The Proceedings

2020 15th International Joint Symposium on
Artificial Intelligence and Natural Language Processing
(iSAI-NLP)

2020 International Conference on
Artificial Intelligence and Internet of Things
(AIoT)

Copyright Page

2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)

IEEE Part Number: CFP20S32-ART
ISBN: 978-0-7381-3068-2

About this Publication

Title:	Proceedings of 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing and 2020 International Conference on Artificial Intelligence and Internet of Things (iSAI-NLP-AIoT)
Editor-in-chief:	Thepchai Supnithi
Editors:	Thanaruk Theeramunkong, Chai Wutiwiwatchai, Tsuyoshi Isshiki, and Denchai Worasawate
Production Assistants:	Thodsaporn Chay-intr
Organizers:	Artificial Intelligence Association of Thailand (AIAT), Thailand Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU), Thailand National Electronics and Computer Technology Center (NECTEC), Thailand Tokyo Institute of Technology, Japan Kasetsart University (KU), Thailand
Publisher:	Artificial Intelligence Association of Thailand (AIAT), 160 Moo 5, Tiwanon Road, Bangkadi, Mueang Pathum Thani, Pathum Thani, 12000, Thailand Email: info@aiat.or.th URL: https://aiat.or.th
Date Published:	November 2020

Contents

Welcome Message from the iSAI-NLP-AIoT 2020 General Committee

From 2017, the Symposium on Natural Language Processing (SNLP2018) has been broadened to cover the topics of artificial Intelligence in addition to natural language processing and its name is changed to the Joint International Symposium on Artificial Intelligence and Natural Language Processing series (iSAI-NLP). The first meeting was held in 1993 by Chulalongkorn University in Bangkok, Thailand. The following conferences were held by Kasetsart University in Bangkok (1995), Asian Institute of Technology in Phuket (1997), King Mongkut's University of Technology Thonburi in Chiang Mai (2000), SIIT, Thammasat University in Hua Hin (2002), Chulalongkorn University in Chiang Rai (2005), Kasetsart University in Pattaya (2007), Dhurakij Pundit University in Bangkok (2009), King Mongkut's Institute of Technology Ladkrabang in Bangkok (2011), SIIT in Phuket (2013), Thammasat University in Ayutthaya (2016), King Mongkut's University of Technology Thonburi and Rangsit University in Hua Hin (2017), Mahidol University and AIAT in Pattaya (2018), and Muban Chombueng Rajabhat University and AIAT in Chiang Mai (2019).

The iSAI-NLP-AIoT 2020 is the joint event of the 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2020) and the International Conference on Artificial Intelligence and Internet of Things (AIoT 2020). It is hosted by National Electronics and Computer Technology Center (NECTEC), Sirindhorn International Institute of Technology (SIIT, Thammasat University) and Tokyo Institute of Technology (Tokyo Tech) with the great support from Artificial Intelligence Association of Thailand (AIAT). Due to the COVID-19 pandemic, the iSAI-NLP-AIoT 2020 is held online during November 18-20, 2020, together with the International Multi-Conference on Advances in Engineering, Technology, and Management (IMC-AETM 2020). The first International Conference on Artificial Intelligence and Internet of Things (AIoT 2020) is the continuous event of the International Conference on Information and Communication Technology for Embedded Systems (ICICTES). The previous series were ended with the number 10. Therefore, the first AIoT conference can be counted as the eleventh event of the ICICTES conferences. Being a joint conference under the theme of Artificial Intelligence of Things (AIoT), the aim of iSAI-NLP-AIoT 2020 is to promote artificial intelligence research in the five main topics; (1) natural language processing and knowledge engineering, (2) robotics/IoT/embedded system, (3) data analytic and machine learning (4) signal, image and speech processing and (5) smart industrial technologies

On behalf of the iSAI-NLP-AIoT 2020 committee, we would like to thank our keynote speakers, Siriwan Suebnukarn (Thammasat University, also co-speaker of IMC-AETM 2020), Chidchanok Lursinsap (Chulalongkorn University, Thailand), Hiroshi Sasaki (Tokyo Institute of Technology, also co-speaker of IMC-AETM 2020), for their insightful keynote and invited speeches. We would also like to express our appreciation to Thepchai Supnithi (NECTEC, Thailand), a program chair of iSAI-NLP-AIoT 2020) for their support. We also appreciate the co-hosts: Sirindhorn International Institute of Technology, Thammasat University (SIIT, TU, Thailand), Tokyo Institute of Technology (TiTech, Japan), Artificial Intelligence Association of Thailand (AIAT, Thailand), and Thailand National Electronics and Computer Technology Center (NECTEC, Thailand). We also appreciate the presenters, the reviewers, the organizing team for their contribution.

We look forward to seeing you soon at iSAI-NLP-AIoT 2020 online during November 18-20, 2020.

iSAI-NLP2020 General Co-Chairs
Thanaruk Theeramunkong *(SIIT, Thammasat University, Thailand and AI Association of Thailand)*
Chai Wutiwiwatchai *(NECTEC, Thailand)*
Denchai Worasawate (*Kasetsart University, Thailand)*
Tsuyoshi Isshiki *(Tokyo Institute of Technology, Japan)*

Welcome Message from the iSAI-NLP-AIoT 2020 Program Chairs

The International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), formerly named the International Joint Symposium on Natural Language Processing (SNLP), is one of the oldest series of international conferences dedicated to natural language processing and artificial intelligence in general originating from Thailand since 1993. This year, the iSAI-NLP is the fifteenth of the series and it will be held online during November 18-20, 2020. We are pleased to have iSAI-NLP 2020 co-located with AIoT 2020. The iSAI-NLP 2020, whose theme is Artificial Intelligence of Things (AIoT), consists of five tracks: natural language processing; robotics, IoT and Embedded System; data analytic and machine learning, signal image and speech processing and smart industrial technology. The first International Conference on Artificial Intelligence and Internet of Things (AIoT 2020) is the continuous event of the International Conference on Information and Communication Technology for Embedded Systems (ICICTES). It focuses on AI and embedded systems.

Because of the COVID-19 pandemic, it is the first time we organize a virtual conference. Thanks for interests from many parts of the world, iSAI-NLP-AIoT 2020 received 75 submissions from Bangladesh, India, Pakistan, Japan, Myanmar, and Thailand. All the papers are rigorously reviewed by at least three anonymous reviewers, and finally we accepted 34 regular papers and 6 short papers. For this review process, we appreciate the reviewers who spent their valuable time to review and evaluate the submitted papers. We also have the 2nd Joint Myanmar-Thai NLP/AI R&D Workshop which aims to promote R&D collaboration among Myanmar and Thailand and will extend to regional collaboration in the near future.

We would like to show our sincere gratitude to our keynote speakers, we would like to thank our keynote speakers, Siriwan Suebnukarn (Thammasat University, also co-speaker of IMC-AETM 2020), Chidchanok Lursinsap (Chulalongkorn University, Thailand), Hiroshi Sasaki (Tokyo Institute of Technology, also co-speaker of IMC-AETM 2020), for their excellent keynote and invited speeches. We also appreciate the co-hosts: Sirindhorn International Institute of Technology, Thammasat University (SIIT, TU, Thailand), Tokyo Institute of Technology (TiTech, Japan), Artificial Intelligence Association of Thailand (AIAT, Thailand), and Thailand National Electronics and Computer Technology Center (NECTEC, Thailand). We also appreciate the presenters, the reviewers, and the organizing team for their contribution. Finally, we express our biggest gratitude to all the authors who submitted their valuable papers, who were accepted and made their presentations, and who joined active discussion at the sessions.

The iSAI-NLP-AIoT 2020 will be held online during November 18-20, 2020 and see you soon.

Thank you for all your effort, cooperation, and kindness.

iSAI-NLP2020 Program Co-Chairs
Thepchai Supnithi *(NECTEC, Thailand)*

iSAI-NLP-AIoT 2020 Committee

Honorary Co-Chair

Vilas Wuwongse
Sakon Nakhon Rajabhat University, Thailand

Chongrak Wachrinrat
President, Kasetsart University, Thailand

Gasinee Witoonchart
Thammasat University, Thailand

Chadamas Thuvasethakul
NSTDA, Thailand

Thanya Kiatiwat
Kasetsart University, Thailand

Pruettha Nanakorn
SIIT, Thammasat University, Thailand

Kiyota Hashimoto
Prince of Songkla University, Thailand

Conference Co-chair

Thanaruk Theeramunkong
SIIT, Thammasat University, Thailand

Chai Wutiwiwatchai
NECTEC, Thailand

Tsuyoshi Isshiki
Tokyo Institute of Technology, Japan

Denchai Worasawate
Kasetsart University, Thailand

Technical Program Chairs

Thepchai Supnithi
NECTEC, Thailand

Technical Program Committee

Track 1: Natural Language Processing

Prachya Boonkwan
NECTEC, Thailand

Aw Ai Ti
Institute for Infocomm Research, Singapore

Rachada Kongkachandra
Thammasat University, Thailand

Ye Kyaw Thu
NECTEC, Thailand

Track 2: Data Analytics and Machine Learning

Worawut Yimyam
Phetchaburi Rajabhat University, Thailand

Natsuda Kaothanthong
SIIT, Thammasat University, Thailand

Olarik Surinta
Mahasarakham University, Thailand

Track 3: Signal, Image and Speech Processing

Narit Hnoohom
Mahidol University, Thailand

Anuchit Jitpattanakul
King Mongkut 's University of Technology North Bangkok, Thailand

Ngoc Hong Tran
University College Dublin, Ireland

Sakorn Mekruksavanich
University of Phayao, Thailand

Thach-Thao Nguyen Duong
University of Burgundy, France

Track 4: Robotics IoT and Embedded System

Mahasak Ketcham
King Mongkut's University of Technology North Bangkok, Thailand

Michael Pecht
University of Maryland, USA

Thaweesak Yingthawornsuk
King Mongkut's University of Technology Thonburi, Thailand

Track 5: Smart Industrial Technologies

Narumol Chumuang
Muban Chombueng Rajabhat University, Thailand.

