and natural rubber latex. It was found that the ratio of clay and rice husk at 3:1 with 20% wt prevulcanized natural rubber latex based on the total quantity of liquid for brickmaking was appropriate to give the highest flexural resistance and compressive strength at 570 kg and 955 kg, respectively. Dove, C. [3] studied unfired earth bricks using seaweed biopolymers. Chantaramee, N. and Rattanaplome, T. [4] introduced pseudo-ceramic material from clay-compound latex composite that required low processing temperature at only 150°C resulting in lower energy consumption than conventional earthenware ceramic. Nevertheless, pseudo ceramic results in unpleasant odor probably due to thermal degradation of natural rubber latex. The method of reducing unbearable odor from natural rubber is in fact to change the offensive odor to be more pleasant by masking and/or reducing the odor intensity to a more acceptable level. Many researchers added odor adsorbent filler to reduce unbearable odor in natural rubber [4, 5]. Rattanaplome, T and Chantaramee, N. [6], by using gas chromatography-mass Spectrometry technique, reported that perlite and carbon black caused a reduction in the average of relative peak area of N.N diethylformamide which is odor chemical compound generated from clay-compound latex composite. Therefore, it is our interest to investigate the utilization of natural filler as odor reduction filler in clay-compound latex composite. Those are ground coffee and defatted rice bran, which are waste material from agriculture process.

2. Material and Methods

Material

A commercial grade clay (Compound Clay Co., Ltd, Thailand) containing kaolinite ($Al_2Si_2O_5(OH)_4$), microcline ($KAlSi_3O_8$) and quartz (SiO_2) was used in this study. High ammonia concentrated latex was purchased from Lucky Four Co., Ltd, Thailand. Other mixing ingredient such as sulfur (the vulcanizing agent), zinc diethyldithiocarbamate (the accelerator), zinc oxide (the activator) and Lovinox® (CPL, the antioxidant) were also supplied from Lucky Four Co., Ltd, Thailand. Potassium hydroxide was provided by Fluka Co., Ltd, Germany. Natural odor reducing fillers are ground coffee from Hillkoff and defatted rice bran from Thai edible oil company. They are sieved by 80mesh grid (177micron). The compound latex is shown in Table 1.

Table 1. The formulation of NR latex compound

Ingredient	Quantity (phr)*
60% High ammonia concentrated latex	100
10% Potassium hydroxide (KOH)	0.4
50% Sulfur (S)	1.5
50% Zinc diethyldithiocarbamate (ZDEC)	1.0
50% Zinc oxide (ZnO)	1.5
50% CPL	1.5

* parts per hundred of rubber

Methods

2.1 Evaluation of the odor reduction efficiency

The sensory method was evaluated by 20 semi-well trained persons according to guidelines for the selection and training of sensory panel members according to ASTM STP775 [7]. All ground samples about 50 gram were filled in each test bottle and covered with aluminum foil. Each semi-well trained person evaluated the odor intensity on a scale of 1–5 ranging from low to high odor intensity. Blind test was needed for each person to provide validity. Furthermore, each person was requested to smell the odor of coffee beans before each evaluation to eliminate the previous experience of the former odor determination and prevent odor confusion. Data was statistically shown in terms of mode. The gas chromatograph-mass

spectrometry technique (model GC-6890/MS-5793, Agilent Technology, USA) was used for the examination of odor compounds from samples (5g.). Ten microliters of 10 ppm methyl valerate were also filled into all NR sample test tubes as an internal standard. Before trapping air by using Polydimethylsiloxane/Divinylbenzene fiber above the samples, preheating at 600°C for 2 hours was done. The HP-5MS column containing polysiloxane (0.25 m.x25.0 m) was held at 50°C for 2 minutes, then heated at the rate of 5°C/min from 50°C to 230°C and held at 230°C for 4 minutes. The velocity of carrier gas (helium) was 1.0 ml/min. [4, 5, 6, 8]

2.2 Mechanical testing

The flexural strength was examined using a universal testing machine (LLOYD LRX 5K, UK) using a 3-point bending technique with a support distance of 80 mm and at a crosshead speed of 10 mm/min. The flexural strength (σ_f) was computed from the following equation:

$$\sigma_f = \frac{3FL}{2Bh^2} \qquad (1)$$

where F is the fracture load (kgf), L is the support distance (cm), B is the width of the sample (cm) and h is the thickness of sample (cm). Weibull statistical analysis was performed on each set with at least 20 samples resulting in Weibull's modulus (m) and the characteristic strength of 3% failure probability (σ_o) [4]

Abrasion resistance was determined using 5 samples according to ISO 10545-6, and fused Al_2O_3 with a particle size of 185 μm was used as the abrasive material. This method involves the groove length on the sample surface to the volume loss reported in cubic millimeters. The higher volume loss relates with the lower abrasion resistance. The test samples were square surface area (100mm. x 100 mm.) with the thickness about 10 mm. [4]

2.3 Physical testing

The water absorption of the samples was explained as the ratio of the weight of water adsorbed to their dry weight [6]. The samples were totally immersed in water at room temperature (about 28°C). After 15 min, the samples were taken out from the water and the excess moisture on surface was wiped off; the samples were then weighed and placed back in the water. Data were collected using a precision balance with a resolution of 0.01 gram every 15 min thereafter until the weight was constant. The water absorption percentage was evaluated from the following equation:

$$\%WA = \frac{(M_2 - M_1)}{M_1} \times 100 \qquad (2)$$

where M_1 is the initial dry mass and M_2 is the wet mass at a certain time.

2.4 Morphological Study

A morphological study was performed on the fracture surfaces of pseudo-ceramic material using the JSM-5800LV scanning electron microscope (SEM). The surfaces of all samples were sputter coated with gold to remove the electrostatic charge buildup during the observation.

3. Results and Discussion

3.1 The Odor Reduction Efficiency

The odor scale of pseudo ceramic was shown in Table 2. The high scale represents a high intensity of unpleasant odor, whereas the low scale represents a low intensity of odor. It was found that adding natural filler reduces the unpleasant odor. The smell of pseudo ceramic filled with ground coffee was the same as burning coffee, especially pseudo ceramic filled with 30wt% of ground coffee (gcf30). While, those of pseudo ceramic filled with defatted rice bran

was unique and pleasant smell, especially pseudo ceramic with 20wt% of defatted rice bran (drb20). The level of unpleasant odor of drb30 is higher than that of drb 20 will be discussed later. Thus, gcf30 and drb20 were selected for further investigation of the odor reduction efficiency by using GC-MS technique. The example of GC data was shown in figure 1. The major chemical compounds of odor smell pseudo ceramic were nonanal, N.N diethyl formamide, cyclohexasiloxane, dodecamethyl, and ketone. Especially, N.N diethyl formamide is a major odor compound found in STR5L and STR20 reported by Hoven et. al.[8] and Rattanaplome et. al. [5,6] Each peak area of chemical compound was compared with the peak area of internal standard (methyl valerate or methyl ester pentanoic acid). The average of two relative peak areas of chemical compound of pseudo ceramic were reported in table 3. In addition, the average of two peak areas of ground coffee and defatted rice bran oil were shown in table 4.

Table 2. The odor scale of pseudo ceramic.

Natural odor-reduction filler content (wt%)	The level of unpleasant odor	
	Ground coffee (gcf)	Defatted rice bran (drb)
0	4	4
10	4	1
20	3	0
30	1	2

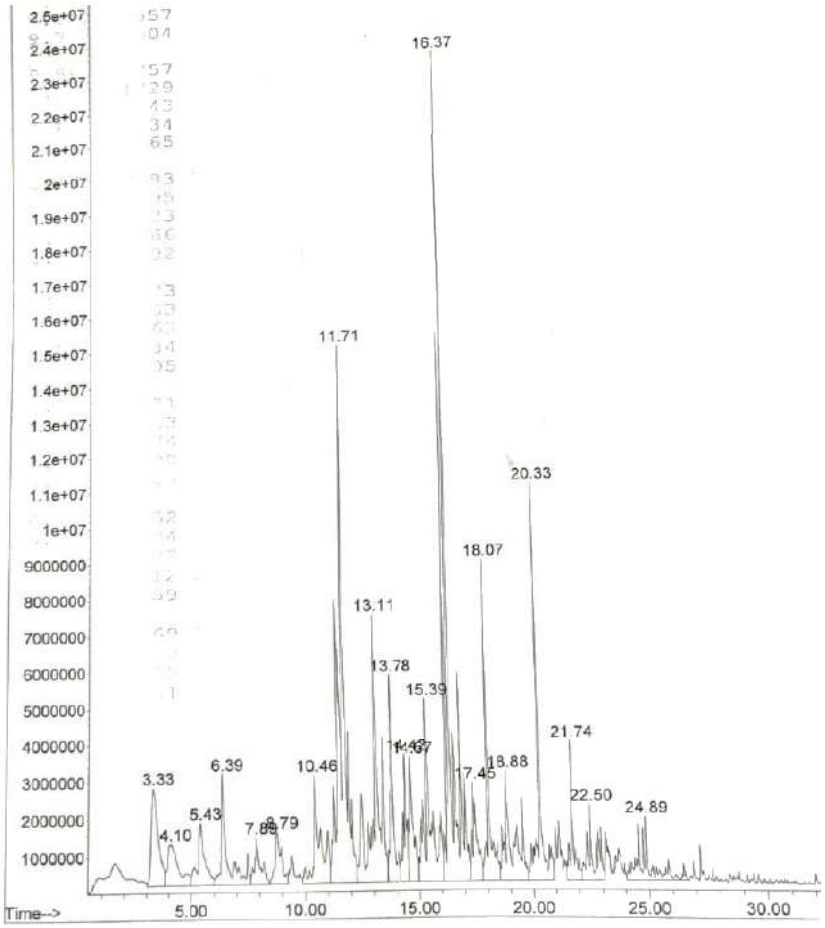


Figure 1. GC data of pseudo ceramic filled with 30wt% ground coffee. X axis represents retention time, while Y axis represents abundance of each peak.

Table 3. The average peak area of pseudo-ceramic samples

pseudo ceramic .		
Odor chemical compound	Peak time	Average peak area
Methyl valerate	2.03	100.0
nonanal	8.57	22.3
N.N diethyl formamide	9.60	140.4
cyclohexasiloxane, dodecamethyl	10.07	10.8
Ketone	16.94	8.5
pseudo ceramic filled with 30wt% ground coffee		
Methyl valerate	3.25	100.0
Nonanal	10.45	69.8
Benzaldehyde	13.12	203.5
Ketone	18.90	90.4
Hexanoic acid	20.34	98.6
Heptanoic acid	22.50	45.5
pseudo ceramic filled with 20wt% defatted of rice bran		
Methyl valerate	3.37	100.0
Nonanal	10.46	99.0
N.N diethyl formamide	11.43	133.9
Ketone	15.47	25.5

It was found that adding natural filler caused an increase in the average peak area of nonanal and ketone of pseudo ceramic samples. Furthermore, pseudo ceramic filled with 30wt% ground coffee showed benzaldehyde, hexanoic acid and heptanoic acid generated from natural rubber. It seemed that ground coffee and defatted rice bran was unable to adsorb typical odor chemical compound of natural rubber. However, incorporation of natural filler caused a decrease in peak area of N.N diethyl formamide. Especially, adding ground coffee gave the absent of average peak areas of N.N diethyl formamide. This could be due to hydroxyl group (–OH) grou on natural filler surface adsorbing N.N diethyl formamide through hydrogen bond. [5, 6, 9]. Further investigation on odor chemical compound of pseudo ceramic with 30wt% defatted rice bran oil, the average peak areas of N.N diethyl formamide was about 1343.3. This could be resiponsible fo the high level of unpleasant odor from sensory method.

Table 4. The average peak area of natural fillers

Ordor Chemical compound	Peak time	Average relative peak area
Ground coffee		
Methyl valerate	3.30	100.0
D-Limonene	5.60	57.6
2,2,4,4,6,8,8-heptamethyl nonane	8.94	64.5
Nonanal	10.45	129.1
Ketone	18.91	121.6
Tetraethyl thiourea	22.04	72.1
Defatted rice bran oil		
Methyl valerate	2.43	100.0
2-Heptanal, (z)	8.13	1267.6
Nonanal	8.87	1380.3
1-octen-3-ol	101.53	1527.4
Benzaldehyde	11.81	1381.

3.2 Mechanical and Physical properties

Fracture surface of all samples after flexural test were shown in figure 2. It could be observed that crack surface of all samples are relatively smooth. The Weibull plots of all samples are shown in figure 3 and 4. Table 4 summarizes the flexural strength, the Weibull moduli (m), the abrasion resistance and the percentage of water absorption evaluated on samples. Addition of natural filler in pseudo-ceramic samples causes a decrease in strength and abrasion resistance but an increase in water adsorption. The low Weibull moduli means that the significant difference between the highest flexural strength and the lowest flexural strength. Those samples with natural fillers shows low Weibull moduli. This could imply that natural filler might interfere the homogeniety of pseudo ceramic samples.The high water adsorption could be due to hydroxyl group on natural filler surface. Similar observation was also reported by Rattamplome et al. [6].

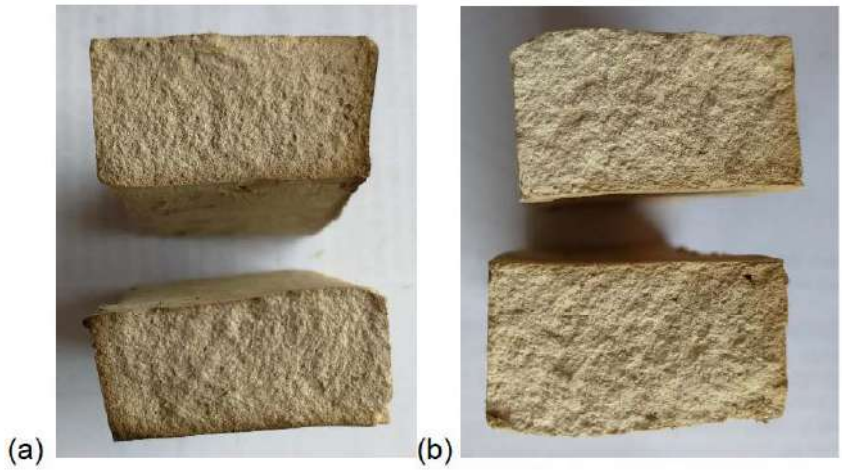
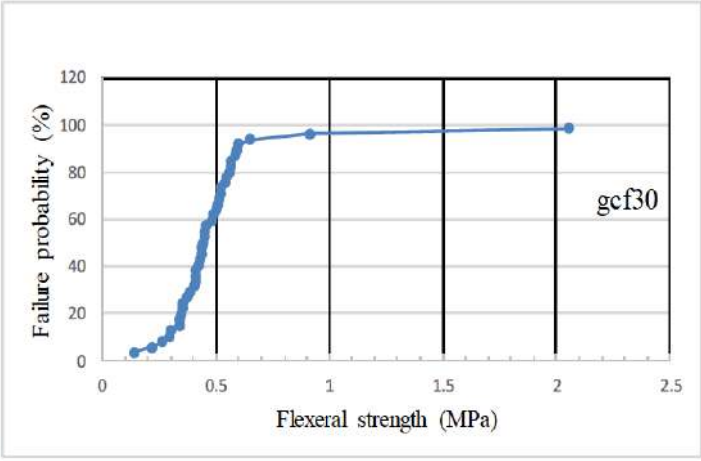


Figure 2. Pseudo-ceramic samples after flexural test of (a) pseudo-ceramic filled with 30wt% of ground coffee (gcf30) (b) pseudo-ceramic filled with 20wt% of defatted rice bran (drb20)

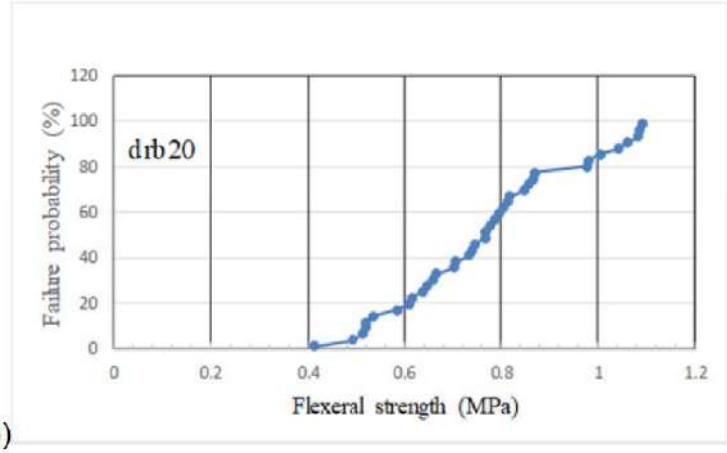
Table 5. Mechanical and physical properties of pseudo-ceramic samples.