Adil Farooq
University of Cyprus, Cyprus.

Supakrit Sukjarern
Muban Chombueng Rajabhat University, Thailand.

AIoT

Thanaruk Teeramunkong
SIIT, Thammasat University, Thailand

Denchai Worasawate
Kasetsart University, Thailand

,
,

iSAI-NLP General Track Reviews

Anchana Muankid
King Mongkut's University of Technology North Bangkok, Thailand

Anocha Rugchatjaroen
NECTEC, Thailand

Aongart Aun-a-nan
Rajamangala University of Technology Suvarnabhumi, Thailand

Apichart Intarapanich
NECTEC, Thailand

Chanon Promsakon
King Mongkut's University of Technology North Bangkok, Thailand

Chenchen Ding
National Institute of Information and Communications Technology, Japan

Chonnikarn Rodmorn
King Mongkut's University of Technology North Bangkok, Thailand

Chutima Beokhaimook
Rangsit University, Thailand

Hnin Aye Thant
University of Technology, Myanmar

Jaruwat Pailai
Kasetsart University, Thailand

Kiyota Hashimoto
Prince of Songkla University, Thailand

Konlakorn Wongpatikaseree
Mahidol University, Thailand

Kreangsak Tamee
Naresuan University, Thailand

Krit Kosawat
NECTEC, Thailand

Masao Utiyama
National Institute of Information and Communications Technology, Japan

Mathuros Panmuang
Thonburi University, Thailand

International Advisory Committee

Publication Co-Chair

Thodsaporn Chay-intr
Tokyo Institute of Technology, Japan

Toshiaki Kondo
SIIT, Thammasat University, Thailand

Waree Kongprawechnon
SIIT, Thammasat University, Thailand

Secretary General

Choermath Hongakkaraphan
AIAT, Thailand

Sasiporn Usanavasin
SIIT, Thammasat University, Thailand

Financial Co-Chair

Chutima Beokhaimook
Rangsit University, Thailand

Webmaster

Suchathit Boonnag
AIAT, Thailand

Thanasan Tanhermhong
AIAT, Thailand

Wirat Chinnan
AIAT, Thailand

Co-hosts

Artificial Intelligence Association of Thailand (AIAT)
Thailand

Sirindhorn International Institute of Technology (SIIT),
Thammasat University (TU)
Thailand

National Electronics and Computer Technology Center (NECTEC)
Thailand

Tokyo Institute of Technology
Japan

Kasetsart University (KU)
Thailand

Supporters

IEEE Thailand Section
Thailand

Keynote Speakers

Keynote Speakers

Siriwan Suebnukarn

Vice Rector for Research and Innovation,
Thammasat University, Thailand

Intelligent Clinical Training during the COVID-19 Pandemic

Abstract Clinical training is one of the most challenging areas for education especially during the COVID-19 pandemic. There are limited access to apprenticeship training in the complex scenarios with corresponding difficulty training in a time-effective manner. Professor Suebnukarn's work on intelligent clinical training systems provides one effective solution to this problem by introducing intelligent clinical training systems that can supplement tutoring sessions by expert clinical instructors. The Bayesian representation techniques and algorithms for generating tutoring feedback in medical problem-based learning group problem solving made important contributions to the field of Intelligent Tutoring Systems. In particular, it was one of the first systems for tutoring groups of students and the first intelligent tutoring systems for medical problem-based learning. The virtual reality simulator she developed is one of the most sophisticated dental simulators. It stands out as the first dental simulator to integrate sophisticated analysis of the surgical procedure. Particularly noteworthy is also the creative way to understand important issues such as differences in expert and novice performance, the effectiveness of virtual pre-operative practice, and the teaching effectiveness of the simulator. The systems have been implemented in undergrad pre-clinical training and postgrad pre-surgical training with strong scientific evidence of their effectiveness.

Bibliography Professor Siriwan Suebnukarn serves as Vice Rector for Research and Innovation at Thammasat University, Thailand. Professor Suebnukarn's combined background in Dentistry and Computer Science gives her a rather unique set of skills to tackle some important outstanding problems in Medical Informatics and Education. Her research work has included Artificial Intelligence in Education, Intelligent User Interfaces, and User Modeling. She developed an Intelligent Virtual Reality Clinical Training Simulator for which she won the prestigious International Federation of Inventor Association's (IFIA) Lady Prize for the Best Women's Invention and the National Outstanding Researcher Award in Education.

Keynote Speakers

Dr. James O. B. Rotimi
Associate Professor James O. B. Rotimi, PhD
Academic Dean (Construction), School of Built Environment
College of Sciences, Massey University, New Zealand

A healthy construction sector: A panacea for growth

Abstract The relationship that exists between the construction sector and national economic growth is well known. This relationship is generally referred to as being 'causal', so one affects the other and vice versa. With the aspirations for growth and national developments (especially post Covid-19) there needs to be a conscious bias for a key sectoral stimulant to be healthy. In this keynote presentation, Dr Rotimi will present arguments for both a healthy construction sector and also the health and wellbeing of sector participants. Sector statistics are frightening, evidencing the need for focus on this sector challenges to help future-proof the construction industry.

Bibliography Dr. James O. B. Rotimi is Associate Professor of Construction Project Management and the Academic Dean for Construction at Massey University. James has qualifications in Building, Construction Management, Civil Engineering, Commerce and Education. He has over 27 years tertiary teaching and research experience in academic institutions in Nigeria, UK, South Africa and New Zealand. He also has various building construction industry experiences including a senior associate role in a quantity surveying consultancy in Nigeria. James publishes extensively within peer-reviewed Journals and conference proceedings. He is the Founding Editor of the International Journal of Construction Supply Chain Management (IJCSCM) which was established in 2011. James research interests are in: construction project management, construction supply chain management, disaster legislation and post disaster reconstruction and management.

Keynote Speakers

Hiroshi Sasaki

Associate Professor Department of Information and Communications Engineering,
Tokyo Institute of Technology, Japan

Achieving Practical Byte-Granular Memory Safety

Abstract Memory safety issues have been a serious threat which have provided a significant opportunity for exploitation by attackers. I would like to share my experience in building a hardware-based fine-grained memory safety solution, based on a simple idea that prohibits a program from accessing certain memory regions based on program semantics.

Bibliography Hiroshi Sasaki is an Associate Professor of Information and Communications Engineering at the Tokyo Institute of Technology, Japan. His research interests include computer architecture, computer systems, and computer security.

Conference Overall Schedule (Day 1)

<table>
<tr><th></th><th colspan="4">Wednesday, November 18, 2020 (Day 1)</th></tr>
<tr><td>11:30-13:30</td><td colspan="4">Registration to the ZOOM</td></tr>
<tr><td></td><td colspan="4">Room A</td></tr>
<tr><td>13:15-13:30</td><td colspan="4">Opening Ceremony
General Chair representatives
IMC-AETM 2020: Pruetha Nanakorn, SIIT, Thammasat University, Thailand
iSAI-NLP & AIoT 2020: Tsuyoshi Isshiki, Tokyo Institute of Technology, Japan
Program Chair Representatives
IMC-AETM 2020: Thepchai Supnithi, NECTEC, Thailand
iSAI-NLP & AIoT 2020: Erwin Oh, Griffith University, Australia</td></tr>
<tr><td>13:30-14:30</td><td colspan="4">Keynote Speech I (iSAI-NLP-AIoT)
Intelligent Clinical Training during the COVID-19 Pandemic
Siriwan Suebnukarn
Thammasat University, Thailand</td></tr>
<tr><td>14:30-14:50</td><td colspan="4">Break</td></tr>
<tr><td></td><td>Room A</td><td>Room B</td><td>Room C</td><td>Room D</td></tr>
<tr><td>14:50-17:00</td><td>iSAI-Workshop
(Thepchai Supnithi)
(15.30-18.20)</td><td>IMC-01: CE
(Warangkana Saengsoy)</td><td>IMC-02: ICT
(Nguyen Duy Hung)</td><td>IMC-03: CHE
(Wanwipa Siriwatwechakul)</td></tr>
</table>

Conference Overall Schedule (Day 2)

	Thursday, November 19, 2020 (Day 2)			
08:00-09:00	**Registration**			
	Room A			
09:00-10:00	**Keynote Speech II (IMC-ATEM)** ***A healthy construction sector: A panacea for growth*** James Rotimi *Massey University, New Zealand*			
10:00-10:30	**Break**			
	Room A	**Room B**	**Room C**	**Room D**
10:15-12:00	iSAI-01: NLP *(Ponrudee Netisopakul)*	iSAI-02: Signal *(Narit Nhoohom)*	iSAI-03: DA & ML *(Konlakorn Wongpatikaseree)*	IMC-04: CE *(Ganchai Tanapornraweekit)*
12:00-13:30	**Lunch Time**			
13:30-14:30	**Keynote Speech III (iSAI-NLP-AIoT)** ***Achieving Practical Byte-Granular Memory Safety*** Hiroshi Sasaki *Tokyo Institute of Technology, Japan*			
14:30-14:50	**Break**			
	Room A	**Room B**	**Room C**	**Room D**
14:50-17:00	iSAI-04: NLP *(Rachada Kongkachandra)*	iSAI-05: Robot *(Patiyuth Pramkeaw and Denchai Worasawate)*	iSAI-06: Smart *(Thaweesak Yingthawornsuk)*	IMC-05: MT *(Warut Panna-kkong, Jirachai Buddhakulsomsiri, Pisal Yenradee)*

CE: Civil Engineering CHE: Chemistry and Environment DA: Data Analytics
ICT: Information & Communication ML: Machine Learning MT: Management Technology
NLP: Natural Language Processing Robot: Robotics, IoT and Embedded System Smart: Smart Industrial Technologies

Conference Overall Schedule (Day 3)

	Friday, November 20, 2020 (Day 3)			
08:00-09:00	**Registration**			
	Room A	**Room B**	**Room C**	**Room D**
09:00-10:00	iSAI-07: DA &ML *(Sanparith Marukatat)*	iSAI-08: Signal *(Nongnuch Ketui or Narumol Chumuang)*	IMC-06: CHE *(Pakorn Opaprakasit)*	IMC-07: MT *(Akaranan Pong-sathornwiwat)*

Detailed Schedule

DAY 1 AM WEDNESDAY, NOVEMBER 18

Main Session [Day 1 - AM/PM]

[11:30-13:30]
Registration to the Zoom
[13:15-13:30]
Opening Ceremony at Room A
General chair representatives:
IMC-AETM 2020: Pruetha Nanakorn, *SIIT, Thammasat University, Thailand*
iSAI-NLP & AIoT 2020: Tsuyoshi Isshiki, *Tokyo Institute of Technology, Japan*
Program Chair Representatives:
IMC-AETM 2020: Thepchai Supnithi, *NECTEC, Thailand*
iSAI-NLP & AIoT 2020: Erwin Oh, *Griffith University, Australia*
[13:30-14:30]
[iSAI-NLP-AIoT2020] Keynote Speech I at Room A:
Intelligent Clinical Training during the COVID-19 Pandemic
Siriwan Suebnukarn, *Thammasat University, Thailand*
[14:30-14:50]
Break

DAY 1 - PM (1) WEDNESDAY, NOVEMBER 18

iSAI-Workshop [Day 1 - PM] Room A

Session Chair: Thepchai Supnithi

[15:30-15:35]
Welcome Address
Thepchai Supnithi, *NECTEC, Thailand*

[15:35-16:15]
Keynote Speech
Khmer NLP at NIPTICT
Sam Sethserey,
Vice President, National Institute of Posts, Telecoms and ICT (NIPTICT), Cambodia

[16:15-16:35]
An Approach of Network Analysis Enhancing Knowledge Extraction in Thai Newspapers Contexts (W1-01)
Akkharawoot Takhom, Dhanon Leenoi, Chotanunsub Sophaken, Prachya Boonkwan, and Thepchai Supnithi

[16:35-16:55]
Improve Accuracy of Word Suggestion by Location of Word Search: A case study of Regional Thai Dialects (W1-02)
PNattapol Kritsuthikul, Witchaworn Mankhong, Wasan Na-Chai, and Thepchai Supnithi

[16:55-17:15]
Myanmar Text (Burmese) and Braille (Mu Thit) Machine Translation applying IBM Model 1 and 2 (W1-03)
Zun Hlaing Moe, Thida San, Ei Thandar Phyu, Hlaing Myat Nwe, Hnin Aye Thant, Thepchai Supnithi, and Ye Kyaw Thu

[17:15-17:35]
Grapheme-to-IPA Phoneme Conversion for Burmese (myG2P Version 2.0) (W1-04)
Honey Htun, Ni Htwe Aung, Shwe Sin Moe, Wint Theingi, Nyein Nyein Oo, Thepchai Supnithi, and Ye Kyaw Thu

[17:35-17:55]
Grapheme to Syllable Sequence Phoneme Conversion for Myanmar Language Spelling TTS (W1-05)
Hnin Yu Hlaing, Ye Kyaw Thu, Hlaing Myat Nwe, Thepchai Supnithi, and Hnin Aye Thant

[17:55-18:15]
Improve Neural Machine Translation (NMT) with Conjoined Twin Model (W1-06)
Nattapol Kritsuthikul, Peerachet Porkaew, and Thepchai Supnithi

[18:15-18:20]
Closing Remarks
Kyaw Thu, *NECTEC, Thailand*

DAY 1 - PM (2) WEDNESDAY, NOVEMBER 18

IMC-01: Civil Engineering [Day 1 - PM] Room B

Session Chair: Warangkana Saengsoy

[14:50-15:05]
Cloud-based Virtual Reality Application for Real Estate Industry (IMC-AETM2020-0004)
Daluch Sinoeurn and Kriengsak Panuwatwanich

[15:05-15:20]
Free Shrinkage and Restrained Shrinkage Behaviors of Multi-binder Systems Containing Cement, Limestone Powder, Fly Ash and Calcined Clay (IMC-AETM2020-0008)
Phung Manh Cuong, Krittiya Kaewmanee, and Somnuk Tangtermsirikul

[15:20-15:35]
Use of Bottom Ash as a Subbase Material for Pavement (IMC-AETM2020-0014)
Thet Htet Ye Htun, Parnthep Julnipitawong, Weeraya Chimoye, and Somnuk Tangtermsirikul

[15:35-15:50]
Utilization of Substandard Fly Ash as a Partial Replacement for Fine Aggregate in Concrete (IMC-AETM2020-0015)
Piseth Pok and Parnthep Julnipitawong

[15:50-16:05]
Bond Strength of Concrete Repair Materials with Considering Failure Plane Area by Image Analysis (IMC-AETM2020-0018)
Pakawat Sancharoen and Thitiporn Lertrusdachakul

DAY 1 - PM (3) WEDNESDAY, NOVEMBER 18

IMC-02: ICT [Day 1 - PM] Room C

Session Chair: Nguyen Duy Hung

[14:50-15:05]
Smart-Plug Implementation and Feature Comparisons for Electrical Appliance Recognition (IMC-AETM2020-0011)
Puwaphat Jitket, Seksan Laitrakun, and Sasiporn Usanavasin

[15:05-15:20]
Optimal Discount Rate in Multiagent Deep Reinforcement Learning for Traffic Signal Control By the Simulation of Urban Mobility (SUMO) (IMC-AETM2020-0033)
Paschol Supradith Na Ayudhyam and Mongkut Piantanakulchai

[15:20-15:35]
Deep Learning Approach on Photovoltaic Electricity Generation Capacity Prediction (IMC-AETM2020-0038)
Noravee Sungpuag, Sasiporn Usanavasin, and Waree Kongprawechnon

[15:35-15:50]
Marker and IMU-based registration for mobile augmented reality (IMC-AETM2020-0041)
Pansavuth Khehasukcharoen and Teerayut Horanont

[15:50-16:05]
Marker and IMU-based registration for mobile augmented reality (IMC-AETM2020-0041)
Pansavuth Khehasukcharoen and Teerayut Horanont

[16:05-16:20]
A Smoothing Photovoltaic Power Injection Using Double Moving Average with Efficient Battery Energy Storage System (IMC-AETM2020-0047)
Possawee Boonkerd, Waree Kongprawechnon, and Muhammad Jamil

DAY 1 - PM (4) WEDNESDAY, NOVEMBER 18

IMC-03: Chemistry and Environment [Day 1 - PM] Room D

Session Chair: Wanwipa Siriwatwechakul

[14:50-15:05]
Micro-plastics contamination in drinking water treatment plants: case study of Phnom Penh (IMC-AETM2020-0005)
Dork Hakk and Sandhya Babel

[15:05-15:20]
Improve Workability of CSA Cement Additive by Polymer Encapsulation (IMC-AETM2020-0007)
Malinee Nontikansak, Phattarakamon Chaiyapoom, Wanwipa Siriwatwechakul, Passarin Jongvisuttisan, and Chalermwut Snguanyat

[15:20-15:35]
On improving Performance of Artificial Neural Network in Predicting Dissolved Oxygen (DO) by Utilizing Feature Selection Algorithms to Generalize Water Quality Variables (IMC-AETM2020-0010)
Thananya Rinsiri and Akkaranan Pongsathornwiwat

[15:35-15:50]
A Specific Heat Prediction Model for Ternary Binder Pastes (IMC-AETM2020-0025)
Kanin Pinitoppapun, Krittiya Kaewmanee, and Somnuk Tangtermsirikul

[15:50-16:05]
In Situ spectroscopic ellipsometry for the study of ultrathin film poly (N-isopropylacrylamide-co-acrylamide) in aqueous environment (IMC-AETM2020-0040)
Phongphot Sakulaue, Tossaporn Lertvanithphol, and Wanwipa Siriwatwechakul

DAY 2 - AM (1) THURSDAY, NOVEMBER 19

Main Session [Day 2 - AM]

[08:00-09:00]
Registration to the Zoom
[09:00-10:00]
[IMC-ATEM2020] Keynote Speech II at Room A:
A healthy construction sector: A panacea for growth
James Rotimi, *Massey University, New Zealand*
[10:00-10:30]
Break

DAY 2 - AM (2) THURSDAY, NOVEMBER 19

iSAI-01: Natural Language Processing [Day 2 - AM] Room A

Session Chair: Ponrudee Netisopakul

[10:30-10:45]
Survey of Query correction for Thai business-oriented information retrieval (iSAI-NLP-AIoT2020-0143)
Anuruth Lertpiya, Tawunrat Chalothorn, Phongsathorn Kittiworapanya, and Nuttapong Saelek

[10:45-11:00]
A Conversational Agent for Database Query: A Use Case for Thai People Map and Analytics Platform (iSAI-NLP-AIoT2020-0149)
Thikamporn Simud, Somchoke Ruengittinun, Navaporn Surasvadi, Nuttapong Sanglerdsinlapachai, and Anon Plangprasopchok