Properties	Types of pseudo-ceramic samples		
	control*	gcf30	drb20
Flexural strength (MPa)	3.9149 (m=16.40)	0.1652 (m=0.16)	0.0675 (m=0.14)
Abrasion resistance in term of volume loss (mm^3)	196.4±48	389.2±182	242.0±100
%Water adsorption	20.83	23.80	24.93

*control means pseudo-ceramic samples without filler



(a)



(b)

Figure 3. Weibull plots of failure probability vs. flexural strength for pseudo ceramic. (a) pseudo ceramic filled with 30wt% of ground coffee (gcf30) (b) pseudo ceramic filled with 20wt% of defatted rice bran (drb20)

3.3 SEM micrograph

Morphology of various types of fractured pseudo-ceramic samples a e revealed in figure 4. From the micrographs with 500-magnification, the relatively homogeneous surface of samples was observed. This could be because natural fillers were ball milled and homogenously mixed with clay compound.

4. Conclusion

In conclusion, only 30wt% of ground coffee, not 20wt% of defatted rice bran, caused a reduction in the average of relative peak area of N.N diethylformamide generated from psdudo ceramic. Nevertheless, natural fillers seem to be ineffective odor-reducing filler and caused an increase in the percentage of water adsorption, but a decrease in flexural strength and abrasion resistance of pseudo-ceramic.

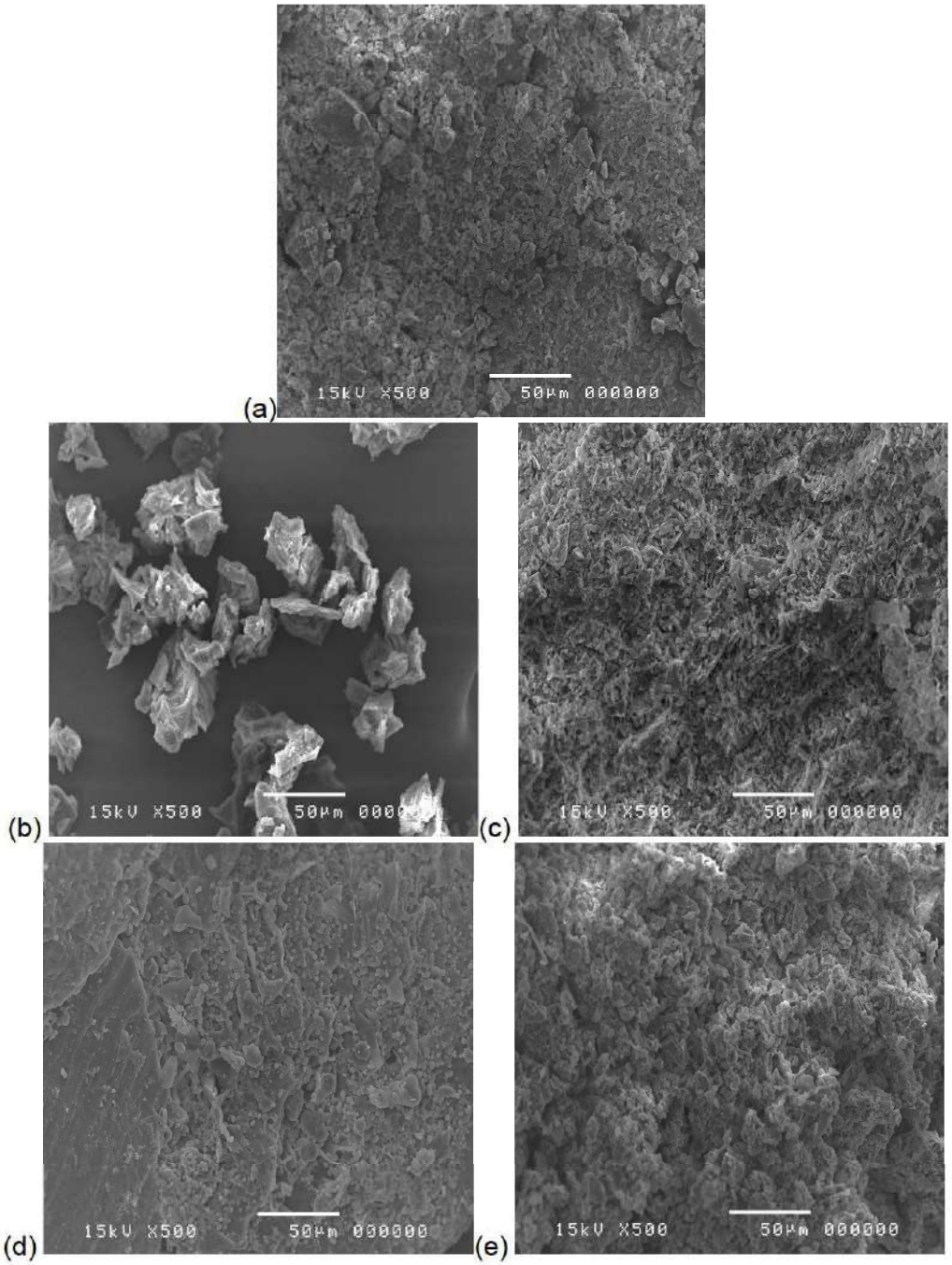


Figure 4. SEM micrograph of (a) pseudo ceramic without natural filler (b) ground coffee (c) pseudo ceramic filled with 30wt% of ground coffee (gcf30) (d) defatted rice bran (e) with 20wt% of defatted rice bran (drb20)

Acknowledgments
The authors would like to express special thanks of gratitude to financial support from the office of Agricultural research and extension, Maejo university (MJU-1-61-033). The authors are also appreciated support from the Faculty of Engineering and Agro-Industry, Maejo University for providing the research facilities. Finally, the authors would like to thank students and coworkers from the Rubber and Polymer Technology unit for their contributions

References

[1] Lee, K. C., Her, J. H. and Kwon, S. K. 2008 Red clay composite reinforced with polymeric binder, *Construction and Building Materials,* 22 p 2292-2298
[2] Hinchiranan, N. 2009 Production of soil brick composite from rice husk and natural rubber latex (Research report). Thailand Research fund.
[3] Dove, C. 2014 The development of unfired earth bricks using seaweed biopolymers, *Eco-Architecture* p142
[4] Rattanaplome, T., Chantaramee, N., and Pornprasit, P. 2015 The potential of perlite as an odour-adsorbing fillers in natural vulcanizates, *Macromolecular Symposium*, 354(1), p197-206
[5] Rattanaplome, T., Chantaramee, N., and Pornprasit, P. 2015 Utilization of perlite as an odor-reducing filler in natural rubber vulcanizates, in Proceeding of the second Asia Pacific Rubber Conference (APRC), Surathani, Thailand, p195-205
[6] Rattanaplome, T. and Chantaramee, N. 2020 The reduction of unpleasant odor in clay-compound latex composite by using odor adsorbent filler Proceeding of the 12th international conference on science, technology and innovation for sustainable well-being, p319-324
[7] Meilgaard, M., Civille, G. V. and Carr, B.T. 1992 Sensory Evaluation Technique, 3rd Ed. Chapman& Hall, Boca Raton, p86-94, p139-140
[8] Hoven, V.P., Rattanakarun, K., and Tanaka, Y. (2004) Reduction of offensive odor from natural rubber by odor-reducing substances, *Journal of Applied Polymer Science*. 92, p2253-2260
[9] Sakdapanich, J. T. and Insom, K. 2006 High-resolution gas chromatography-mass spectrometry: Characterization and mechanism to generate the obnoxious odor in natural rubber *KGK*. 7/8-06, p382-387.
[10] Chantaramee, N. and Rattanaplome, T. 2015 Flexural strength and abrasion resistance of clay-based materials reinforced with vulcanized natural rubber latex, Proceeding in Bangkok international conference on engineering and applied science

Improving the Tensile Strength of Rice Straw Paper by Different Paper Additives

Leulee Nortoualee[1,*], Siraporn Cheunbarn[2], Tapana Cheunbarn[2], Tawat Soitong[3] and Supattra Wongsaenmai[3]
[1]Program in Food Science and Technology, Faculty of Agriculture and Forest Resource, Souphanouvong University, Luang prabang, Lao PDR.
[2]Program in Environmental Technology Program, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand
[3]Program in Materials Science, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand
**Corresponding author's email: leuleemju@gmail.com*

Abstract. This research investigated the improvement of tensile strength of rice straw paper by using two types of additives, which were rice flour and tapioca starch, in order to make the rice straw paper of a better quality. The results of the study found that rice straw paper modified with tapioca starch and rice straw paper modified with rice flour saw as increase in tensile strength compare with unmodified rice straw paper at a 95% confidence level. The highest tensile strength was found in rice flour modified paper at 3.04 ±0.37 N/mm^2, followed by tapioca starch modified paper at 2.12 ± 0.39 N/mm^2 and finally, rice straw paper at 3.04 ± 0.37 N/mm^2. Compression was another way to increase the tensile strength of rice straw paper using a compressive strength of 2,500 pounds per inch at a temperature of 150 ˚C for 5 minutes, the results showed that compression every type of paper had higher tensile strength. However, only rice straw paper and tapioca starch modified paper to have tensile strength that were statistically different at the 95% confidence level. The rice flour modified paper was found that no statistically significant difference in the tensile strength between uncompressed and compressed paper. In conclusion, rice straw paper modified with rice flour is the strongest paper and could by apply for use as packaging in the future.

Keyword: Pulp paper, Rice straw, Rice Flour, Tapioca starch, Tensile strength

1. Introduction

From the summary of The Seventh National Economy and Social Development Plan 5 years (2011-2015), in Lao PRD it was found that in 2015, the country had a rice planting area of approximately 948,535 hectares, yielding 4.1 million tons of rice, divided into annual rice cultivation 755,950 hectares, yield 3.36 million tons, secondary rice fields are 101,252 hectares, the yield is 520,000 tons, and upland rice is 91,060 hectares, the yield is 224,360 tons. Therefore, it is found that there is an average of 5.7 million tons of rice straw from the agricultural sector per year, compared with the data of Vietnam which has about 6 tons of rice straw waste per hectare [1]. The huge quantities of straw were leftover each year and this was a problem for farmer after the harvest season.

Although rice straw can be used for many purposes, the current management of excess straw is poor. only a small fraction is as fuels, fertilizers, and animal feed. In general, after the harvest most of the rice straw is burned in situ to prepare the following crop. However, burning destroys the soil structure, soil nutrient, microorganism, beneficial organisms and causes soil degradation [2]. whereas burning in open area causer extensive atmospheric pollution including greenhouse and gas emissions and smoke [3].

Utilization of waste straw from rice production in paper production is one way to reduce burning and provide income for farmers in Lao PRD. the study of Leulee et al. (2020) seems to show that only straw can be used to make paper. The paper sheet from 50g rice of straw wet pulp had the highest tensile strength at 0.45±0.04 N/mm^2. However, the paper from this process was not strong enough to use for a durable product. The pulp from rice straw only without other ingredients had low tensile strength. Therefore, this study aims to improve the properties of rice straw paper by using different types of starch to improve paper quality for future packaging development.

2. Materials and Methods

Equipment: stainless steel pulp boiler, gas stove, thermometer stand, scissors, spatula, thin rice cloth, gloves, pulp storage tank, pulp spoon sieve for washing pulp Sieve up the pulp, absorbent cloth, rolled up paper sheets.

Raw materials: rice straw, rice flour, tapioca starch, clean water.

Chemical: sodium hydroxide (NaOH)

Tools: 4 position precision balance, Weighing scale, pH Meter, paper compactor paper thickness gauge, Fume Hood, Heater

Glassware: beaker (1000 ml and 200ml), cylinder, vacuum filter set, spatula, glass stirring rod

Method

The pulp used in the experiments came from rice straw, according to the process from Leulee et al., 2020. The experiment was divided into 2 treatments as follows: the control (50g of rice straw pulp), treatment 1 (50g of rice straw pulp and 6% tapioca starch) and treatment 2 (50g of rice straw pulp and 6% rice flour starch). The ingredients were mixed together with a blender at 1800 rpm for 20 minutes, put on a sheet of paper with a frame size 25x25 cm and dried and pressed with a compressor with a compressive strength of 2,500 pounds per inch, at 150 °C for 5 minutes. Then, the modified papers with the two types of starch had the properties tested: grammage (Basis Weight) used standard reference TAPPI T410, thickness used standard TAPPI-T411-om-89. and tensile strength use standard of TAPPI-T494-om-01.

Data analysis

Data were statistically analyzed using SPSS program, compared with mean difference by Duncan's New multiple rang test: (DMRT).

3. Results and discussion

3.1. Properties of rice straw paper modified with tapioca starch and rice flour

Improvement of paper properties occurred by using 50g of paper pulp with 6% boiled tapioca starch and 50g of paper pulp with 6% boiled rice flour of water, wet pulp weight per volume. The pulp was blended with 1 liter of water using a blender at 1800 rpm for 20 minutes. Then put on a sheet of paper with a frame size of 25x25cm, the experimental results are shown in Table 1.

Table 1. Properties of modified rice straw paper

Paper sheet	Properties of a sheet of paper		
	Thickness (mm)	Grammage (g/m²)	Tensile strength (N/mm²)
Rice straw paper (control)	0.63 ±0.03	268.32 ± 1.89	0.32 ± 0.02[c]
Tapioca Starch Modified Paper	0.53 ± 0.05	228.96 ± 0.22	2.12 ± 0.39[b]
Rice Flour Modified Paper	0.57 ± 0.06	242.72 ± 0.25	3.04 ± 0.37[a]
SEM	0.021	0.482	0.41
F-Test	ns	ns	**

ns means no statistically significant difference at the 95% confidence level (P-value ≥ 0.05).
a, b, c means a statistically significant difference at the 95% confidence level (P-value ≤ 0.05).