[11:00-11:15]
Can Tweets predict ICO success? Sentiment Analysis for Success of ICO Whitepaper: evidence from Australia and Singapore Markets (iSAI-NLP-AIoT2020-0151)
Anchaya Chursook, Nathee Naktnasukanjn, Somsak Chaimaim, Piyachat Udomwong, Jutharut Chatsirikul, and Nopasit Chakpitak

[11:15-11:30]
Behavioral Analysis of Transformer Models on Complex Grammatical Structures (iSAI-NLP-AIoT2020-0152)
Kanyanut Kriengket, Kanchana Saengthongpattana, Peerachet Porkaew, Vorapon Luantangsrisuk, Prachya Boonkwan, and Thepchai Supnithi

DAY 2 - AM (3) THURSDAY, NOVEMBER 19

iSAI-02: Signal, Image and Speech Processing [Day 2 - AM] Room B

Session Chair: Narit Nhoohom

[10:30-10:45]
Improving of Pre-Processing Image for Child Cleft Lip by Image Processing (iSAI-NLP-AIoT2020-0113)
Jutturong Charoenrit, Wararat Songpan, and Apisak Pattanachak

[10:45-11:00]
Comparison of Face Classification with Single and Multi-model base on CNN (iSAI-NLP-AIoT2020-0121)
Sarin Watcharabutsarakham, Supphachoke Suntiwichaya, Chanchai Junlouchai, and Apichon Kitvimorat

[11:00-11:15]
Direction of Arrival Identification Using MUSIC Method and NLMS Beamforming (iSAI-NLP-AIoT2020-0129)
Raungrong Suleesathira

[11:15-11:30]
Face Detection System for Public Transport Service Based on Scale-Invariant Feature Transform (iSAI-NLP-AIoT2020-0132)
Narumol Chumuang, Sansanee Hiranchan, Mahasak Ketcham, Worawut Yimyam, Patiyuth Pramkeaw, and Tanapon Jensuttiwetchakult

[11:30-11:45]
Image Classification of Forage Plants in Fabaceae FamilyUsing Scale Invariant Feature Transform Method (iSAI-NLP-AIoT2020-0134)
Worawut Yimyam, Thidarat Pinthong, Narumol Chumuang, Mahasak Ketcham, Patiyuth Pramkeaw, and Sansanee Hiranchan

[11:45-12:00]
A Development Heat Stroke Detection System Integrated with Infrared Camera (iSAI-NLP-AIoT2020-0145)
Worawut Yimyam, Nattavee Utakrit, Mahasak Ketcham, Montean Rattanasiriwongwut, Narumol Chumuang, and Sansanee Hiranchan

[12:00-12:15]
Estimation of Oil Content in Oil Palm Fresh Fruit Bunch by Its Surface Color (iSAI-NLP-AIoT2020-0140)
Sutat Sae-Tang

DAY 2 - AM (4) THURSDAY, NOVEMBER 19

iSAI-03: Data Analytics and Machine Learning [Day 2 - AM] Room C

Session Chair: Konlakorn Wongpatikaseree

[10:30-10:45]
Model proposed to cost reduction in printing [Short] (iSAI-NLP-AIoT2020-0126)
Noppon Mingmuang, Chaleedol Inyasri, and Wirote Jongchanachavawat

[10:45-11:00]
COVID-19: Data Analysis and The situation Prediction Using Machine Learning Based on Bangladesh perspective (iSAI-NLP-AIoT2020-0127)
Abir Abdullha and Sheikh Abujar

[11:00-11:15]
Enhancing and Evaluating an Impact of OCR and Ontology on Financial Document Checking Process (iSAI-NLP-AIoT2020-0137)
Worawut Yimyam, Sansanee Hiranchan, Mahasak Ketcham, Patiyuth Pramkeaw, Tanapon Jensuttiwetchakult, and Narumol Chumuang

[11:15-11:30]
Feature Extraction with SHAP Value Analysis for Student Performance Evaluation in Remote Collaboration (iSAI-NLP-AIoT2020-0170)
Mako Komatsu, Chihiro Takada, Chihiro Neshi, Teruhiko Unoki, and Mikifumi Shikida

DAY 2 - AM (5) THURSDAY, NOVEMBER 19

IMC-04: Civil Engineering [Day 2 - AM] Room D

Session Chair: Ganchai Tanapornraweekit

[10:30-10:45]
Calculation models of fire resistance rate for unprotected hot-rolled steel H-beams and their comparison (IMC-AETM2020-0016)
Laxman khanal and Taweep chaisomphob

[10:45-11:00]
Influence of Fly ash on viscosity and compressive strength of High Strength Concrete (IMC-AETM2020-0024)
Pham Hoai An, Parnthep Julnipitawong, and Somnuk Tangtermsirikul

[11:00-11:15]
Performance of Cracked AFRC and SFRC Beams Exposed to Chloride Environment (IMC-AETM2020-0026)
Suphawit Untimanon, Ganchai Tanapornraweekit, and Somnuk Tangtermsirikul

[11:15-11:30]
Topology Optimization of Planar Trusses with No Crossing and Overlapping Elements by A Firefly Algorithm (IMC-AETM2020-0028)
Sujan Tripathi and Pruettha Nanakorn

[11:30-11:45]
Mechanism of Lattice Frame Reinforcement (LFR) System Used to Delay Settlement of Coastal Rock Embankments (IMC-AETM2020-0031)
Isuru Tharanga Dias, Ganchai Tanapornraweekit, and Somnuk Tangtermsirikul

[11:45-12:00]
Effect of restraining and curing conditions on expansion and shrinkage of expansive concrete (IMC-AETM2020-0032)
Rasla Dumaru, Warangkana Saengsoy, and Somnuk Tangtermsirikul

DAY 2 - PM (1) THURSDAY, NOVEMBER 19

Main Session [Day 2 - PM]

[13:30-14:30]
[iSAI-NLP-AIoT2020] Keynote Speech III at Room A:
Achieving Practical Byte-Granular Memory Safety
Hiroshi Sasaki, *Tokyo Institute of Technology, Japan*
[14:30-14:50]
Break

DAY 2 - PM (2) WEDNESDAY, NOVEMBER 18

iSAI-04: Natural Language Processing [Day 2 - PM] Room A

Session Chair: Rachada Kongkachandra

[14:50-15:05]
A Structured Transformer Neural Machine Translation on Abstractive Text Summarization for Bangla seq2seq Learning [Short] (iSAI-NLP-AIoT2020-0130)
Abu Kaisar Mohammad Masum, Sharun Akter Khushbu, Mumenunnessa Keya, Sheikh Abujar, and Syed Akhter Hossain

[15:05-15:20]
Myanmar POS Resource Extension Effects on Automatic Tagging Methods (iSAI-NLP-AIoT2020-0157)
Zar Zar Hlaing, Ye Kyaw Thu, Myat Myo Nwe Wai, Thepchai Supnithi, and Ponrudee Netisopakul

[15:20-15:35]
A Study of Three Statistical Machine Translation Methods for Myanmar (Burmese) and Shan (Tai Long) Language Pair (iSAI-NLP-AIoT2020-0162)
Nang Aeindray Kyaw, Ye Thu, Hlaing Myat Nwe, Phyu Phyu Tar, Nandar Win Min, and Thepchai Supnithi

[15:35-15:50]
Daily Health Monitoring Chatbot with Linear Regression (iSAI-NLP-AIoT2020-0163)
Konlakorn Wongpatikaseree, Arunee Ratikan, Chaianun Damrongrat, and Katiyaporn Noibanngong

[15:50-16:05]
A Machine Learning Approach for the Classification of Methamphetamine Dealers on Twitter in Thailand (iSAI-NLP-AIoT2020-0165)
Punnavich Khowrurk and Rachada Kongkachandra

[16:05-16:20]
Statistical Machine Translation for Myanmar Language Paraphrase Generation (iSAI-NLP-AIoT2020-0171)
Myint Myint Htay, Ye Kyaw Thu, Hnin Aye Thant, and Thepchai Supnithi

DAY 2 - PM (3) WEDNESDAY, NOVEMBER 18

iSAI-05: Robotics, IoT and Embedded System [Day 2 - PM] Room B

Session Chair: Patiyuth Pramkeaw and Denchai Worasawate

[14:50-15:05]
A Surveillance system for children stuck inside the car with embedded system technology (iSAI-NLP-AIoT2020-0142)
Arthi Yooyen and Siriruang Phatchuay

[15:05-15:20]
Real-life Human Activity Recognition with Tri-axial Accelerometer Data from Smartphone using Hybrid Long Short-Term Memory Networks (iSAI-NLP-AIoT2020-0150)
Narit Hnoohom, Anuchit Jitpattanakul, and Sakorn Mekruksavanich

[15:20-15:35]
Simulation of Autonomous Mobile Robot System for Food Delivery in In-patient Ward with Unity (iSAI-NLP-AIoT2020-0155)
Supachai Vongbunyong, Salil Parth Tripathi, Kitti Thamrongaphichartkul, Nitisak Worrasittichai, Aphisit Takutruea, and Teeraya Prayongrak

[15:35-15:50]
A Pressure Sensor and Depth Camera-Based Monitoring and Alarming System for Bed Fall Detection [Short] (iSAI-NLP-AIoT2020-0161)
Punnavich Khowrurk and Rachada Kongkachandra

[15:50-16:05]
Development of a Web Service to Support the Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations [Short] (iSAI-NLP-AIoT2020-0168)
Chihiro Takada, Yurika Takeuchi, Mari Kinoshita, and Mikifumi Shikida

[16:05-16:20]
A proposal of evaluation method using a pressure sensor for supporting auscultation training (iSAI-NLP-AIoT2020-0174)
Yuki Kodera, Kunimasa Yagi, and Mikifumi Shikida

[16:20-16:35]
Design and Construct Logistic Robot Using Camera to Detect Line Combine with Lidar Sensor (Mecanum Wheel) (iSAI-NLP-AIoT2020-0166)
Tanaporn Anurakwongsri

DAY 2 - PM (4) WEDNESDAY, NOVEMBER 18

iSAI-06: Smart Industrial Technologies [Day 2 - PM] Room C

Session Chair: Thaweesak Yingthawornsuk

[14:50-15:05]
A Memetic Algorithm for Tour Trip Design Problem (iSAI-NLP-AIoT2020-0107)
Apisit Cheng and Aussadavut Dumrongsiri

[15:05-15:20]
Utilization-Weighted Algorithm for Spreading Factor Assignment in LoRaWAN (iSAI-NLP-AIoT2020-0144)
Kasama Kamonkusonman and Rardchawadee Silapunt

[15:20-15:35]
Development a home electrical equipment control device via voice commands for elderly assistance (iSAI-NLP-AIoT2020-0145)
Narumol Chumuang, Worawut Yimyam, Patiyuth Pramkeaw, Mahasak Ketcham, Sansanee Hiranchan, Sakchai Tangwannawit, Montean Rattanasiriwongwut

[15:35-15:50]
Medicine Identification System on Mobile Device for the Elderly (iSAI-NLP-AIoT2020-0148)
Pitchaya Chotivatunyu and Narit Hnoohom

[15:50-16:05]
Cryptocurrencies Asset Pricing Analysis: evidence from Thailand markets (iSAI-NLP-AIoT2020-0153)
Kanyawut Ariya, Nathee Naktnasukanjn, Tanarat Rattanadamrongaksorn, Piyachat Udomwong, Saronsad Sokantika, and Nopasit Chakpitak

[16:05-16:20]
An Efficiency Random Forest Algorithm for Classification of Patients with Kidney Dysfunction (iSAI-NLP-AIoT2020-0159)
Narumol Chumuang, Nuttawoot Meesang, Mahasak Ketcham, Worawut Yimyam, Jiragorn Chalermdit, Nawarat Wittayakhom, and Patiyuth Pramkeaw

[16:20-16:35]
A Framework of IoT Platform for Autonomous Mobile Robot in Hospital Logistics Applications (iSAI-NLP-AIoT2020-0173)
Kitti Thamrongaphichartkul, Nitisak Worrasittichai, Teeraya Prayongrak, and Supachai Vongbunyong

DAY 2 - PM (5) WEDNESDAY, NOVEMBER 18

IMC-05: Management Technology [Day 2 - PM] Room D

Session Chair: Warut Pannakkong, Jirachai Buddhakulsomsiri, and Pisal Yenradee

[14:50-15:05]
Design of Freemium with Usage Limitation (IMC-AETM2020-0006)
Tanet Kato and Aussadavut Dumrongsiri

[15:05-15:20]
Using conjoint analysis to elicit women preferences on personal health insurances: A case of working women in Bangkok, Thailand (IMC-AETM2020-0009)
Praewpaka Chumtong Drucker and Akkaranan Pongsathornwiwat

[15:20-15:35]
Delivery Location Positioning and Delivery Planning Systems for Thai SMEs using Google Maps (IMC-AETM2020-0017)
Sophea Horng and Pisal Yenradee

[15:35-15:50]
Impacts of shift in share of consumption from palm-oil based B10-biodiesel to B20-biodiesel in Thailand (IMC-AETM2020-0019)
Satish Acharya and Mongkut Piantanakulchai

[15:50-16:05]
Problem With Sequence-Dependent Setup Time Under Machines Eligibility Restriction to Minimize Total Earliness and Tardiness: A Case Study in a Company (IMC-AETM2020-0020)
Seyha Thay and Morrakot Raweewan

[16:05-16:20]
Estimation of Remaining Service Life of Expressway in Bangkok based on Inspection Results (IMC-AETM2020-0022)
Hanghoy Tek, Pakawat Sancharoen, Natthapon Suksomklin, Natthawat Suksomklin, and Somnuk Tangtermsirikul

DAY 3 - AM (1) FRIDAY, NOVEMBER 20

Main Session [Day 3 - AM]

[08:00-09:00]
Registration to the Zoom

DAY 3 - AM (2) FRIDAY, NOVEMBER 20

iSAI-07: Data Analytics and Machine Learning [Day 3 - AM] Room A

Session Chair: Sanparith Marukatat

[10:30-10:45]
Mushroom Classification by Physical Characteristics by Technique of k-Nearest Neighbor (iSAI-NLP-AIoT2020-0158)
Narumol Chumuang, Kittisak Sukkanchana, Mahasak Ketcham, Worawut Yimyam, Jiragorn Chalermdit, Nawarat Wittayakhom, and Patiyuth Pramkeaw

[10:45-11:00]
Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm (iSAI-NLP-AIoT2020-0160)
Pronab Ghosh, Saima Afrin, F. M. Javed Mehedi Shamrat, Atqiya Abida Anjum, Shahana Shultana, and Aliza Ahmed Khan

[11:00-11:15]
Developed Credit Card Fraud Detection Alert Systems via Notification of LINE Application (iSAI-NLP-AIoT2020-0131)
Narumol Chumuang, Sansanee Hiranchan, Mahasak Ketcham, Worawut Yimyam, Patiyuth Pramkeaw, and Sakchai Tangwannawit

DAY 3 - AM (3) FRIDAY, NOVEMBER 20

iSAI-08: Signal, Image and Speech Processing [Day 3 - AM] Room B

Session Chair: Nongnuch Ketui

[10:30-10:45]
COVID 19 X-Ray Image Classification using Voting Ensemble CNNs Transfer Learning [Short] (iSAI-NLP-AIoT2020-0138)
Phuwadol Viroonluecha, Thanwarat Borisut, and Jose Santa

[10:45-11:00]
Intelligent medicine identification system using a combination of image recognition and optical character recognition (iSAI-NLP-AIoT2020-0146)
Nagorn Maitrichit and Narit Hnoohom

[11:00-11:15]
Face Recognition System for Financial Identity Theft Protection (iSAI-NLP-AIoT2020-0164)
Thidarat Pinthong, Worawut Yimyam, Narumol Chumuang, Mahasak Ketcham

DAY 3 - AM (4) FRIDAY, NOVEMBER 20

IMC-06: Chemistry and Environment [Day 3 - AM] Room C

Session Chair: Pakorn Opaprakasit

[10:30-10:45]
Input Data Inventory Analysis for Environmental Life Cycle Assessment of Organic Fraction of Solid Waste Treatment by Composting and Anaerobic Digestion: A Case Study in Sri Lanka (IMC-AETM2020-0027)
W.T.Rasangika Thathsaranee and Sandhya Babel

[10:45-11:00]
Vancomycin production from biodiesel-derived crude glycerol and soybean flour (IMC-AETM2020-0042)
Le Phuong Vy and Wanwipa Siriwatwechakul

[11:00-11:15]
Chitosan Thin Film for Food Packaging with Enhanced Antibacterial Properties (IMC-AETM2020-0043)
Nunik Nugrahanti and Wanwipa Siriwatwechakul

[11:15-11:30]
Performance Comparison of CRMC and Conventional Ejector on R141b Ejector Refrigeration System (IMC-AETM2020-0045)
Boriak Kitrattana and Satha Aphornratana

DAY 3 - AM (5) FRIDAY, NOVEMBER 20

IMC-07: Management Technology [Day 3 - AM] Room C

Session Chair: Akaranan Pongsathornwiwat

[10:30-10:45]
A Comparison of Organizational Culture and Performance of Chinese Entrepreneurs with Different Genders and Amount of Work Experience (IMC-AETM2020-0023)
Qiuxue Luo, Mayuree Suacamram, Luckxawan Pimsawadi

[10:45-11:00]
Customer Review and Brand Experience: The case of Smart Product Brand (IMC-AETM2020-0034)
S M Arifuzzaman and Pisit Chanvarasuth

[11:00-11:15]
Elephant Scheduling for Tourism Industry Using a Mathematical Model: A Case Study (IMC-AETM2020-0036)
Rawisara Chongphaisal and Aussadavut Dumrongsiri

[11:15-11:30]
Genetic Algorithm with Reinforcement Learning for Production Scheduling and lot size (IMC-AETM2020-0037)
Vannak Seng, Chawalit Jeenanunta, and Masahiro Nakamura

[11:30-11:45]
Mathematical model of COVID-19 transmission dynamics in Thailand (IMC-AETM2020-0039)
Chirarak Wongthitaporn and Rujira Chaysiri

Papers

(Sorted by paper ID)

A Memetic Algorithm for Tour Trip Design Problem

Apisit Cheng
School of Management Tecnology
Sirindhorn International Institute of Technology, Thammasat University
Pathum Thani, 11120 Thailand
j.cheng.apisit@gmail.com

Aussadavut Dumrongsiri
School of Management Tecnology
Sirindhorn International Institute of Technology, Thammasat University
Pathum Thani, 11120 Thailand
aussadavut@siit.tu.ac.th

***Abstract*—to design a tour plan which provide a maximum satisfaction, before have any experiences with the destination can be hard and time consuming process. The goal of this study is to create an algorithm that efficiently generate a tour plan with high or maximum satisfaction within a reasonable processing time. The memetic algorithm which is a combination of genetics algorithm and local search algorithm would be created to solve this problem. This study used real data gathered from trusted tourist community in Thailand such as TripAdvisor.com, Wongnai.com, etc. The result of this study shown Memetic Algorithm (MA) approach could solve tour trip design problem efficiently since both saving in computation time and %gap are in a good shape and well-balanced.**

Keywords—Tour trip design problem, Memetic algorithm, Mathematical model, Travel routing, Orienteering problem