From the results, it was found that the thickness and standard weight (grammage) of rice straw paper sheets were not significantly different among the 3 treatments, although, the rice straw paper (control) had the highest thickness and standard weight at 0.63 ±0.03 mm and 268.32 ± 1.89 g/m² respectively. The tensile strength of rice straw paper was found to have a statistically significant difference at a 95% confidence level. The highest tensile strength was in rice flour modified paper (treatment 2) at 3.04 ±0.37 N/mm^2, followed by tapioca starch modified paper (treatment 1) at 2.12 ± 0.39 N/mm^2, and finally, rice straw paper (control) at 0.32 ± 0.02 N/mm^2. The results showed that the modified paper with tapioca starch and rice flour made tensile strength increase.

3.2. The property of modified rice straw paper in terms of tensile strength after compression

The quality of the paper was improved after compression at a compressive strength of 2,500 pounds per inch at a temperature of 150 ˚C for 5 min. The comparison of the properties of modified rice straw paper with and without compression in terms of tensile strength among rice straw paper (control), tapioca starch modified paper (treatment 1) and rice flour modified paper (treatment 2) are show in Table 2.

Table 2. The comparison of the property of modified rice straw paper in terms of tensile strength between uncompressed and compressed

Paper sheet	Tensile strength of paper (N/mm²)		SEM	F-Test
	Not compressed	Compressed		
Rice straw paper (control)	0.32 ± 0.02[b]	0.45 ± 0.04[a]	0.03	**
Tapioca Starch Modified Paper	2.12 ± 0.39[b]	4.35 ± 0.18[a]	0.51	**
Rice Flour Modified Paper	3.04 ± 0.37[a]	4.61 ± 0.93[a]	0.43	ns

ns means no statistically significant difference at the 95% confidence level (P-value ≥ 0.05).
a, b means there was a statistically significant difference at the 95% confidence level (P-value ≤ 0.05).

From Table 2, the results show that after compression every type of paper had higher tensile strength, however, only the rice straw paper (control) and the tapioca starch modified paper (treatment 1) had statistically significant differences at the 95% confidence level. The rice flour modified paper (treatment 2) was found to have no statistically significant difference in the tensile strength between uncompressed and compressed.

Discussion

The production of rice straw paper using only 50 g of rice straw pulp had a tensile strength of 0.32 ± 0.02 N/mm^2 which is less than Saa paper which is a commonly used paper made from pulp. The Saa paper had a tensile strength of 0.90 ± 0.36 N/mm^2 [5]. However, when rice straw paper was modified with tapioca starch and rice flour, the paper had a higher tensile strength. The rice flour modified paper had the highest tensile strength at 3.04 ± 0.37 N/mm^2, which agrees with Noppamas who stated that in the wet process when starch is added, the starch serves to help micro fibers and particles to bind together. It also improves the bonding force between fibers and others, including fibers and particles. The resulting paper is stronger [6]. Compression was another way to increase tensile strength of rice straw paper. Due to the compressive strength, the paper fibers overlap more tightly, and the starch contained in the paper binds to the pulp fibers better. Thus, the tensile strength of rice straw paper is increased.

4. Conclusion

From the results of the experiment, it was found that tapioca starch modified rice straw paper and rice flour modified rice straw paper had a statistically significant increase in tensile strength at 95% confidence level compared to non-modified rice straw paper. The rice flour modified paper had the highest tensile strength at 3.04 ± 0.37 N/mm^2. When the paper was compressed with a compactor, it was found that the paper had a tensile strength that was one to two times higher than that of uncompressed paper. However, the tensile strength of rice flour modified straw paper both compressed and uncompressed was not significantly different.

Therefore, to improve the tensile strength of paper a mix of pulp and flour in the paper production process can make the paper stronger. Starch acts to help the fibers bind and also increases the bonding force between the fibers themselves, including the fibers and the boiled starch, thus resulting in a stronger paper, and from the results of this study in this research rice straw paper modified with rice flour is the strongest paper and could be used for packaging in the future.

Acknowledgments

This research was funded under the Second Stage Higher Education Strengthening Program (SSHEP) funded by ADB. Thank you to Faculty of Science Maejo University

References

[1] Ministry of Planning and Investment. 2016. The 8th Five-Year National Economic and Social Development Plan
[2] Department of Land Development. 2005. Refrain from burning tongs create sustainable soil Restore the environment. Department of Lands Ministry of Agriculture and Cooperatives
[3] Kanokkanjana, K and Garivait, S, 2013. Alternative Rice Straw Management Practices to Reduce Field Open Burning in Thailand. International Journal of Environmental Science and Development, 4.
[4] Leulee Nortoualee, Siraporn Cheunbarn, Tapana Cheunbarn, Tawat Soitong and Supattra Wongsaenmai. 2020. Effect of Lao's Local Rice Cultivars on Properties of Pulp

and Paper from Rice Straw. The 17th National Kasetsart University Kamphaeng Saen Conference

[5] Leulee Nortoualee. 2020. Increasing the value of rice straw by producing packaging paper. Master's Thesis Maejo University

[6] Noppamas Yiamsawat. 2010. Effect of external application of starch solution on surface paper properties.

The Parameters Estimation for the Discrete Weibull Regression Model With Type-I Right Censored Data

Dusit Chaiprasithikul[1,*], Monthira Duangsaphon[1]
[1]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathumthani, Thailand.

**Corresponding Author Email: dusit1994@hotmail.com*

Abstract. This research aims to study the parameters estimation for the discrete Weibull regression model with type-I right censored data by linking both the shape parameters which are the log-log and the log link functions. Moreover, this study compared the performance of the Bayesian estimation under the three different prior distributions; uniform noninformative prior, Laplace prior, and normal prior using the random walk Metropolis algorithm and the maximum likelihood estimation. A simulation study was conducted to compare the performance of four different estimation methods by using the mean square error under three explanatory variables and over-dispersion response data. The results from the simulation study showed that the Bayesian estimation with Laplace prior is more appropriate for the discrete Weibull regression model with type-I right censored data than other methods.

Keywords: Bayesian estimation; discrete Weibull regression; maximum likelihood estimation; random walk Metropolis algorithm; type-I right censored.

1. Introduction

Regression analysis for count data is used in realistic contexts when other variables have an effect on the count response variable. Count data refer to the number of times an event or an item occurs over a fixed period of time, which can take only the non-negative integer values. Examples include the number of times cardiac arrest happens over a fixed period of time, the number of postoperative complications over a fixed period of time, the number of claims in an insurance company over a fixed period of time, and the number of recurrent circuit breaker failures over a fixed period of time. Individual count data is called a count variable, which is treated as a random variable: the Poisson, negative binomial, and discrete Weibull distributions are widely used to represent its distribution. Modeling count data may present three types of dispersion: equi-dispersion, under-dispersion, and over-dispersion. The Poisson regression model is commonly used to evaluate the relationship between the count response variable and explanatory variables, e.g., [4,10]. However, its use is limited of the equality of the mean and variance assumption with real data. The negative binomial regression model is used to account for over-dispersion, e.g., [1,9]. Moreover, the discrete Weibull regression model is utilized to manage for under-dispersion and over-dispersion. The discrete Weibull regression is an interesting subject for study and development, e.g., [7,15]. The cumulative distribution function and the probability mass function of a discrete Weibull random variable Y are given by

$$F_Y(y;q,\beta)=\begin{cases}1-q^{(y+1)^\beta} & ;y=0,1,\ldots\\ 0 & ;\text{otherwise}\end{cases} \tag{1}$$

and

$$p_Y(y;q,\beta)=\begin{cases}q^{y^\beta}-q^{(y+1)^\beta} & ;y=0,1,\ldots\\ 0 & ;\text{otherwise}\end{cases} \tag{2}$$

respectively, where $0<q<1$ and $\beta>0$ are the shape parameters.

Focusing on over-dispersion data wherein the variance is larger than the mean may limit the effectiveness of a standard model. When the value of an observation or measurement is only partly known we refer to this observation as being censored. In some cases, the response variable takes large values or outliers affecting its mean and variance, causing over-dispersion that potentially having a negative effect on the performance of a regression model. Thus, censoring the large values of this response can control this over-dispersion. Therefore, appropriate censoring is applied to solve problem with over-dispersion data. Censored count data can appear in many applications where recording the count response variable is available for a limited range while the covariate values are always observed. Moreover, the censoring data is natural interest in survival analysis, reliability analysis, and many studies due to cost or time considerations. There are three types of right censoring: type-I right censoring that fixes a predetermined censoring value; type-II right censoring that fixes a predetermined number of uncensored data; and random censoring, extended from the type-I right censoring, in which the predetermined censoring value is random. The challenge faced by practitioners is the selection of censored data, e.g., [14,15]. In this research, we consider the type-I right censoring models for count data which n independent observations y_i^* censored from the right at a fixed censoring value C. The observed response variables y_i can be determined as

$$y_i=\min\left(y_i^*,C\right)=\begin{cases}y_i^* & ;y_i^*<C\\ C & ;y_i^*\geq C\end{cases} \tag{3}$$

The censor indicator for type-I right censoring δ_i can be specified as

$$\delta_i=I\left(y_i^*\geq C\right)=\begin{cases}0 & ;y_i^*<C\\ 1 & ;y_i^*\geq C\end{cases} \tag{4}$$

The likelihood function for type-I right censored data is given by

$$L_C=\prod_{i=1}^{n}\left[p_{Y^*}(y_i)\right]^{1-\delta_i}\left[1-F_{Y^*}(C-1)\right]^{\delta_i} \tag{5}$$

where $p(\cdot)$ and $F(\cdot)$ are, respectively, the probability mass function and the cumulative distribution function for y_i^*, $i=1,2,\ldots,n$.

Methods to estimate the regression model parameters precisely and efficiently are very important. The maximum likelihood estimation is valid for an asymptotically large sample size of data. Additionally, the maximum likelihood estimation is used for only empirical knowledge from the likelihood function. Alternatively, the Bayesian estimation is an interesting method because it uses information from both prior knowledges about the parameters from the prior probability distribution and empirical knowledge from the likelihood function. Hence, the performance of the Bayesian estimation depends upon the prior distribution that is defined. According to determining the prior distribution, it is very important in the Bayesian estimation, if the researchers have no prior knowledge of the parameters; then they can use the

noninformative prior distribution. Contrastingly, the researchers use the informative prior distribution when knowing about prior knowledge of the parameters. Furthermore, the Bayesian estimation is offered for small sample problems [13]. However, the disadvantage of the Bayesian estimation is that it takes a long time to compute.

Kalktawi [15] performed the maximum likelihood estimation of parameters for the discrete Weibull regression model with type-I right censoring data via the log-log link function under a shape parameter. Moreover, there are many paper works considering the Bayesian inference for estimation in discrete regression model. Unfortunately, there are no conjugate priors in the context of discrete Weibull regression model. Haselimashhadi et al. [7] proposed the Bayesian estimation for the discrete Weibull regression model under a uniform noninformative prior, Jeffreys prior, and Laplace prior by linking both the shape parameters which are the log-log and the logit link functions and the log link function. It is often more natural to express prior information directly in term of the parameters, the regression coefficients that can be a real number which correspond to the possible values of a normal distribution. There are papers selecting the prior distribution of the regression coefficients as a normal distribution [3,5,6].

The main objective of this research is to perform the parameters estimation for the discrete Weibull regression model with type-I right censored data by linking both the shape parameters which are the log-log and the log link functions. Moreover, this research compared the performance of the Bayesian estimation under the three different prior distributions; uniform noninformative prior, Laplace prior, and normal prior and the maximum likelihood estimation. This research constructed the Bayes estimators of the parameters under squared error loss function which is the expected value of the joint posterior density function. The main difficulty faced when dealing with the Bayesian estimation comes from the integral of the posterior probability distribution without a closed form. Therefore, in this case, we choose one of the Markov chain Monte Carlo (MCMC) methods which is the random walk Metropolis algorithm in order to estimate the parameters.

The remainder of this research is organized as follows. In Section 2, we perform the parameters estimation for the discrete Weibull regression model with type-I right censored data. In Section 3, we investigate the performance of the estimations through a simulation study. Finally, we conclude our findings in Section 4.

2. Materials and Methods

In this section, we perform the parameters estimation for the discrete Weibull regression model with type-I right censored data by linking both the shape parameters.

2.1. Maximum likelihood estimation

According to Kalktawi [15] performed the maximum likelihood estimation for the discrete Weibull regression model with type-I right censored data by linking only the parameter q. In this section, we perform the maximum likelihood estimation for the discrete Weibull regression model with type-I right censored data by linking both parameters q and β. Let $Y_1^*, Y_2^*, \ldots, Y_n^*$ be the count response variables which take only the non-negative integer values and let $x_{i1}, x_{i2}, \ldots, x_{ik}$, $i = 1, 2, \ldots, n$, be k explanatory variables. Assume that the conditional distribution of Y_i^* given $\mathbf{x}_i$ follows the discrete Weibull distribution with the parameters q_i and β_i, where the parameters q_i and β_i are related to k explanatory variables via the log-log and the log link functions:

$$\log\left(-\log\left(q_i\right)\right) = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_k x_{ik} = \mathbf{x}_i \boldsymbol{\alpha} \tag{6}$$

and

$$\log\left(\beta_i\right) = \zeta_0 + \zeta_1 x_{i1} + \ldots + \zeta_k x_{ik} = \mathbf{x}_i \boldsymbol{\zeta} \tag{7}$$

respectively, where $q_i = e^{-e^{(\mathbf{x}_i\boldsymbol{\alpha})}}$, $\beta_i = e^{\mathbf{x}_i\boldsymbol{\zeta}}$, $\mathbf{x}_i = \begin{pmatrix}1 & x_{i1} & \cdots & x_{ik}\end{pmatrix}$, $\boldsymbol{\alpha} = \begin{pmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_k\end{pmatrix}'$, and $\boldsymbol{\zeta} = \begin{pmatrix}\zeta_0 & \zeta_1 & \cdots & \zeta_k\end{pmatrix}'$; $i = 1,2,\ldots,n$. Thus, the conditional probability mass function of Y_i^* given $\mathbf{x}_i$ can be written as

$$p_{Y_{DW}^*|\mathbf{X}}\left(y_i^*|\mathbf{x}_i\right) = \begin{cases} \left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{y_i^* e^{\mathbf{x}_i\boldsymbol{\zeta}}} - \left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{\left(y_i^*+1\right)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}} & ; y_i^* = 0,1,2,\ldots \\ 0 & ;\text{otherwise} \end{cases} \tag{8}$$

Given n independent observations y_i^*, $i = 1,2,\ldots,n$ from Equation 8 censored from the right at a fixed censoring value C, the likelihood function and the log-likelihood function of the discrete Weibull regression model with type-I right censored data are given by

$$\begin{aligned} L_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right) &= \prod_{i=1}^{n}\left[p_{Y_{DW}^*|\mathbf{X}}\left(y_i|\mathbf{x}_i\right)\right]^{1-\delta_i}\left[1-F_{Y_{DW}^*|\mathbf{X}}\left(C-1|\mathbf{x}_i\right)\right]^{\delta_i} \\ &= \prod_{i=1}^{n}\left[\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}} - \left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{(y_i+1)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}}\right]^{1-\delta_i}\left[\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{C^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}}\right]^{\delta_i} \end{aligned} \tag{9}$$

and

$$\begin{aligned} l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right) &= \sum_{i=1}^{n}(1-\delta_i)\log\left[\left(e^{-y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right) - \left(e^{-(y_i+1)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)\right] - \sum_{i=1}^{n}\delta_i C^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}} \\ &= \sum_{i=1}^{n}(1-\delta_i)\log\left[w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)\right] - \sum_{i=1}^{n}\delta_i C^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}} \end{aligned} \tag{10}$$

respectively, where $w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right) = e^{-y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}} - e^{-(y_i+1)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}$ and $\boldsymbol{\theta} = \left(\alpha_0,\ldots\alpha_k,\zeta_0,\ldots,\zeta_k\right)$.