I. INTRODUCTION

According to report from World Travel & Tourism Council (WTTC), a total contribution of travel & tourism to Thailand GDP was 19.7% of GDP in 2019 (Those number calculated from total spending on Travel & Tourism within Thailand for business and leisure purposes by residents and non-residents, and spending by government on Travel & Tourism service directly linked to visitors). This show that Travel & Tourism is one of the largest Thailand economic sectors, therefore the goal for Thailand is to increase or at least maintain the same level of tourism. One possible solution for Thailand to achieve this goal is to create a tour plan that could leads to a high overall consumer tour satisfaction by using an information from past consumption experiences. Many studies, shows that overall consumer's tour satisfaction is an important factor directly affect intension of tourist to recommend (positive-review) a tour to their peer and relatives, and also affect intension to re-buy the tour or re-visit a region ([1]; [2]; [3]). In this study, an approach to efficiently plan a tour for consumer would be studied. A complete tour in this study is defined as a tour which consists of a D trip which connect initial and final location. Each trip must have start location, tourist attraction, restaurant for lunch and dinner, and end location. And the tour have a time and budget constraint as a limits. According to the above condition, in this study, tour planning process/ tour trip design problem (TTDP) could be seen as Orienteering Problem with Hotel and Restaurant Selection and Time Window (OPHRSTW). Since its shared the following similarities: 1. Objective of an orienteering is to maximize the score which tour planning process also aim to visit a place that have a high consumer's satisfaction which could be measured by review score from past consumer experiences on the internet, 2. Objective value could only be increase by visiting as much place as possible, 3. There is a time limit in each trip. Based on the similarity described above, the real world tour planning problem could be converted to a mathematical model for computer to solve this kind of problem, which called "Optimization approach", to the optimal point (the best solution).

The OP was proved to be NP-hard by [4]. Since the OP is NP-hard, it extensions such as OPHRSTW will also be NP-hard. The term NP-hard mean that to solve the problem to optimality is very time consuming. Thus, heuristic approach would also be studied along with optimization approach to compare performances of both approach.

II. REVIEW OF LITERATURE

A. Orienteering problem and it extensions

Sport game named Orienteering is the origin of the Orienteering Problem (OP) [5]. The objective of this sport game is to maximize a total collected score that obtained from visiting each check points and come back to starting point within a time limit. [6] considered it as a combination between the Knapsack Problem and the Travelling Salesperson Problem (TSP), and also known as the Selective Traveling Salesperson Problem [7] and Maximum Collection Problem [8].

There are many extension of the OP such as the OP with Time Window (OPTW), The Team OP (TOP), the generalized OP (GOP), the OP with Hotel Selection (OPHS), and the OP with Mandatory visits and Exclusionary Constraints (OPMVEC). For a recent survey, see ([9]; [10]; [11]). Due NP-hard nature of the OP and it extensions, many study have develop a heuristic algorithm to solve them. For example, Tabu Search (TS), Greedy heuristics, Variable Neighborhood Search (NVS), Skewed Variable Neighborhood Search (SVNS), Simulated Annealing (SA), Artificial Neural Network (ANN), Ant Colony Optimization (ACO), and Memetic Algorithm (MA). See ([6]; [10]; [12]; [13]) for a recent survey.

B. Orienteering problem with hotel and restaurant selection and time window

As OPHRSTW is one of a variant of OP, so it shares some similarities and have some differences with several other extension of OP. In the OPHS [12] a set of hotel and attraction are given with assigned score but in this study, the score would also assigned to each hotel. The tour consist of D trip which start and end at one of the hotel.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

Each hotel can be visited more than once. Each attraction can be visited only once. The time t_{ij} required to travel from location i to j is given for all pairs of location. The objective is to maximize a total collected score from all D trip in the tour, while OPTW have the same objective and almost the same constraint as OPHS. Except that OPTW have just one trip in the tour and each trip must be start and end at the same location. To visit each location in OPTW, the time window must be check whether arrive time fall within a given time window [14].

Thus, OPHRSTW is a research that would be a combination of two existing research fields and also have more node type (restaurant node) to be considered than those existing problems.

C. Other related research

Many studies propose method and algorithm to solve OP and its variant in field of tour planning to create a tour planning that could support tourism decisions which could leads to success tourism. An overview of related research of TTDP are shown on Table I.

TABLE I. LIST OF RELATED STUDIES

Author	Hotel	Restaurant	Multi-Day	Time Window
[15]	-	-	-	-
[16]	X	-	X	X
[17]	X	X	-	-
[18]	X	-	-	X
[19]	-	-	-	X
[20]	X	-	X	X

[17] creates a mathematical model for tour planning process which result produced by the model show priority of the tourist attraction site. Information of available tourism route played an important role in the tour planning process when choosing which tourist attraction to be included to the tour. They used a combination of genetic algorithm (GA) and difference evolution algorithm (DEA) to solve the problem. The result generated by this algorithm could provide an effective tour plan that could satisfied the stochastic condition of the problem. Similarly, [20] used one variant of GA algorithm, which is non-dominated sorting genetic algorithm II (NSGAII), to deal with bi-objective covering tour routing problem. Result of the study shows that NSGAII could solve the bi objective problem within a short-period of time while also provide excellence result in terms of tour plan quality.

III. MATHEMATICAL MODEL

A. Objective:

Maximize: $\sum_{d=1}^{D}\sum_{j=0}^{Att} SjYjd + \sum_{d=1}^{D}\sum_{j=0}^{H} SjZjd + \sum_{d=1}^{D}\sum_{j=0}^{RL} SjLUjd + \sum_{d=1}^{D}\sum_{j=0}^{RD} SjDIjk$ (1)

B. Subjected to:

$\sum_{j=0}^{AnL} x_{1j1} = 1$ d = 1,…,D (2)

$\sum_{i=0}^{AnD} x_{i2D} = 1$ d = 1,…,D (3)

$\sum_{j=0}^{V}\sum_{i=0}^{H} X_{ijd} = 1$ d = 1,…,D (4)

$\sum_{i=0}^{V}\sum_{j=0}^{H} X_{ijd} = 1$ d = 1,…,D (5)

$\sum_{j=0,j\neq i}^{V} X_{ijd} = \sum_{j=0,i\neq j}^{V} X_{jid}$ d = 1,…,D-1; i = 1,…, AnR (6)

$\sum_{i=0,i\neq j}^{V}\sum_{d=1}^{D} X_{ijd} \leq 1$ j = 1,…,AnR (7)

$Y_{jd} = \sum_{i=0,i\neq j}^{V} X_{ijd}$ j = 1,…,Att; d = 1,…,D (8)

$Z_{jd} = \sum_{i=0}^{AnD} X_{ijd}$ j = 1,…,H; d = 1,…,D (9)

$\sum_{j=0}^{V} X_{jid} = \sum_{j=0}^{V} X_{ijd+1}$ i = 0,…,H; d = 1,…,D-1 (10)

$LU_{jd} = \sum_{i=0}^{AnH} X_{ijd}$ j = 0,…,RL; d = 1,…,D (11)

$\sum_{i=0}^{V}\sum_{j=0}^{RL} X_{ijd} = 1$ d = 1,…,D (12)

$DI_{jd} = \sum_{i=0}^{AnL} X_{ijd}$ j = 0,…,RD; d = 1,…,D (13)

$\sum_{i=0}^{V}\sum_{j=0}^{RD} X_{ijd} = 1$ d = 1,…,D (14)

$ArrT_{id} - ArrT_{jd} + TrT_{ij} + TS_i \leq (1\text{-}X_{ijd})\, M$ i = 0,…,V; j= 0,…,AnR; d = 1,…,D (15)

$ArrT_{id} - ArrTW_{jd} + TrT_{ij} + TS_i \leq (1\text{-}X_{ijd})\, M$ i = 0,…,AnR; j= 0,…,H; d = 1,…,D (16)

$OpT_i \leq ArrT_{id} \leq ClT_i$ i = 0,…,V; d = 1,…,D (17)

$OpT_i \leq ArrTW_{id} \leq ClT_i$ i = 0,…,H; d = 1,…,D (18)

$\sum_{d=1}^{D}\sum_{i=0}^{Att} EC_i Y_{id} + \sum_{d=1}^{D}\sum_{j=0}^{H} EC_j Z_{jd} + \sum_{d=1}^{D}\sum_{lu=0}^{RL} EC_{lu} LU_{lud} + \sum_{d=1}^{D}\sum_{di=0}^{RD} EC_{di} DI_{did} \leq B$ d = 1,…,D; i = 0,…,Att; lu = 0,…,RL; di = 0,…,RD (19)

Equation (1) maximize total customer satisfaction score. Equation (2) and (3) guarantee that the tour start and end at desired location. Equation (4) and (5) ensure that each trip start and end in one of the available hotels. Equation (6) ensure connectivity of the path. Equation (7) ensure that each attraction and restaurant can be visited at most once. Equation (8) use for attraction score calculation. Equation (9) use for hotel score calculation. Equation (10) ensure that

the first trip and the second trip will end and start at the same hotel, respectively. Equation (11) use to calculate score of restaurants for lunch. Equation (12) ensure that each trip will have one and only one restaurant for lunch. Equation (13) use for restaurant for dinner score calculation. Equation (14) ensure that each trip will have one and only one restaurant for dinner. Equation (15) ensure timeline of the path for attraction and restaurant in each trip. Equation (16) ensure timeline of the path for ending location/hotel in each trip. Equation (17) limit that each attraction and restaurant must visited within time window. Equation (18) limit that each ending location/hotel must visited within time window. The reason to separate equation 17 and 18 is to allow the same hotel to be selected since combining these two constraints together would resulting in a constraint unbounded issue when number of day is more than number of available hotel in the list. Equation (19) limit total cost of the whole tour to be less than available budget.