The maximum likelihood estimation of the parameters is obtained by set the partial derivatives of the log-likelihood function in Equation 10 with respect to each unknown parameter to zero. Hence, we obtain the first partial derivative of $l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)$ with respect to parameters α_j and ζ_j, $j = 0,1,\ldots,k$, as follows:

$$\frac{\partial}{\partial\alpha_j}l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right) = \sum_{i=1}^{n}\frac{(1-\delta_i)}{w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}\frac{\partial w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}{\partial\alpha_j} - \sum_{i=1}^{n}\delta_i C^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}x_{ij}e^{\mathbf{x}_i\boldsymbol{\alpha}} \tag{11}$$

$$\frac{\partial}{\partial\zeta_j}l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right) = \sum_{i=1}^{n}\frac{(1-\delta_i)}{w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}\frac{\partial w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}{\partial\zeta_j} - \sum_{i=1}^{n}\delta_i\log\left(C\right)C^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}x_{ij}e^{\mathbf{x}_i\boldsymbol{\zeta}}e^{\mathbf{x}_i\boldsymbol{\alpha}} \tag{12}$$

where

$$\frac{\partial w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}{\partial\alpha_j} = -x_{ij}e^{\mathbf{x}_i\boldsymbol{\alpha}}\left[e^{-y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}} - e^{-(y_i+1)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}\left(y_i+1\right)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}\right]$$

$$\frac{\partial w_i\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)}{\partial\zeta_j} = -x_{ij}e^{\mathbf{x}_i\boldsymbol{\zeta}}e^{\mathbf{x}_i\boldsymbol{\alpha}}\left[e^{-y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}\log\left(y_i\right) - e^{-(y_i+1)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}e^{\mathbf{x}_i\boldsymbol{\alpha}}}\left(y_i+1\right)^{e^{\mathbf{x}_i\boldsymbol{\zeta}}}\log\left(y_i+1\right)\right]$$

It can be seen that the maximum likelihood estimators of the parameters α_j and ζ_j, $j=0,1,\ldots,k$, does not have a closed form. Therefore, we estimate the maximum likelihood estimators by using the numerical method that applied in the function `optim( )` from package `stats` in R, that is, minimizing the negative log-likelihood function of the discrete Weibull regression model with type-I right censored data $-l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)$. Then, we get the estimated values $\hat{\theta}_{MLE,CDW}$ of the maximum likelihood estimators of the parameters α_j and ζ_j, $j=0,1,\ldots,k$, where θ is a parameter in vector $\boldsymbol{\theta}=\left(\boldsymbol{\alpha},\boldsymbol{\zeta}\right)$.

2.2. Bayesian estimation

In this section, we perform the Bayesian estimation for the discrete Weibull regression model with type-I right censored data by linking both parameters q and β which are the log-log link function on the parameter q and the log link function on the parameter β. This research investigates the performance of the estimation through uniform noninformative prior, Laplace prior, and normal prior distributions under the parameters are assumed to act independently.

i) Uniform noninformative prior distribution, if no prior information is available, we can resort to a default flat prior then it is easy to focus on the uniform noninformative prior distribution, that is

$$\pi\left(\alpha_j\right)\propto 1,\quad j=0,1,\ldots,k$$
$$\pi\left(\zeta_j\right)\propto 1,\quad j=0,1,\ldots,k$$

ii) Laplace prior distribution, if prior information is available, we can perform the informative prior distribution that should include all possible values of parameter. The possible values of α_j and ζ_j are real number which correspond to the possible values of a Laplace distribution, we selected the prior distribution of α_j and ζ_j as a Laplace distribution with the hyperparameters are $\left(0,1/\lambda\right)$ and $\left(0,1/\tau\right)$respectively, that is

$$\pi\left(\alpha_j\right)=\frac{\lambda}{2}e^{-\lambda\left|\alpha_j\right|},\quad \lambda>0, j=0,1,\ldots,k$$
$$\pi\left(\zeta_j\right)=\frac{\tau}{2}e^{-\tau\left|\zeta_j\right|},\quad \tau>0, j=0,1,\ldots,k$$

iii) Normal prior distribution, as stated earlier, the possible values of α_j and ζ_j are real number which correspond to the possible values of a normal distribution, we selected the prior distribution of α_j and ζ_j as a normal distribution with the hyperparameters are $\left(\mu_{\alpha_j},\sigma^2_{\alpha_j}\right)$ and $\left(\mu_{\zeta_j},\sigma^2_{\zeta_j}\right)$, $j=0,1,\ldots,k$, respectively, that is

$$\pi\left(\alpha_j\right)=\frac{1}{\sqrt{2\pi\sigma^2_{\alpha_j}}}e^{-\frac{1}{2\sigma^2_{\alpha_j}}\left(\alpha_j-\mu_{\alpha_j}\right)^2},\quad \mu_{\alpha_j}\in\mathbb{R},\sigma^2_{\alpha_j}>0, j=0,1,\ldots,k$$

$$\pi\left(\zeta_j\right)=\frac{1}{\sqrt{2\pi\sigma^2_{\zeta_j}}}e^{-\frac{1}{2\sigma^2_{\zeta_j}}\left(\zeta_j-\mu_{\zeta_j}\right)^2},\quad \mu_{\zeta_j}\in\mathbb{R},\sigma^2_{\zeta_j}>0, j=0,1,\ldots,k$$

The joint prior distribution of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ under the independence assumption is

$$\pi(\boldsymbol{\theta}) = \pi(\alpha_0)\cdots\pi(\alpha_k)\pi(\zeta_0)\cdots\pi(\zeta_k) \tag{13}$$

where $\boldsymbol{\theta} = (\alpha_0,\ldots\alpha_k,\zeta_0,\ldots,\zeta_k)$.

The choice of the hyperparameters' values will generally be modified by available information of dataset to improve the Bayes estimators. At this moment, they are left unspecified. The joint posterior density function of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ can be written as:

$$\begin{aligned} & p_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}) \\ & = \frac{L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta})}{\iint\cdots\int L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta})d\alpha_0\cdots d\alpha_k d\zeta_0\cdots d\zeta_k} \\ & \propto L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta}) \end{aligned} \tag{14}$$

where $L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})$ is the likelihood function of the discrete Weibull regression model with type-I right censored data by linking both parameters q and β in Equation 9.

The Bayes estimator of function $h(\boldsymbol{\theta})$ of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ under squared error loss function is the expected value of function $h(\boldsymbol{\theta})$ under the joint posterior density function. Therefore, the Bayes estimator of function $h(\boldsymbol{\theta})$ is given by

$$\hat{h}(\boldsymbol{\theta}) = \iint\cdots\int h(\boldsymbol{\theta})p_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})d\alpha_0\cdots d\alpha_k d\zeta_0\cdots d\zeta_k \tag{15}$$

Since the integral in Equation 15 does not have a closed form, this research chose the random walk Metropolis algorithm, which is a special case of the Metropolis-Hastings algorithm to estimate the Bayes estimators.This research determines the joint posterior density function of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$, $p_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})$, in Equation 14 as the target distribution, while $\boldsymbol{\theta}$ is the current state value, and $\boldsymbol{\theta}^*$ is the proposal value generated from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$. Then, the proposal value $\boldsymbol{\theta}^*$ is accepted with the probability $p = \min(1, R_{\boldsymbol{\theta},CDW})$, where

$$R_{\boldsymbol{\theta},CDW} = \frac{L_{CDW}(\boldsymbol{\theta}^*|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta}^*)}{L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta})} \times \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})} \tag{16}$$

In the random walk Metropolis algorithm, the proposal distribution is symmetrical, depending only on the distance between the current state value and the proposal value. Then, the proposal value $\boldsymbol{\theta}^*$ is accepted with probability $p = \min(1, R_{\boldsymbol{\theta},CDW})$, where

$$R_{\boldsymbol{\theta},CDW} = \frac{L_{CDW}(\boldsymbol{\theta}^*|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta}^*)}{L_{CDW}(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta})\pi(\boldsymbol{\theta})} \tag{17}$$

The iterative steps of the random walk Metropolis algorithm can be described as follows:

Step 1: Initialize the parameters $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\alpha}^{(0)},\boldsymbol{\zeta}^{(0)})$ for the algorithm using the maximum likelihood estimators of the parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha},\boldsymbol{\zeta})$.

Step 2: For $l = 1, 2, \ldots, L$ repeat the following steps;

a. Generate random error vector $\boldsymbol{\varepsilon}$ from a multivariate normal distribution with a zero-mean vector and variance-covariance matrix as a diagonal matrix in which the diagonal elements are the diagonal of the inverse of the observed Fisher's information matrix; $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = diag\left(I_{CDW}^{-1}(\boldsymbol{\theta})\right)\right)$. Then, set $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(l-1)} + \boldsymbol{\varepsilon}$.

b. Calculate $p = \min\left(1, R_{\boldsymbol{\theta},CDW}\right)$ by using Equation 17.

c. Generate u from a uniform distribution; $u \sim U(0,1)$.

If $u \le p$, accept $\boldsymbol{\theta}^*$ and set $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^*$ with probability p.

If $u > p$, reject $\boldsymbol{\theta}^*$ and set $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^{(l-1)}$ with probability $1 - p$.

Step 3: Remove B of the chain for *burn-in*.

Step 4: Calculate the estimated values of the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ from the average of the generated values given by

$$\hat{\theta}_{Bayes,CDW} = \frac{1}{L-B}\sum_{l=B+1}^{L} \theta^{(l)} \tag{18}$$

where θ is a parameter in vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\zeta})$ and B is the length of the *burn-in* period.

3. Simulation study

In this section, the Monte Carlo simulation is conducted to assess and compare the performance of the maximum likelihood estimation and the Bayesian estimation via the random walk Metropolis algorithm for the discrete Weibull regression model with type-I right censored data by linking both parameters q and β, which are the log-log link function on the parameter q and the log link function on the parameter β, under uniform noninformative prior, Laplace prior, and normal prior with various selected sample sizes (n) are 60, 90, 120, and 150. In particular, we construct the over-dispersion response data, we select $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \zeta_0, \zeta_1, \zeta_2, \zeta_3) = (-2.8, 0.01, 0.4, -0.2, 0.1, -0.3, 0.2, 0.2)$. The three explanatory variables are considered: a normal distribution with mean 0 and variance 1 $(x_1 \sim N(0,1))$, a uniform distribution that lies between -0.3 and 0.3 $(x_2 \sim U(-0.3, 0.3))$, and a Bernoulli distribution with probability of success 0.4 $(x_3 \sim Ber(0.4))$.

We fix the hyperparameters' values of α_j and ζ_j, $j = 0,1,2,3$, as 0.5 for Laplace prior. Also, we fix the hyperparameters' values of α_j and ζ_j, $j = 0,1,2,3$, as the maximum likelihood estimators and the variance of the maximum likelihood estimators for normal prior.

We compute q_i and β_i for each type of data from the log-log link function in Equation 6 and the log link function in Equation 7. We then generate the count response variables $Y_1^*, Y_2^*, \ldots, Y_n^*$ from Equation 8 using function `rdw( )` from package `DWreg` in R. We censor the data using type-I right censoring with different percentages of censored data: censoring at $C = 16, 20, 26, 38$ for the over-dispersion response data. Then we get the response variables $y_1, y_2, \ldots, y_n$ as observed data from Equation 3 and the indicator $\delta_1, \delta_2, \ldots, \delta_n$ is the censor indicator from Equation 4.

Next, we calculate the maximum likelihood estimators of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ by minimizing the negative log-likelihood function of the discrete Weibull regression model with the type-I right censored data $-l_{CDW}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)$ in Equation 10. Then, we get $\hat{\boldsymbol{\theta}}^{(m)}_{MLE,CDW}$ using function `optim( )` from package `stats` in R. We calculate the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ with uniform noninformative prior, Laplace prior, and normal prior under the squared error loss function using the random walk Metropolis algorithm with $L=10,000$ replicates and 10% of the chain for *burn-in*; $B=1,000$, respectively. Then, we get $\hat{\boldsymbol{\theta}}^{(m)}_{BayesU,CDW}$, $\hat{\boldsymbol{\theta}}^{(m)}_{BayesL,CDW}$, and $\hat{\boldsymbol{\theta}}^{(m)}_{BayesN,CDW}$ for uniform noninformative prior, Laplace prior, and normal prior from Equation 18, respectively.

For each sample sizes n, it is repeated the previous steps for $m=1,2,\ldots,M$, $M=1,000$ times. The parameter estimates and the mean squared error $\left(MSE\right)$ of estimators from the maximum likelihood estimation (MLE) and the Bayesian estimation under uniform noninformative prior (Bayes(Uniform)), Laplace prior (Bayes(Laplace)), and normal prior (Bayes(Normal)) are calculated as

$$\hat{\theta}_{method,CDW}=\frac{1}{M}\sum_{m=1}^{M}\hat{\theta}^{(m)}_{method,CDW}$$

and

$$MSE\left(\hat{\theta}_{method,CDW}\right)=\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}^{(m)}_{method,CDW}-\theta\right)^{2}$$

respectively, where θ is a parameter in vector $\left(\alpha_0,\alpha_1,\alpha_2,\alpha_3,\zeta_0,\zeta_1,\zeta_2,\zeta_3\right)$ and $method$ is $MLE, BayesU, BayesL,$ and $BayesN$, respectively. The results are reported in Table 1 to Table 4 for censoring at $C=38,26,20,$ and 16, respectively.

Table 1. Parameter estimates (Est.) and MSE at $C = 38$.

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes(Uniform) Est.	Bayes(Uniform) MSE	Bayes(Laplace) Est.	Bayes(Laplace) MSE	Bayes(Normal) Est.	Bayes(Normal) MSE
60 (10.09%)	α_0	-1.4607	0.3584	-1.8012	0.2100	-1.7122	**0.1365**	-1.5084	0.2189
	α_1	-0.0454	0.1149	-0.0407	0.0939	-0.0452	**0.0746**	-0.0457	0.0867
	α_2	0.5669	1.6341	0.2417	2.9610	0.1996	**0.7552**	0.3009	1.2331
	α_3	-0.2492	0.6621	-0.1411	0.4953	-0.1490	**0.2818**	-0.2582	0.4141
	ζ_0	-0.0294	0.0537	0.0726	0.0183	0.0539	**0.0145**	0.0049	0.0289
	ζ_1	-0.1614	0.0119	-0.1677	0.0088	-0.1653	**0.0075**	-0.1584	0.0087
	ζ_2	0.1564	0.2853	0.1373	0.2778	0.1451	**0.1050**	0.0973	0.1478
	ζ_3	0.1617	0.0741	0.1160	0.0414	0.1198	**0.0261**	0.1568	0.0462
90 (10.06%)	α_0	-1.3098	0.3481	-1.5946	0.1403	-1.5483	**0.0989**	-1.4430	0.1629
	α_1	-0.0419	0.0861	-0.0131	0.0579	-0.0192	**0.0494**	-0.0587	0.0619
	α_2	0.4494	1.1633	0.2170	1.7688	0.1687	**0.5362**	0.2602	0.9488
	α_3	-0.2177	0.5186	-0.1300	0.2967	-0.1214	**0.1886**	-0.1824	0.3323
	ζ_0	-0.0309	0.0531	0.0603	0.0124	0.0505	**0.0102**	0.0182	0.0201
	ζ_1	-0.1433	0.0088	-0.1566	0.0052	-0.1543	**0.0048**	-0.1420	0.0064
	ζ_2	0.1348	0.2297	0.1145	0.1604	0.1236	**0.0696**	0.1080	0.1090
	ζ_3	0.1487	0.0580	0.1096	0.0238	0.1078	**0.0165**	0.1331	0.0331
120 (10.01%)	α_0	-1.2315	0.3141	-1.5077	0.0884	-1.4757	**0.0734**	-1.3186	0.1567
	α_1	-0.0516	0.0754	-0.0237	0.0456	-0.0233	**0.0367**	-0.0440	0.0530
	α_2	0.4584	1.1277	0.1393	1.2818	0.1389	**0.5140**	0.2582	0.7977
	α_3	-0.2251	0.5110	-0.1306	0.2019	-0.1282	**0.1506**	-0.1930	0.2930
	ζ_0	-0.0410	0.0460	0.0525	0.0093	0.0455	**0.0084**	0.0074	0.0199
	ζ_1	-0.1336	0.0081	-0.1479	0.0045	-0.1480	**0.0033**	-0.1365	0.0051
	ζ_2	0.1404	0.2000	0.1338	0.1249	0.1284	**0.0519**	0.0872	0.0844
	ζ_3	0.1445	0.0570	0.1075	0.0177	0.1077	**0.0143**	0.1307	0.0302
150 (9.99%)	α_0	-1.2036	0.3102	-1.7553	0.0815	-1.4565	**0.0610**	-1.2897	0.1523
	α_1	-0.0352	0.0564	-0.0130	0.0446	-0.0163	**0.0318**	-0.0405	0.0395
	α_2	0.4717	1.0902	0.2451	1.2659	0.1285	**0.4812**	0.2788	0.6722
	α_3	-0.1823	0.4725	-0.1422	0.1961	-0.1298	**0.1233**	-0.2006	0.2925
	ζ_0	-0.0497	0.0457	0.0604	0.0086	0.0445	**0.0068**	-0.0011	0.0186
	ζ_1	-0.1363	0.0060	-0.1779	0.0040	-0.1481	**0.0030**	-0.1377	0.0042
	ζ_2	0.1280	0.1907	0.1349	0.1137	0.1282	**0.0517**	0.0847	0.0743
	ζ_3	0.1368	0.0559	0.1271	0.0172	0.1095	**0.0114**	0.1365	0.0291

Note: the boldface identifies the smallest MSE for each case.

Table 2. Parameter estimates (Est.) and MSE at $C = 26$.

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes(Uniform) Est.	Bayes(Uniform) MSE	Bayes(Laplace) Est.	Bayes(Laplace) MSE	Bayes(Normal) Est.	Bayes(Normal) MSE
60 (15.25%)	α_0	-1.7792	0.3889	-2.1399	0.2675	-2.0162	**0.1627**	-1.7378	0.2313
	α_1	-0.0011	0.1344	-0.0444	0.1052	-0.0411	**0.0743**	-0.0273	0.1011
	α_2	0.5948	1.4529	0.2844	3.3846	0.2328	**0.7171**	0.4293	1.2225
	α_3	-0.3264	0.8022	-0.1816	0.5631	-0.1878	**0.3254**	-0.2946	0.5520
	ζ_0	-0.0211	0.0592	0.0820	0.0245	0.0547	**0.0194**	0.0141	0.0319
	ζ_1	-0.2111	0.0148	-0.1995	0.0107	-0.2009	**0.0087**	-0.1895	0.0111
	ζ_2	0.2091	0.2955	0.1535	0.3142	0.1613	**0.1059**	0.0954	0.1665
	ζ_3	0.2050	0.0877	0.1470	0.0486	0.1491	**0.0318**	0.1803	0.0592
90 (15.33%)	α_0	-1.6064	0.3619	-1.9408	0.1725	-1.8763	**0.1267**	-1.6857	0.2065
	α_1	-0.0213	0.1055	-0.0290	0.0782	-0.0282	**0.0618**	-0.0207	0.0688
	α_2	0.5229	1.3838	0.3035	2.2799	0.2140	**0.7137**	0.3302	1.0958
	α_3	-0.2953	0.6609	-0.1612	0.3784	-0.1702	**0.2638**	-0.3296	0.4623
	ζ_0	-0.0353	0.0571	0.0740	0.0161	0.0595	**0.0136**	-0.0005	0.0302
	ζ_1	-0.1864	0.0112	-0.1871	0.0073	-0.1871	**0.0063**	-0.1922	0.0076
	ζ_2	0.1891	0.2640	0.1301	0.2097	0.1485	**0.0870**	0.1148	0.1431
	ζ_3	0.1958	0.0725	0.1349	0.0314	0.1384	**0.0233**	0.1948	0.0494
120 (15.32%)	α_0	-1.4919	0.3321	-1.7769	0.0941	-1.7365	**0.0738**	-1.5846	0.1799
	α_1	-0.0185	0.0794	-0.0174	0.0495	-0.0209	**0.0408**	-0.0311	0.0544
	α_2	0.4509	1.2593	0.2388	1.5356	0.1893	**0.5865**	0.2750	0.9843
	α_3	-0.2550	0.5859	-0.1358	0.2291	-0.1354	**0.1670**	-0.2568	0.3276
	ζ_0	-0.0326	0.0569	0.0655	0.0089	0.0571	**0.0079**	0.0031	0.0265
	ζ_1	-0.1742	0.0081	-0.1769	0.0048	-0.1763	**0.0041**	-0.1772	0.0061
	ζ_2	0.1921	0.2595	0.1407	0.1359	0.1510	**0.0633**	0.1307	0.1204
	ζ_3	0.1731	0.0673	0.1215	0.0199	0.1208	**0.0157**	0.1667	0.0369
150 (15.50%)	α_0	-1.4548	0.3169	-1.4819	0.0689	-1.7250	**0.0688**	-1.4902	0.1720
	α_1	-0.0221	0.0755	-0.0144	0.0363	-0.0123	**0.0356**	-0.0204	0.0436
	α_2	0.4120	1.1920	0.1425	1.1843	0.2044	**0.5319**	0.2628	0.6966
	α_3	-0.2452	0.5722	-0.1333	0.1623	-0.1306	**0.1399**	-0.2815	0.3183
	ζ_0	-0.0494	0.0497	0.0501	0.0079	0.0536	**0.0076**	-0.0005	0.0250
	ζ_1	-0.1727	0.0080	-0.1489	0.0037	-0.1788	**0.0034**	-0.1693	0.0052
	ζ_2	0.1919	0.2327	0.1277	0.1059	0.1424	**0.0631**	0.1152	0.0817
	ζ_3	0.1772	0.0667	0.1101	0.0139	0.1248	**0.0129**	0.1706	0.0367

Note: the boldface identifies the smallest MSE for each case.

Table 3. Parameter estimates (Est.) and MSE at $C = 20$.

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes(Uniform) Est.	Bayes(Uniform) MSE	Bayes(Laplace) Est.	Bayes(Laplace) MSE	Bayes(Normal) Est.	Bayes(Normal) MSE
60 (19.94%)	α_0	-1.8742	0.3874	-2.2190	0.3166	-2.0879	**0.1979**	-1.9482	0.2542
	α_1	0.0033	0.1589	-0.0375	0.1362	-0.0390	**0.1075**	0.0083	0.1116
	α_2	0.5743	1.5242	0.3297	3.7867	0.2360	**0.8569**	0.3479	1.5456
	α_3	-0.3489	0.8737	-0.2152	0.7085	-0.2208	**0.4183**	-0.3463	0.6566
	ζ_0	-0.0142	0.0616	0.0841	0.0282	0.0551	**0.0223**	0.0196	0.0383
	ζ_1	-0.2223	0.0186	-0.2083	0.0130	-0.2073	**0.0113**	-0.2221	0.0130
	ζ_2	0.2070	0.3156	0.1571	0.3518	0.1691	**0.1249**	0.1530	0.2091
	ζ_3	0.2147	0.0936	0.1585	0.0559	0.1603	**0.0375**	0.2069	0.0697
90 (20.37%)	α_0	-1.8431	0.3579	-2.1631	0.1820	-2.0840	**0.1265**	-1.9507	0.2133
	α_1	0.0067	0.1227	-0.0381	0.0848	-0.0371	**0.0669**	0.0066	0.0784
	α_2	0.5892	1.4734	0.3912	2.6204	0.2834	**0.7888**	0.4578	1.1513
	α_3	-0.3098	0.7688	-0.1559	0.3957	-0.1594	**0.2556**	-0.3132	0.4583
	ζ_0	-0.0222	0.0604	0.0843	0.0178	0.0657	**0.0146**	0.0133	0.0324
	ζ_1	-0.2194	0.0124	-0.2046	0.0086	-0.2053	**0.0072**	-0.2258	0.0089
	ζ_2	0.1889	0.2948	0.1286	0.2447	0.1497	**0.0957**	0.1126	0.1416
	ζ_3	0.2111	0.0825	0.1434	0.0356	0.1455	**0.0244**	0.2018	0.0510
120 (20.37%)	α_0	-1.7795	0.3275	-2.1070	0.1315	-2.0449	**0.0945**	-1.9452	0.1967
	α_1	0.0233	0.0905	-0.0287	0.0636	-0.0234	**0.0533**	-0.0135	0.0656
	α_2	0.5224	1.4404	0.2140	1.8130	0.1962	**0.6154**	0.3899	1.0331
	α_3	-0.2925	0.6066	-0.1505	0.2616	-0.1588	**0.1786**	-0.2602	0.4164
	ζ_0	-0.0323	0.0587	0.0801	0.0131	0.0658	**0.0110**	0.0255	0.0284
	ζ_1	-0.2227	0.0096	-0.2043	0.0062	-0.2061	**0.0055**	-0.2153	0.0076
	ζ_2	0.2174	0.2533	0.1818	0.1788	0.1790	**0.0801**	0.1297	0.1304
	ζ_3	0.1972	0.0688	0.1397	0.0253	0.1421	**0.0179**	0.1833	0.0434
150 (20.39%)	α_0	-1.8032	0.3010	-2.2805	0.1213	-2.0591	**0.0818**	-1.8791	0.1886
	α_1	-0.0154	0.0844	-0.0363	0.0553	-0.0253	**0.0426**	0.0033	0.0531
	α_2	0.5982	1.2578	0.2259	1.7968	0.2551	**0.6138**	0.3102	0.8509
	α_3	-0.2612	0.5896	-0.1467	0.2569	-0.1497	**0.1569**	-0.2670	0.3484
	ζ_0	-0.0306	0.0521	0.0892	0.0130	0.0689	**0.0091**	0.0111	0.0276
	ζ_1	-0.2117	0.0091	-0.2208	0.0059	-0.2093	**0.0045**	-0.2168	0.0066
	ζ_2	0.1787	0.2477	0.1965	0.1759	0.1575	**0.0681**	0.1550	0.1095
	ζ_3	0.1953	0.0713	0.1482	0.0240	0.1403	**0.0150**	0.1883	0.0401

Note: the boldface identifies the smallest MSE for each case.

Table 4. Parameter estimates (Est.) and MSE at $C=16$.

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes(Uniform) Est.	Bayes(Uniform) MSE	Bayes(Laplace) Est.	Bayes(Laplace) MSE	Bayes(Normal) Est.	Bayes(Normal) MSE
60 (25.26%)	α_0	-2.1469	0.4203	-2.4901	0.3301	-2.3483	**0.2028**	-2.1679	0.3120
	α_1	0.0449	0.1642	-0.0266	0.1488	-0.0246	**0.1098**	0.0216	0.1184
	α_2	0.7561	1.6600	0.2408	4.2706	0.2262	**0.9225**	0.4704	1.5203
	α_3	-0.3758	0.9146	-0.2466	0.6925	-0.2524	**0.4320**	-0.3622	0.7577
	ζ_0	-0.0086	0.0697	0.0897	0.0336	0.0585	**0.0270**	0.0213	0.0471
	ζ_1	-0.2677	0.0189	-0.2440	0.0153	-0.2446	**0.0128**	-0.2527	0.0149
	ζ_2	0.1954	0.3552	0.2046	0.4385	0.1986	**0.1537**	0.1441	0.2272
	ζ_3	0.2379	0.1053	0.1761	0.0637	0.1793	**0.0443**	0.2282	0.0786
90 (25.60%)	α_0	-2.0807	0.3513	-2.3889	0.2411	-2.2986	**0.1677**	-2.1204	0.2261
	α_1	0.0258	0.1271	-0.0312	0.0975	-0.0284	**0.0755**	0.0082	0.0912
	α_2	0.6514	1.5463	0.3310	2.7760	0.2573	**0.8760**	0.4442	1.2601
	α_3	-0.3215	0.7750	-0.1858	0.4473	-0.1864	**0.2970**	-0.3204	0.5260
	ζ_0	-0.0074	0.0653	0.0911	0.0256	0.0695	**0.0213**	0.0174	0.0354
	ζ_1	-0.2544	0.0150	-0.2325	0.0105	-0.2343	**0.0088**	-0.2455	0.0106
	ζ_2	0.1919	0.3267	0.1738	0.2753	0.1878	**0.1186**	0.1341	0.1830
	ζ_3	0.2212	0.0887	0.1625	0.0434	0.1640	**0.0317**	0.2148	0.0568
120 (25.52%)	α_0	-1.9825	0.3332	-2.2976	0.1467	-2.2331	**0.1069**	-2.1026	0.2060
	α_1	0.0293	0.1044	-0.0252	0.0648	-0.0255	**0.0545**	0.0111	0.0716
	α_2	0.6955	1.4621	0.3229	2.0960	0.2260	**0.7359**	0.4577	1.0062
	α_3	-0.3336	0.6988	-0.1720	0.3256	-0.1775	**0.2260**	-0.3331	0.4683
	ζ_0	-0.0304	0.0592	0.0809	0.0166	0.0661	**0.0142**	0.0159	0.0328
	ζ_1	-0.2482	0.0127	-0.2279	0.0069	-0.2288	**0.0062**	-0.2464	0.0089
	ζ_2	0.1916	0.2716	0.1685	0.2108	0.1887	**0.0953**	0.1284	0.1325
	ζ_3	0.2278	0.0818	0.1577	0.0314	0.1593	**0.0231**	0.2170	0.0522
150 (25.31%)	α_0	-1.9913	0.3172	-2.1056	0.1128	-2.2301	**0.0888**	-2.0936	0.1855
	α_1	0.0033	0.0883	-0.0285	0.0521	-0.0292	**0.0463**	0.0170	0.0593
	α_2	0.6198	1.2316	0.3199	1.5797	0.2074	**0.6957**	0.4291	0.8145
	α_3	-0.3048	0.6740	-0.1447	0.2292	-0.1591	**0.1926**	-0.3539	0.4527
	ζ_0	-0.0160	0.0562	0.0799	0.0110	0.0768	**0.0109**	0.0176	0.0295
	ζ_1	-0.2355	0.0099	-0.2079	0.0053	-0.2234	**0.0052**	-0.2424	0.0071
	ζ_2	0.1986	0.2483	0.1440	0.1474	0.1956	**0.0825**	0.1410	0.1115
	ζ_3	0.2169	0.0758	0.1387	0.0201	0.1524	**0.0197**	0.2208	0.0493

Note: the boldface identifies the smallest MSE for each case.