IV. PROPOSED ALGORITHM

A. Structure of Memetic Algorithm for OPHRSTW

[21] was the first to introduced MA and defined it as a combination of genetic algorithm and local search technique such that have more capability to explore and exploit promising region of the search space. It was successfully applied to solve VRPs by [22]. For the OP, MA was first applied by [23]. Many study successfully applied MA to solve OP extension and related VRP such as a Memetic Algorithm for the Orienteering Problem with Hotel Selection [6], The Memetic algorithm for the optimization of urban transit network [24], and A Memetic Algorithm for Orienteering Problem with Mandatory Visits and Exclusionary Constraints [10]. The structure of MA for OPHRSTW compose of two parts: The first part, Population initialization which start the process by preprocessing step that calculate a score for each pair of hotel then used it to create initial population. After initial population is created, it will be processed through the second part called Main-loop. This process will repeated until it reach Maximum iteration (*MaxIteration*). See Fig. 1 below for more detailed MA structure description.

1. Population Initialization:
  1.1. Pre-Processing Step
  1.2. Creating initial population
2. Main-loop:
  Iter = 1
  While (*Iter* <= *MaxIteration*) **Do:**
    1. Populate the *Pool*
    2. Sort the *Pool* according to solution quality
    3. Save the *Best-Found-Solution*
    4. Population management
    Iter += 1
  End
Output: *Best-Found-Solution*

Fig. 1. Structure of Memetic Algorithm

B. Population Initialization

In population initialization, the estimated score for each pair of hotel is created in preprocessing step. After that it is used to generate the initial population. Note that the term "population" in this study refers to a tour which can consist of one trip or multiple trip. A detailed algorithm of the population initialization part on Fig. 2 and Fig. 3.

C. Matrices of pairs of hotels

a) Insert and Add (a) Algorithm: The different between insert and add is that insert will be used to determine a position to insert attraction, while add will be used to add restaurant to a slot position. For each non-included attraction in the trip, the best position of insertion that have minimum association cost is determined. Once leave time fall within a specified range of lunch and dinner time the restaurant will be added. Note that an attraction and restaurant that have the maximum Node Position Ratio (NPR), which calculated from (Score / Traveling time + waiting time at target node if its not available at the time of arrival), and does not make the trip infeasible is inserted/added.

b) Replacement Algorithm: For all non-included attraction, the NPR of each position in the trip is calculated to check whether there are any position that non-included attraction node have a larger NPR than the attraction node in the trip. Then it is to be checked whether replacing a target non-included attraction node to the trip would make the trip feasible. Otherwise no replacement executed. Only the attraction node with largest NPR value would be selected.

c) Two-Opt (a): This heuristic determined the inversion in position of each pair of attraction in the trip that leads to highest overall trip time reduction without making the trip infeasible. On the other hand, if there is no solution that could reduce overall trip time or it leads to an infeasible solution then the original solution is returned.

d) Move-best (a): For each attraction, it is checked that if move this attraction to another position in the trip will resulting in overall trip time reduction, and feasible solution. The solution that have the most time reduction would be selected. However, if there is no solution that could reduce overall trip time or leads to an infeasible solution then the original solution would be selected.

Best = the current incumbent solution
Set of four greedy moves (Gr_k): *Insert (Att) and Add (Res)*, *Replacement (Both)*, *Two-Opt (Both)*, *Move-best (Att)*
NoI = 0
While (*NoI*<*4*) **Do:**
  Best' = *Best* after applied greedy move Gr_k on
  If *X'* > *X* **then**
    Best = *Best'*
    NoI = 0
  Else
    NoI += 1
End

Fig. 2. Logic structure of pre-processing step

D. Creating Initial Population

This process create an initial population by appending sequence of hotel to the tour one by one based on probability which proportioned to their score. For the tour with more than 1 day long, the end hotel of each day and a start hotel of the next day must be the same. If it makes an infeasible solution, the algorithm will try to improve the solution using feasibility improvement algorithm to make it feasible. After each individual population is created, it is improved by a local search technique. The whole process is repeated to create solution equals to the specified limit *PopSize*.

```
Creating Initial Population:
    k = 1
    While (k < PopSize) Do:
        G_k = Generate an initial solution
        Apply solution feasibility improvement algorithm on G_k to ensure feasibility of the solution
        If initial solution is a feasible solution then:
            Apply local search on G_k to improve solution quality
            k += 1
    End
Current Population = Initial Population
Calculate and save the Best-Found-Solution
```

Fig. 3. Logic structure of creating initial population

E. Feasibility Improvement

Feasibility improvement algorithm checks and improve solution feasibility. The following conditions would be checked to ensure the feasibility of the solution. 1. Duplication of attraction and restaurant (lunch/dinner) node across the tour, 2. Check number of restaurant whether it less than or exceed 1 per type (lunch/dinner), 3. Time window of an included node in each trip of the tour, 4. Total cost must lower than specified budget. Fig. 4, the algorithm to check feasibility of the above condition was shown and would be described on below section

```
Feasibility Improvement (Attraction/Restaurant):
    For each trip in tour:
        If number of attraction/ restaurant node duplications > 0 then:
            Remove duplication across the tour
        If number of lunch/dinner node in trip < 1 then:
            Insert lunch/dinner node into the trip
            If lunch/dinner node insertion couldn't process then:
                Replace attraction with lowest ratio with lunch/dinner node
        If number of lunch/dinner node in trip > 1 then:
            Remove lunch/dinner node with the lowest NPR value from the trip
        Improve solution time window
End
```

Fig. 4. Logic structure of feasibility improvement algorithm

a) Remove duplication: Logic to remove duplicate in attraction and restaurant node across the tour: the algorithm would check whether there are any duplication in attraction and restaurant node across the tour. If duplication occurred the node with the lowest NPR value would be removed from the tour.

b) Check number of restaurant node (lunch/dinner): this is an algorithm to check whether it equals to 1 per trip across the tour. If number of restaurant (lunch/dinner) in some trip is less than 1 then the algorithm would insert a restaurant node of type that were missing to the trip. But if the restaurant insertion could not be processed due to the time limit of each trip then the algorithm would remove some attraction node from the tour to create time slot. On the other hand, if the number of restaurant node is more than 1 then the algorithm would remove the restaurant node that have the lowest NPR value from the trip.

F. Local Search

After each solution processed through the feasibility improvement algorithm, then local search heuristics is used to improve attractions and restaurants between each pair of hotel

```
Best = the current incumbent solution
Set of four greedy moves (Gr_k):
Insert (Att) and Add (Res), Replacement (Both), Two-Opt (Both),
Move-best (Att), Swap (Res) and Swap-Best (Att), Extract-Insert,
Extract2-Insert
NoI = 0
While (NoI<4) Do:
    Best' = Best after applied local search Gr_k on
    If X' > X then
        Best = Best'
        NoI = 0
    Else
        NoI += 1
End
```

Fig. 5. Logic structure of Variable Neighborhood Descent with seven local search heuristics

a) Insert and Add (b): Main logic of this algorithm is the same as Insert and Add (a) algorithm in pre-processing step. The only different is that this algorithm considered to insert and add node to a multiple trip within the tour.

b) Two-Opt (b): Main logic of this algorithm is the same as Two-Opt (a) algorithm in pre-processing step. The only different is that this algorithm would try to do node inversion in every trip within the tour, while the algorithm in pre-processing step could be used with only one trip. This algorithm would iterated through each trip in tour to check whether Two-Opt (b) should be executed.

c) Move-best (b): Main logic of this algorithm is the same as Move-best (a) algorithm in pre-processing step. The only different is that this algorithm could be used to find the best target position across multiple trip in the tour for each attraction node to be moved to.

d) Swap-Best: The purpose of Swap-Best algorithm is to reduce overall trip time. Swap-Best will be used to determine the best pair of attraction node to be swapped. It could be pair of attraction node from the same trip or pair of attraction node from different trip that could provide the most reduction on overall tour time.

e) Extract-Insert-Replace: This algorithm would be applied to 2 types of node which are attraction and restaurant node. For attraction node, the algorithm would remove one attraction node from each trip in the tour then insert other unvisited attraction node as much as possible. Only attraction node with the highest NPR value that would be selected in each round of the algorithm. If there are any improvement in total score of a solution then the score would be considered as master solution, otherwise original solution would still be considered as master solution. The same approach also applied to each restaurant in each trip of the tour, but the different is that only one node of type restaurant for lunch/dinner that would be insert to the trip.

G. Main-Loop

In this part, the current population will be applied two types of genetic operator which are crossover and mutation. After each genetic operator applied, the feasibility improvement algorithm and local search technique will be applied to improve quality of the solution. After that the population management is processed to increase diversity of a solution pool. Below section explain algorithm with a more detailed.

a) Selection: The selection method used in both crossover and mutation process is the roulette wheel. This method assigns a probability P_i to each solution in the population, where probability calculated from their score over total score [25]:

$$P_i = f_i / \sum_{i=1}^{V} fi \quad (20)$$

b) Termination: In this study, *MaxIterations* is used to limit maximum generation created by this algorithm, this parameter need to be adjusted to balance the solution quality and computation time

c) Crossover: Crossover method used in this study is a one-point crossover method, two parent solutions is selected for crossover process to create two offspring solutions (O_1 and O_2). The algorithm randomly selected one trip in the tour and then randomly select a time in the selected trip to be used as a chromosome cut point for crossover process. This method would be called "trip and time cut" method and a target node that would be cut called "node to be cut". There are 2 parameters that used to control the number of crossover process as follows: 1. Current Population size (*PopSize*) 2. *CrossOver* rate (*Cr*).