The results of the simulation study show that all estimators have monotonic behaviors according to the MSE, namely, when n increases, the estimated MSE values decrease. Moreover, the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ with Laplace prior for the over-dispersion response data show the best performance for all cases in terms of the MSE of the estimators.

4. Conclusion

In this research, we have considered the classical and Bayesian inference for the discrete Weibull regression model with type-I right censored data where the shape parameters q and β are related to explanatory variables via the log-log and the log link functions, respectively. Moreover, we choose the random walk Metropolis algorithm to estimate the Bayes estimators with uniform noninformative prior, Laplace prior, and normal prior.

The results of the simulation showed that as the sample size increases, the MSE of the estimators decreases for all methods, indicating that the estimators are consistent. In addition, the Bayesian estimation with Laplace prior is more appropriate for the discrete Weibull regression model with type-I right censored data than other methods in terms of the MSE of the estimators for over-dispersion response data.

However, there is some computational challenge to be faced while implementing the Bayesian approach which is the selection of hyperparameters' values of the prior distribution that may affect the parameter estimates include of the sampling method in the MCMC techniques. Future research will apply the estimators to real data, explore other link functions on parameters, and consider other types of response data: excessive zeros data and under-dispersion data.

References

[1] Allison PD and Waterman RP 2002 Fixed-effects negative binomial regression models. (*Sociological Methodology* vol 32(1)) pp 247-265

[2] Cameron AC and Trivedi PK 2013 *Regression Analysis of Count Data*, 2nd ed (Cambridge University Press)

[3] Chanialidis C, Evers L, Neocleous T and Nobile A 2018 Efficient Bayesian inference for COM-Poisson regression models (*Statistics and Computing* vol 28(1)) pp 595–608

[4] Frome EL 1983 The analysis of rates using Poisson regression models (*Biometrics* vol 39(3)) pp 665-674

[5] Fu S 2016 Hierarchical Bayesian LASSO for a negative binomial regression (*Journal of Statistical Computation and Simulation* vol 86(11)) pp 2182-2203

[6] Gelman A, Jakulin A, Pittau MG and Su Y 2008 A weakly informative default prior distribution for logistic and other regression models (*The Annals of Applied Statistics* vol 2(4)) pp 1360-1383

[7] Haselimashhadi H, Vinciotti V and Yu K 2018 A novel Bayesian regression model for counts with an application to health data (*Journal of Applied Statistics* vol 45(6)) pp 1085-1105

[8] Lee ET and Wang JW 2013 *Statistical methods for survival data analysis*, 3rd ed (Hoboken: John Wiley & Sons) p 259-263

[9] Liu H, Davidson RA, Rosowsky DV and Stedinger JR 2005 Negative binomial regression of electric power outages in hurricanes (*Journal of Infrastructure Systems* vol 11(4)) pp 258-267

[10] Lovett A and Flowerdew R 1989 Analysis of count data using Poisson regression (*The Professional Geography* vol 41(2)) pp 190–198

[11] Nakagawa T and Osaki S 1975 The discrete Weibull distribution (*IEEE Transactions on Reliability* vol R-24(5)) pp 300-301

[12] Nash JC 1990 *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, 2nd ed (Bristol: Adam Hilger)
[13] Obeidat M, Al-Nasser A and Al-Omari AI 2020 Estimation of generalized Gompertz distribution parameters under ranked-set sampling (*Journal of Probability and Statistics* ID 7362657) pp 14
[14] Saraiva EF and Suzuki AK 2017 Bayesian computational methods for estimation of two-parameters Weibull distribution in presence of right-censored data (*The Chilean Journal of Statistics* vol 8(2)) pp 25-43
[15] Kalktawi HS 2017 *Discrete Weibull regression model for count data*, PhD dissertation (London: Brunel University London)
[16] Hilbe JM 2016 *COUNT: Functions, Data and Code for Count Data* (R package version 1.3.4)
[17] R Core Team 2020 *R: A language and environment for statistical computing* (Vienna: R Foundation for Statistical Computing)
[18] Vinciotti V 2016 *DWreg: Parametric Regression for Discrete Response* (R package version 2.0)

Development of Electrochemical Dopamine Sensor based on Conducting Polymer-Gold Nanoparticle Composites

Rapiphun Janmanee[1,*], Sosista Chatwet[1], Nutsarin Sukonpakdee[1], Anchana Preechaworapun[1], Kulwadee Pinwattana[1], Padarat Ninjiaranai[1], Saengrawee Sriwichai[2]
[1]Chemistry program, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok 65000, Thailand
[2]Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

**Corresponding Author Email: rapiphun16@psru.ac.th*

Abstract. In this research, polypyrrole/polypyrrole-3-carboxylic acid/gold nanoparticle (PPy/PP3C/Au) composites material was developed for electrochemical determination of dopamine (DA). PPy/PP3C/Au composites thin film was synthesized by electropolymerization on fluorine doped tin oxide (FTO) coated glass slide. The cyclic voltammetry and amperometry techniques were used to study the efficiency of the obtained conducing polymer-gold nanoparticle composites thin film for the determination of dopamine. The results indicated that PPy/PP3C/Au composites thin film was successfully fabricated by electropolymerization method. The cyclic voltammetry measurements of the synthesized PPy/PP3C/Au composites thin film in neutral phosphate-buffered saline (PBS) solution indicated that the presented conducing polymer-gold nanoparticle composites thin film exhibited a good electroactivity property in PBS solution. Moreover, the amperometric results revealed that the proposed electrochemical sensor showed good sensitivity and selectivity for the detection of dopamine with a wide detection linear range from 20 to 200 μM. Therefore, it can be concluded that the electrochemically fabricated PPy/PP3C/Au composites thin film can be applied as an electrochemical sensor for the detection of dopamine in real sample for the early diagnosis of various neurological disorders.

Keywords: conducting polymer; gold nanoparticle; dopamine; electrochemical; sensor

1. Introduction

Dopamine is a type of neurotransmitter. It plays important roles in mood, learning, sleep, attention, memory, movement and anticipation [1]. Dopamine dysfunction is the cause of a handful of diseases, most notably Parkinson's disease which is caused by the death of dopamine-producing cells [2]. Therefore, quantification of dopamine in biological fluids is a very important task.

So far, various analytical methods have been developed for dopamine level monitoring such as Calorimetric methods, Spectroscopy methods and Chromatography methods. These methods are complex, time consuming, expensive and have selective detection limits [3-4]. Recently, an electrochemical detection methods have been employed in the determination of dopamine owing to its very compact, inexpensive, highly sensitive and good selective for detection of dopamine and other neurotransmitters [5].
Moreover, the preparation of conducting polymer films by electropolymerization on an electrode surface is another common method to improve the performance of modified electrodes for the electrochemical sensor application, which has high selectivity and sensitivity towards different analytes [6]. Recently, composite nanomaterials based on conducting polymers (CPs) and metal nanoparticles (NPs) of different metals such as Ag Pt Pu and Au have been developed. Gold nanoparticles (Au) have been increasingly interested for the fabrication of electrochemical sensors to improve their analytical performance attributed to its unique properties including excellent charge transfer capability, great chemical stability, rich surface chemistry and good catalytic ability [7-9].
Considering the individual advantages of the CPs and Au, this study aims to develop an electrochemical sensor based on polypyrrole/polypyrrole-3-carboxylic acid/gold nanoparticle (PPy/PP3C/Au) composites material for the determination of dopamine. We believe that the proposed composite material opens a new avenue towards application in electrochemical dopamine biosensors.

2. Materials and methods

2.1. Materials and apparatus

Pyrole (Py), pyrrole-3-carboxylic acid (P3C) and gold solution (Au) were purchased from Sigma-Aldrich Company. All other chemicals were analytical grade with highest purity and used as received.
All electropolymerization and electrochemical measurements were performed on a PC-controlled AUTOLAB PGSTAT204 potentiostat-galvanostat (Metrohm Autolab B.V., The Netherlands). A standard three-electrode cell consisting of an Ag/AgCl (3M KCl) reference electrode, a platinum wire counter electrode and fluorine doped tin oxide (FTO) coated glass slide working electrode (area 0.25 cm^2) were employed.

2.2. Electrochemical synthesis of PPy/PP3C/Au composites thin film

The precursor solution containing 40 µl of Au in a monomer solution of 10 mM Py/P3C in 0.5 M H_2SO_4 solution was used for electropolymerization of PPy/PP3C/Au composites thin film on FTO, which was used as a working electrode. Platinum wire and Ag/AgCl were used as the counter and reference electrode respectively. The electropolymerization was performed by cyclic voltammetry in a range from -1.0 to 1.0 V at a scan rate of 10 mV/s for 5 cycles. The fabrication of PPy/PP3C/Au composites thin film was monitored by cyclic voltammetry (CV).

2.3. Determination of dopamine

The cyclic voltammetry and amperometry techniques were used to study the efficiency of the obtained composites thin film for the determination of dopamine. In addition, the electrochemical behavior of the composites thin film in a neutral phosphate-buffered saline (PBS) solution (pH 7.4) was also studied.

3. Results and discussion

3.1. Electropolymerization of PPy/PP3C/Au composites thin film

The electrochemical property during the electrochemical fabrication of PPy/PP3C/Au composites thin film was monitored using an electrochemical measurement. Figure 1 shows the cyclic voltammograms (CVs) during the electrochemical fabrication of PPy/PP3C/Au composites thin film. The current in the anodic scan increased about 0.6 V, indicating the beginning of the formation of polymer film on FTO electrodes. The current slightly decreased

about 0.3 V in the cathodic scan, which indicated the dedoping process of the deposited copolymer films [10]. The peak observed at -0.3 V can be assumed that, the presence of gold nanoparticles on the modified working electrode. However, the morphological characterization of PPy/PP3C/GO composites thin film will be investigated with scanning electron microscope (SEM).

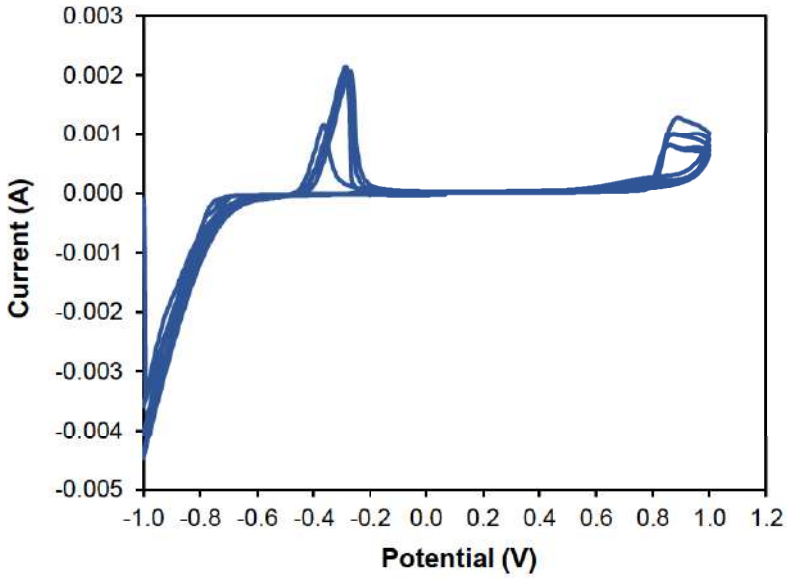


Figure 1. CV of the electrochemical fabrication of PPy/PP3C/Au composites thin film.

3.2. Electrochemical characterization of PPy/PP3C/Au composites thin film

Figure 2 shows the CVs of the PPy/PP3C/Au composites thin film deposited on an FTO glass substrate in PBS solution at different scan rates of 5, 10, 20, 50 and 100 mV/s. The couple of redox peak currents increased with the increase of potential scan rate (figure 2 (a)), indicating that the mass and electron transfer occur at the electrode surface [11]. In addition, the plot of peak current versus scan rate exhibited good linearity relationship (figure 2 (b)), suggesting a surface confined redox process [12].

The current responses of PPy/PP3C/Au composites thin film modified electrode on 20 μM of dopamine compared to the bare FTO electrode is shown in figure 3. The bare does not exhibit any obvious current peaks in the electrochemical processes. The high current response obtained for the modified electrode, is due to the further increase of redox species interactions between electrode and dopamine [3]. When an electrode potential is applied, dopamine is easily oxidized to dopamine-quinone with the exchange of two electrons and protons between electrode and dopamine, the changes in currents due to the electrons released by dopamine during its oxidation are measured [1, 13].

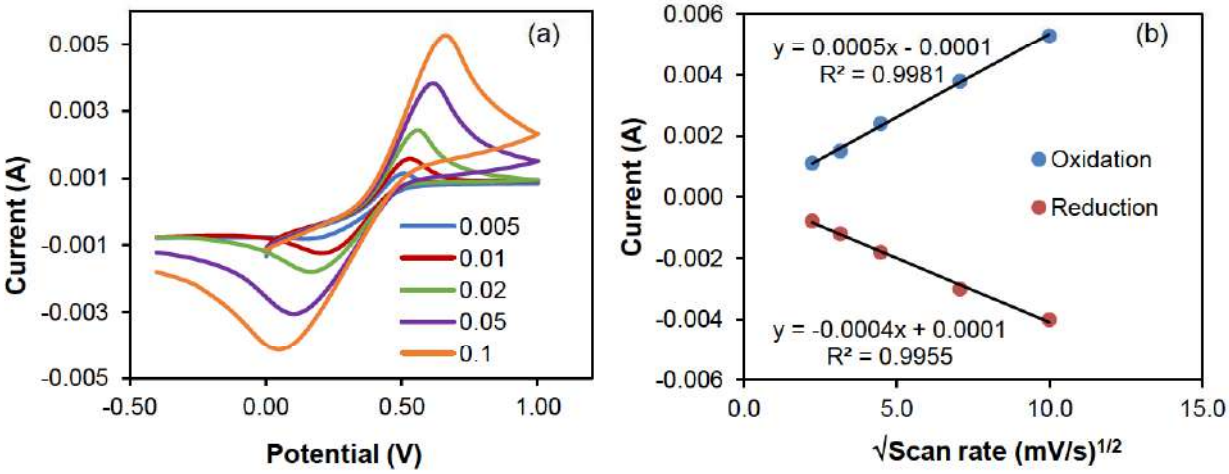


Figure 2. (a) CVs of PPy/PP3C/Au composites thin film in PBS solution (pH 7.4) at different scan rates (5, 10, 20, 50 and 100 mV/s) and (b) plots of peak current versus scan rate.

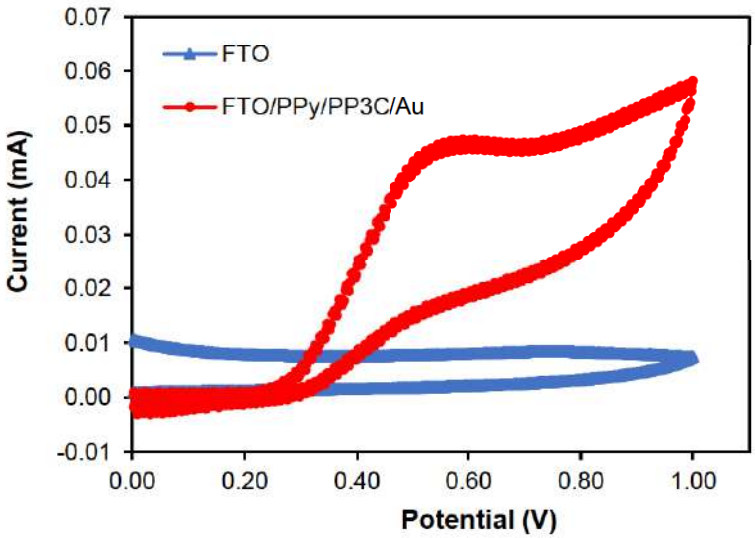


Figure 3. CV of FTO and PPy/PP3C/Au in the presence of 20 μM dopamine.

3.3. Performance of PPy/PP3C/Au modified electrode for the detection of dopamine

The amperogram of the electrochemical dopamine sensor based PPy/PP3C/Au composites thin film for various concentrations of dopamine from 20-200 μM at an applied potential of +0.65 V is shown in figure 4. The rapid and well defined responses were observed for each addition. The response time is about 10 s and the recovery time is found to be 15 s.

In figure 5 shows the clear amperometric current responses of dopamine in the wild linear range of 20-200 μM. The linear relationship between the concentration of dopamine and the current signal response was observed for two concentration ranges of 20-60 μM and 80-200 μM with correlation coefficients are (R^2) 0.9998 and 0.9945 respectively. The results indicated precise sensitivity and wide sensing ability towards the detection of dopamine which is superior to or similar to the previously reported sensing platform [14].

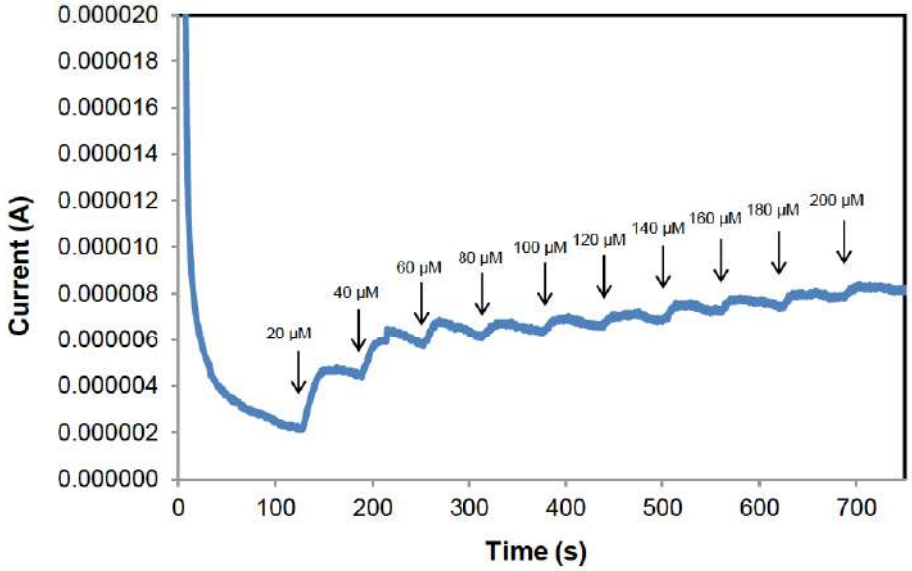


Figure 4. Amperometric response of PPy/PP3C/Au modified electrode at an applied potential of +0.65 V for the detection of dopamine.

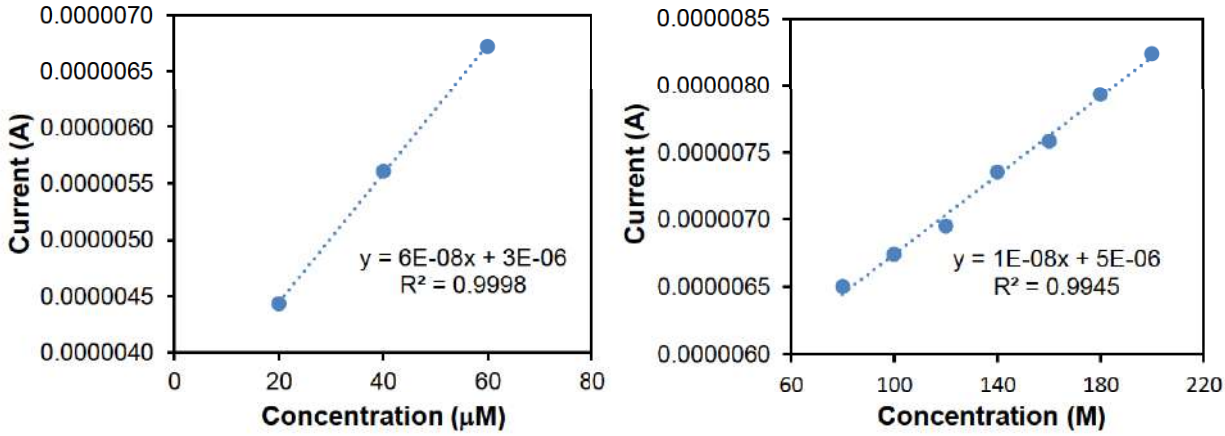


Figure 5. The linear relationship of the response current versus dopamine concentration.

4. Conclusions

In this work, PPy/PP3C/Au composites thin film was successfully prepared by an electrochemical polymerization. The presented PPy/PP3C/Au composites thin film exhibited a good electroactivity property in neutral PBS solution which may possess potential applications in various systems of the biosensors. The prepared PPy/PP3C/Au electrode proved to be sensitive and selective detection of dopamine by means of electrochemical measurements. Amperometric results conveyed the wide detection linear range from 20 to 200 μM. The obtained results revealed that the electrochemically fabricated PPy/PP3C/Au composites thin film can be applied as an electrochemical sensor for the detection of dopamine in real sample with good precision and accuracy for the early diagnosis of various neurological disorders.

Acknowledgements
We gratefully acknowledge Faculty of Science and Technology, Pibulsongkram Rajabhat University and the Research and Development Institute of Pibulsongkram Rajabhat University (RDI-2-64-34), Thailand for financial support.

References
[1] Choo S S, Kang E S, Song I, Lee D, Choi J W and Kim T H 2017 *Sensors* 17 pp 861
[2] Lakard S, Pavel I A and Lakard B 2021 *Biosensors* 11 pp 179
[3] Anbumannan V, Rajendra-Kumar R T and Suresh K 2021 *Materials Science in Semiconductor Processing* 127 pp 105696
[4] Özcan A, İlkbaş S and Özcan A A 2017 *Talanta* 165 pp 489–495
[5] Damphathik C, Kalcher K, Raberb G, Ortner A and Samphao A 2019 *Journal of Science and Technology Ubon Ratchathani University* 21(2) pp 171-180
[6] Fernández I, Trueba M, Núñez C A and Rieumont J 2002 *Revista CENIC Ciencias Químicas* 33(3) pp 121-125
[7] García-Hernández C, García-Cabezón C, Medina-Plaza C, Martín-Pedrosa F, Blanco Y, Antonio de Saja J and Rodríguez-Méndez M L 2015 *Beilstein Journal of Nanotechnology* 6 pp 2052–2061
[8] Sofia Anuar N S, Basirun W J, Shalauddinab M and Akhtera S 2020 *The Royal Society of Chemistry* 10 pp 17336–17344
[9] Thunyakontirakun W, Sriwichai S, Phanichphant S and Janmanee R 2019 *Materials Today: Proceedings* 17(4) pp 2070-2077
[10] Janmanee R, Baba A, Phanichphant S, Sriwichai S, Shinbo K, Kato K and Kaneko K 2011 *Japanese Journal of Applied Physics* 50 pp 01BK02–01– 01BK02–06
[11] Madani A, Nessark B, Brayner R, Elaissari H, Jouini M, Mangeney C and Chehimi M M 2010 *Polymer* 51 pp 2825–2835
[12] Peng H P, Hu Y, Liu P, Deng Y N, Wang P, Chen W, Liu A L, Chen Y Z and Lin X H 2015 *Sensors and Actuators B* 207 pp 269–276
[13] Lakard S, Pavel I A and Lakard B 2021 *Biosensors* 11 pp 179
[14] Eddin F B K and Fen Y W 2020 *Sensors* 20 pp 1039

Purification humic acid as precursor to synthesize reduced graphene oxide by hydrothermal with microwave assisted method

Kanyawee Kaewpradub, Ratchadaporn Puntharod*
Chemistry program, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand
**Corresponding Author Email: ratchadaporn_p@mju.ac.th*

Abstract. Humic acid is a naturally occurring substance. It is inexpensive natural material. The value of humic acid must be increased by used as the precursor to synthesize reduced graphene oxide, which has electrical conductivity property. According to the humic acid from laboratory section, Mae Moh mine, Lampang province, has 11 % of carbon. Therefore, humic acid must be purified by soaked in 2.0 M of sodium hydroxide and left at room temperature for 24, 48 and 72 hr. The precipitation was occurred by using 2.0 M of sulfuric acid as precipitant then adjusted the pH to 1-2. After that the mixture was centrifuged at 3000 rpm, filtered and dried. The results revealed that the soaking of humic acid in sodium hydroxide for 72 hr provided the highest the percentage of yield as 5.79. The purified humic acid was confirmed by X-ray diffractometry. Further, it was used as the precursor to synthesis reduced graphene oxide by hydrothermal with microwave assisted method. The parameters of synthesis were studied the watt of microwave in range of 300-800 watt for 3-12 min. The products were investigated by UV-Vis spectroscopy which found the λ_{max} at 242-246 nm. The products were analyzed by Fourier transform infrared spectroscopy which showed the characteristic of reduced graphene oxide at 1618 cm^{-1}. X-ray diffraction (XRD) showed the 2θ at 23.5°, 32.0° and 45.0°.

Keywords: Humic acid; Reduced graphene oxide; Hydrothermal with microwave assisted method.

1. Introduction

In the mining process, there is subproduct of mining. which is a layer of soil mixed with coal called "leonardite". There are many elements both of useful organic and inorganic for agriculture [1]. Humic substances is a part of leonardite. It has good properties for agriculture such as high cation exchange capacity. Also, humic substance can retain water and plant nutrients. As well as there are many types of plant nutrients because humic substances contain large amounts of carbon atoms. The researchers reported the utilization of humic substances as a precursor for synthesis of reducing graphene oxide (rGO) and graphene oxide (GO) [2,3,4]. Also, there are methods for synthesis of rGO and GO using humic substances. Such as Huang *et.al.* [4] reported the hydrothermal synthesis of rGO from humic acid. The area capacitance was 28 $\mu F\ cm^{-2}$. Smith *et.al.* [5] synthesized rGO from humic acid

by Hummer's method which obtained electrical conductivities of 4.65 x 10^3 S/m. Regarding to the previous research, the rGO and GO could be synthesized by hydrothermal processes. In this research the rGO was synthesized from humic acid by hydrothermal with microwave assisted method because it was easily, economically, quickly and environmental friendly method. The products were investigated by UV-Vis spectroscopy and confirmed by Fourier transform infrared spectroscopy and X-ray diffractometry.

2. Experiment

2.1 Purification of humic acid.

The 5.00 g of humic acid was immersed into 50.00 mL of 2M NaOH for 24, 48 and 72 hr. Then the precipitation was filtered and 2M H_2SO_4 wad added into solution followed by adjusted pH to 1-2. After that the centrifugation was 3000 rpm for 20 min. The precipitate was heated at 60 °C until dried. Finally, the black powder was obtained as purified humic acid. It was analyzed by X-ray diffractometry. The percentage of purified humic acid yield was determined as equation (1).

$$\%\ yield = \frac{wt.of\ dried\ precipitate\ (g)}{wt.of\ humic\ acid\ (g)} \times 100 \qquad (1)$$

2.2 Synthesis of reduced graphene oxide from humic acid using the hydrothermal with microwave assisted method.

The 1.00 g of purified humic acid was put into 10.00 mL of deionized water and followed by sonicated in ultrasonic bath (GT sonic Brand R-series model) for 10 min to make sure the dissolution was occurred properly. The solution was transferred to a Teflon autoclave and put in the microwave oven. The parameters of synthesis were studied the watt of microwave in range of 300-800 watt and reaction time for 3-12 min. The products were investigated by UV-Vis spectroscopy (Hitachi brand UH5300 model), Fourier transform infrared spectroscopy (PerkinElmer brand spectrum RX model) and X-ray diffractometry (Bruker brand D2 phaser model).

3. Results

3.1 The immersing time of purified humic acid from humic substance.

The extraction time was 72 hr provided the highest percentage of yield of purified humic acid equal to 5.79% as in Table 1. The product solid was a dark brown to black powder shown in Figure 1.

Table 1. The percentage of yield of purified humic acid was extracted by 2M NaOH with various immersing time and precipitated by 2M H_2SO_4.

Humic : 2M NaOH	Time (hours)	% yield of purified humic acid			
		1	2	3	$\bar{x}$
5.00 g : 50.00 mL.	24	3.27	3.48	2.99	3.24
	48	3.43	3.45	2.63	3.27
	72	5.60	6.34	5.42	5.79

Figure 1. The purified humic acid.

3.2. XRD of purified humic acid.

The characterization using X-ray diffraction was shown in Figure 2. The identify of purified humic acid at 2θ = 17°, 25°, 32° and 38°. Based on experimental results compared to diffractogram of the purified humic acid show at 2θ =18°, 25°, 31° and 38° reported by Amanda M.P. *et.al.* [6]. The XRD results indicated that the purified humic acid was crystalline and more purer. However, silica was contained ad shown at 2θ = 28°.

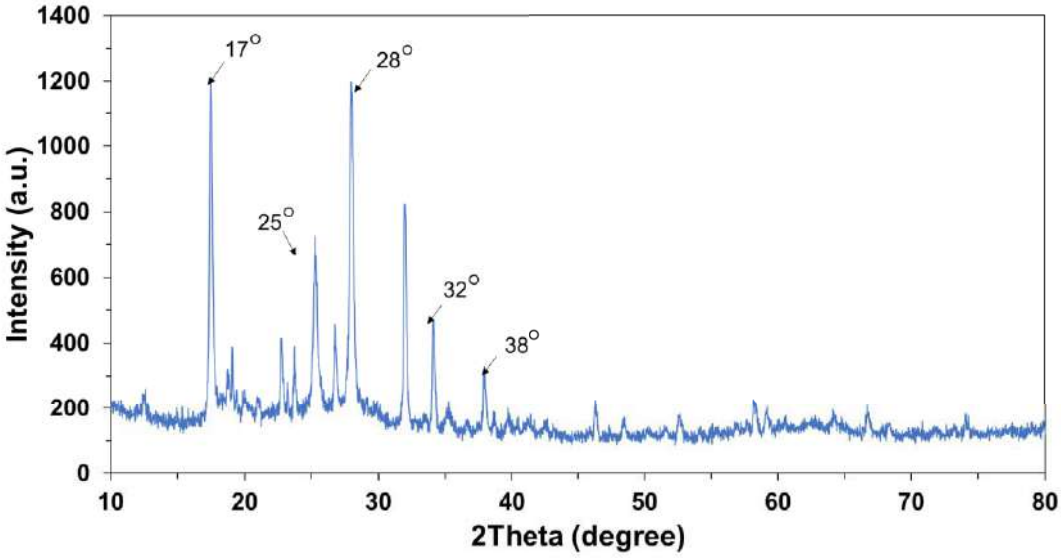


Figure 2. Diffractogram of the purified humic acid.

3.3 UV-Vis spectra of reduced graphene oxide.

UV-Vis spectra of products synthesized shown in Figure 3 and Table 2. It was found the bands at λ_{max} 242-246 nm which was found at power of 800 watts and heated for 12 min. This band would be suggested the band of reduced graphene oxide. The UV-Vis result was compared to UV-Vis spectra of Adere T. H. *et.al.* reported the synthesis of rGO by Hummer method [7]. It was indicated rGO formed at ~226 nm. The increased temperature affected the formation of the product [8].

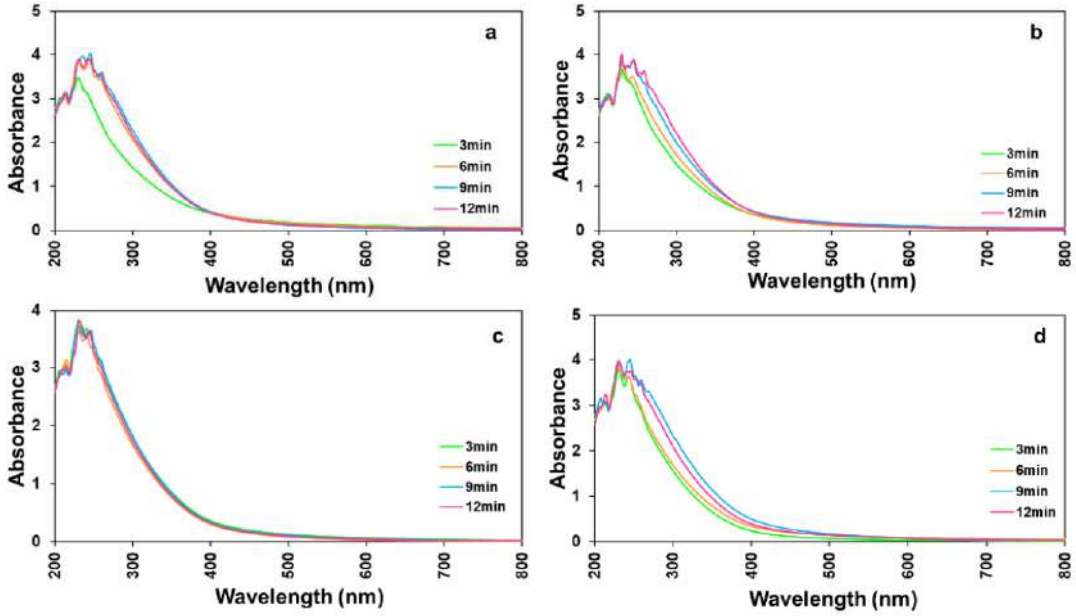


Figure 3. UV-Vis spectra of reduced graphene oxide for 3-12 min at (a) 300 (b) 500 (c) 700 and (d) 800 watt.

Table 2. The λ_{max} of reduced graphene oxide synthesized by hydrothermal with microwave assisted method.

Power (watt)	Time (min)	λ_{max}	Power (watt)	Time (min)	λ_{max}
300	3	242	700	3	244
	6	244		6	244
	9	246		9	246
	12	246		12	246
500	3	242	800	3	242
	6	246		6	242
	9	246		9	246
	12	246		12	246

3.4 FTIR spectra of reduced graphene oxide.

The characterization using FTIR spectroscopy was identified the functional groups and bonding vibrations present in the products during the synthesis process. The FTIR result of products synthesized for 9 and 12 min in Figure 4 and 5, respectively show the presence 1618 cm^{-1} assigned to C=C stretching, 1115 cm^{-1} assigned to C–O. The vibration at 994 and 619 cm^{-1} was assigned to C–OH bending and C=C bending, respectively. The widening of band was –OH stretching at 3436 cm^{-1}. The band at 449 cm^{-1} was assigned to Si–O of silica which was the composition of leonardite. Based on experimental results compared to FTIR spectrum of the rGO show the band at 1605 and 1619 cm^{-1} of Eka *et al.* [9] and Jiao *et al.* [10]

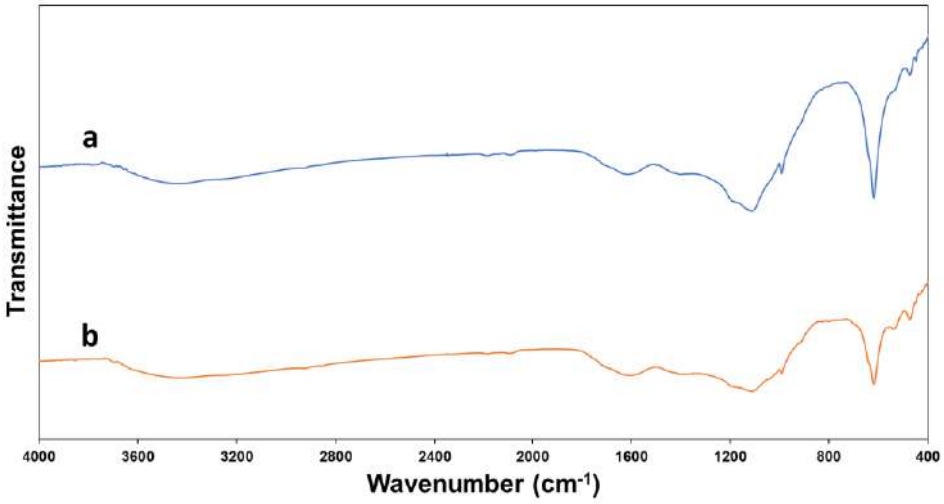


Figure 4. FTIR spectra of reduced graphene oxide for 9 min at (a) 700 and (b) 800 watt.

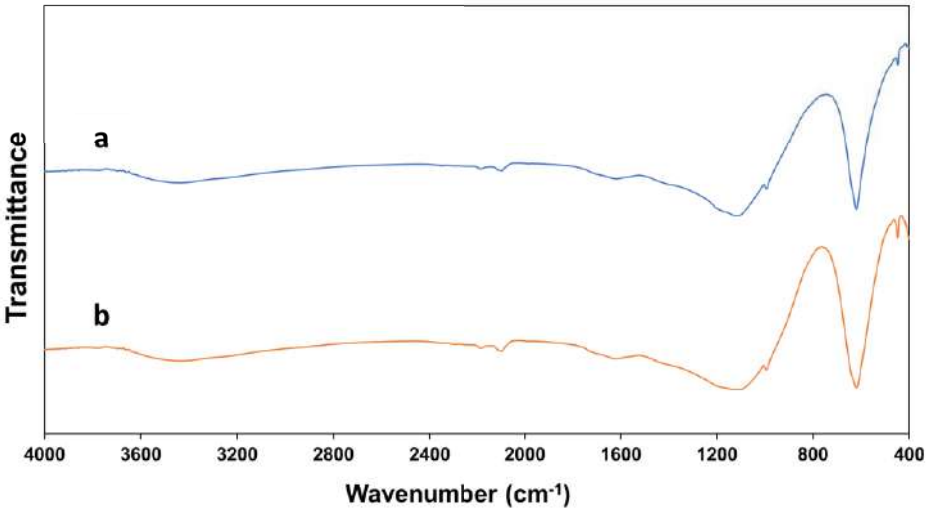


Figure 5. FTIR spectra of reduced graphene oxide for 12 min at (a) 700 and (b) 800 watt.

3.5 XRD results.

The products from condition heating at 700 and 800 watt for 9 min was shown in Figure 6. The peaks displayed 2θ at 23.1º, 32.3º and 45.9º. The products from condition heating at 700 and 800 watt for 12 min was shown in Figure 7. The peaks displayed 2θ at 23.6º, 31.8º and 45.8º. The 2θ of the products were difference when compared to the humic acid as precursor in Figure 2. The rGO from natural graphite by Hummer's method show 2θ = 24.5º and 43º [10]. It indicated the 2θ of products was not corresponded with rGO from natural graphite. It would be suggested that the product would form as rGO uncompleted.

Base on UV-Vis and FTIR results would be assigned the product could be rGO. While, XRD results could not confirm the formation of rGO.

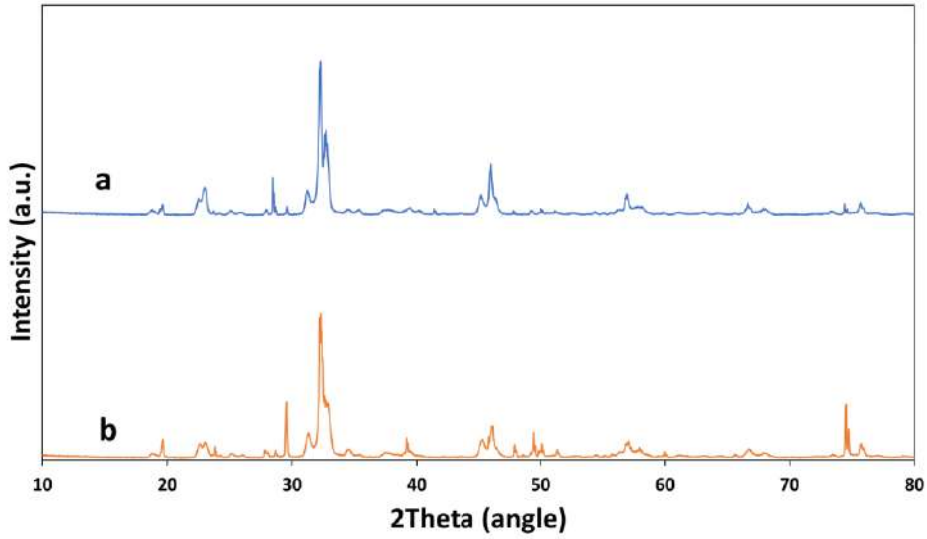


Figure 6. XRD pattern of product heating for 9 min at (a) 700 and (b) 800 watt.

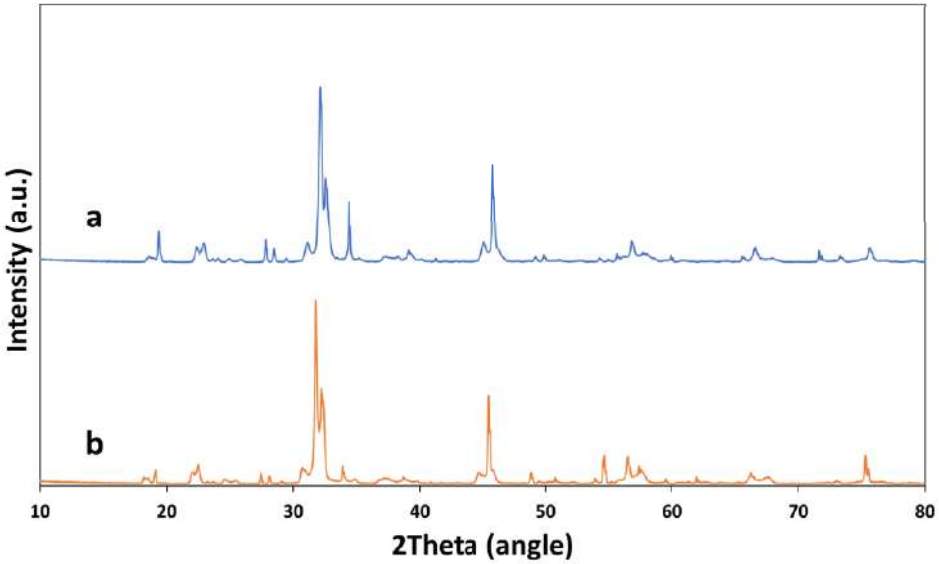


Figure 7. XRD pattern of of product heating for 12 min at (a) 700 and (b) 800 watt.

4. Conclusion

The humic acid dissolved in 2.0 M of sodium hydroxide, left at room temperature for 72 hr and precipitated by 2.0 M of sulfuric acid provided purified humic acid was 5.79 % of yield and more crystallinity. Base on UV-Vis and FTIR results would be assigned reduced graphene oxide could be synthesized by the hydrothermal with microwave method at 700 and 800 watt for 9 and 12 min. While, XRD results could not confirm the formation of rGO.

Acknowledgments

The authors express their sincere gratitude to Mae Moh mine, Lampang province for providing humic acid.

References

[1] Ratanaprommanee C., Chinachanta K., Chaiwan F. and Shutsrirung A. 2017 Chemical characterization of leonardite and its potential use as soil conditioner and plant growth enhancement. (*Asia-Pacific Journal of Science and Technology vol 22*) pp 1-10.

[2] Xing B., Yuan R., Zhang C., Huang G., Guo H., Chen Z., Chen L., Yi G., Zhang Y. and Yu J. 2017 Facile synthesis of graphene nanosheets from humic acid for supercapacitors (*Fuel Processing Technology vol 165*) pp. 112-122.

[3] Konch T.J., Gogoi R.K., Gogoi A., Saha K., Deka J., Reddy K.A. and Raidongia K. 2018 Nanofluidic transport through humic acid modified graphene oxide nanochannels (*Materials Chemistry Frontiers vol. 2*) pp. 1647-1654.

[4] Huang G., Kang W., Geng Q., Xing B., Liu Q., Jia J. and Zhang C. (2018) One-step green hydrothermal synthesis of few-layer graphene oxide from humic acid (*Nanomaterials vol 8*) pp. 2-9.

[5] Smith A. T., LaChance A. M., Zeng S., Liu B and Sun L. 2019 *Synthesis, properties, and applications of graphene oxide/reduced graphene oxide and their nanocomposites*. (*Nano Materials Science vol 1*) pp. 31-47.

[6] Amanda M.P., Alexandre C.B., Ana C.C., Raphael A.B., Jerusa S.G. and Marcello G.T. 2018 *New organomineral complex from humic substances extracted from poultry wastes: Synthesis, characterization and controlled release study* (*Journal Brazil chemistry vol 29*) pp 140-150.

2. [7] Adere T. H. and Delele W. A. 2019 *Synthesis and characterization of reduced graphene oxide (rGO) started from graphene oxide (GO) using the tour method with different parameters* (*Advances in Materials Science and Engineering*) pp. 1-9.

[8] Al-Hazmi F.S., Ghada H.A., Gary W.B., Al-Ghamdi A.A., Obaid A.Y and Waleed E.M. 2015 *One pot synthesis of graphene based on microwave assisted solvothermal technique* (*Synthetic Metals vol 200*) pp 54-57.

[9] Eko A., Shoerya S., Gatot S., and Sadur R. 2015 *Facile synthesis of graphene from graphite using ascorbic acid as reducing agent* (*The 3rd International Conference on Advanced Materials Science and Technology (ICAMST 2015)*) pp. 020003-1.

[10] Jiao X., Qiu Y., Zhang L. and Zhang X. 2017 Comparison of the characteristic properties of reduced graphene oxides synthesized from natural graphite with different graphitization degrees. (*The Royal Society of Chemistry vol 7*) pp 52337-52344.

ICSTI-MJU 2022

THE 2ND INTERNATIONAL CONFERENCE ON
SCIENCE TECHNOLOGY & INNOVATION-MAEJO UNIVERSITY