OffSpringCount = 1
While (*Cr* <= *CrP* * *Popsize*) **Do (Time cross over):**
Select solution P_1 and P_2 from the Current Population (Roulette Wheel)
Apply Feasibility Improvement on O_1 and O_2
Apply Local Search on O_1 and O_2
Add O_1 and O_2 to the Pool
OffSpringCount += 1
End

OffSpringCount = 1
While (*Mut* <= *(Mut** *Popsize*) **Do:**
Select solution P_1' from the Current Population (Roulette Wheel)
O_1' = Apply Mutation on P_1'
Apply Feasibility Improvement on O_1'
Apply Local Search on O_1'
Add O_1' to the *Pool*
OffSpringCount += 1
End

Fig. 6. Algorithm for Crossover and Mutation

d) Mutation: In mutation process, one parent solution is randomly selected and mutated to create one offspring. There are two main cases for mutation algorithm as follow: The first scenario, multiple trip in one tour - One trip in the parent solution with multiple trip will randomly selected and the last hotel of it will be replaced with a new end hotel from matrices pairs of hotels based on their score. If new end hotel is not matches with the start hotel of the next trip then the algorithm would also randomly selected a new start hotel of the next trip but fixed the end hotel of the next trip to be the same hotel. The second scenario: one trip in one tour (one day tour) – One trip in the parent solution will randomly selected then trip and time cut would be executed to split a parent solution into two parts then combined it with another solution from matrices of pair of hotel that also split into two parts by the same method. There are 2 parameters that used to control the number of crossover process as follows: 1. Current Population size (PopSize) 2. Mutation rate (Mut). The mutation algorithm will be explained below and show on Fig. 6.

H. Population Management

Empty the Current Generation
Ind_{c1} = 1
While (Ind_{c1} <= PopSize) **Do:**
i = Ind_{c1}
Select the i[th] solution from the *top* of the Pool and add it to the Current Population (Roulette Wheel; top-down)
Select the i[th] solution from the *bottom* of the Pool and add it to the Current Population (Roulette Wheel: bottom-up)
Ind_{c1} += 2
End

Fig. 7. Algorithm for Population Management

This algorithm will randomly select solution based roulette wheel by using top-down and bottom-up approach in each iteration of the algorithm. Top-down approach is the approach that a solution with higher score tent to have more chance to be selected. On the other hand, Bottom-up approach is the approach that a solution with lower score would have more chance to be selected.

V. EXPERIMENTS AND RESULTS

A. Dataset Explanation

A review/customer satisfaction score was gathered from Agoda.com, Wongnai.com, and TripAdvisor.com. The reason to select these three websites is that these website is a creditable source of information based on number of users provided a review on their website. Data collection for destination in the model of this study was selected from the central province region of Thailand which are Kanchanaburi, Nakhon Pathom, and Ratchburi. These province have many natural and historical tourist attraction, which could represent cultures and history of Thailand, so that tourist could learn and enjoy Thai culture and history.

B. Result and Discussion

TABLE II. EXAMPLE OF OPHRSTW SOLUTION BY OPTIMIZATION AND MA APPROACHES

Test case	Approach	Score	Total Cost	Computation time (second)
1	Optimization	24.32	475.50	10.00
	MA	23.40	470.00	3.30
	difference	3.78%		6.70 sec.
2	Optimization	25.54	620.50	33.00
	MA	24.77	900.50	2.50
	difference	3.01%		30.50 sec.
3	Optimization	24.70	951.00	29.00
	MA	24.70	951.00	2.90
	difference	0.00%		26.10 sec.
4	Optimization	26.32	375.00	351.00
	MA	24.65	670.00	3.70
	difference	6.34%		347.30 sec.
5	Optimization	27.98	1160.00	47.57
	MA	27.98	1160.00	3.64
	difference	0.00%		43.93 sec.
6	Optimization	28.65	595.50	746.20
	MA	28.48	675.50	3.55
	difference	0.59%		742.65 sec.
7	Optimization	27.98	825.00	66.27
	MA	27.98	825.00	3.09
	% difference	0.00		63.18 sec.

To test effectiveness and efficiency of the memetic algorithm, an experiment of computation was conducted. There are 51 test cases were created. Each of test case would have different number of node in each type

assigned. Each of the test case was solved using the ILOG CPLEX and the memetic algorithm (Coded in Python) running on the same computer with specifications as follow Operating System – Window 10 64-bit OS; Processor – Intel® Core ™ i7 – 6700HQ CPU @ 2.60GHz; RAM 16.00 GB. There are 51 test cases, the ILOG CPLEX could solve only 14 test cases. The computation times ranged from 10.00 to 746.20 seconds. For test cases that CPLEX unable to solve was due to it reached time limit which is a pre-set time limit of 28,800 seconds (8 hours). On the other hand, the memetic algorithm (MA) could solve all 51 test cases. The longest computation was only about 10 seconds with a difference in total score range from 0 to 6.34%. Below is an example of the result from the experiment. Fig. 8 and Fig. 9 provide a summary of performance of MA approach compared with Optimization approach. It is shown that saving in computation time when using MA approach on average is 158.92 seconds, while the maximum is 742.65 seconds. And for %gap in total tour score on average is 1.59 with standard deviation of 2.24. As a summary, the MA approach could solve OPHRSTW efficiently since both saving in computation time and %gap are in a good shape and well-balanced.

Computation Time (in second)			
Algorithm/Value	Min	Avg	Max
Optimization	10.00	162.80	746.20
MA	1.47	3.88	3.55
Saving in computation time	8.53	158.92	742.65

Fig. 8. Summary of performance comparison

% Gap between Optimization and MA Approach			
Min	S.D.	Avg	Max
0.00	2.24	1.59	6.34

Fig. 9. Summary of %gap between Optimization and MA approach

C. Conclusions

The result shows that the MA could solve OPHRSTW efficiently as shown on TABLE II. The MA could provide a solution that is optimal or close to optimal point within a reasonable runtime when compared to optimization approach. Thus, MA is a tool that would satisfy users in terms of quality and time usage in solving OPHRSTW and its related problem with avg. %gap and avg. saving in computation time of 1.59% and 158.92 seconds, respectively.

REFERENCES

[1] Marin, J. A., & Taberner, J. G. (2008). Satisfaction and dissatisfaction with destination attributes: Influence on overall satisfaction and the intention to return. *Retrieved December*, *18*, 2011. Retrieved from

[2] Valle, P. O. D., Silva, J. a., Mendes, J., & Guerreiro, M. (2006). Tourist satisfaction and destination loyalty intention : A structural and categorical analysis. *International Journal of Business Science and Applied Management*, *1*(1), 25–44. Retrieved from http://business-and-management.org/library/2006/1_1--25-44--Oom_do_Valle,Silva,Mendes,Guerreiro.pdf

[3] Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management*, *26*(1), 45–56. https://doi.org/10.1016/j.tourman.2003.08.016

[4] Golden, B. L., Levy, L., & Vohra, R. (1987). The orienteering problem. *Naval Research Logistics (NRL)*, *34*(3), 307–318. https://doi.org/10.1002/1520-6750(198706)34:3<307::AID-NAV3220340302>3.0.CO;2-D

[5] Chao, I. M., Golden, B. L., & Wasil, E. A. (1996). A fast and effective heuristic for the orienteering problem. *European Journal of Operational Research*, *88*(3), 475–489. https://doi.org/10.1016/0377-2217(95)00035-6

[6] Divsalar, A., Vansteenwegen, P., Sörensen, K., & Cattrysse, D. (2014). A memetic algorithm for the orienteering problem with hotel selection. *European Journal of Operational Research*, *237*(1), 29–49. https://doi.org/10.1016/j.ejor.2014.01.001

[7] Gendreau, M., Laporte, G., & Semet, F. (1998). A Branch-and-Cut Algorithm for the Undirected Selective Traveling Salesman Problem. *European Journal of Operational Research*, *106*(2), 539–545.

[8] Butt, S. E., & Cavalier, T. M. (1994). A heuristic for the multiple tour maximum collection problem. *Computers & Operations Research*, *21*(1), 101–111. https://doi.org/10.1016/0305-0548(94)90065-5

[9] Gunawan, A., Lau, H. C., & Vansteenwegen, P. (2016). Orienteering Problem: A survey of recent variants, solution approaches and applications. *European Journal of Operational Research*, *255*(2), 315–332. https://doi.org/10.1016/j.ejor.2016.04.059

[10] Lu, Y., Benlic, U., & Wu, Q. (2018). A Memetic Algorithm for the Orienteering Problem with Mandatory Visits and Exclusionary Constraints. *European Journal of Operational Research*, *268*(1), 54–69. https://doi.org/10.1016/j.ejor.2018.01.019

[11] Vansteenwegen, P., Souffriau, W., & Oudheusden, D. Van. (2011). The orienteering problem: A survey. *European Journal of Operational Research*, *209*(1), 1–10. https://doi.org/10.1016/j.ejor.2010.03.045

[12] Divsalar, A., Vansteenwegen, P., & Cattrysse, D. (2013). A variable neighborhood search method for the orienteering problem with hotel selection. *International Journal of Production Economics*, *145*(1), 150–160. https://doi.org/10.1016/j.ijpe.2013.01.010

[13] Kataoka, S., & Morito, S. (1988). AN ALGORITHM FOR SINGLE CONSTRAINT MAXIMUM COLLECTION PROBLEM. *Journal of the Operations Research Society of Japan*, *31*(4). Retrieved from http://www.orsj.or.jp/~archive/pdf/e_mag/Vol.31_04_515.pdf

[14] Kantor, M. G., & Rosenwein, M. B. (1992). The Orienteering Problem with Time Windows. *The Journal of the Operational Research Society*, *43*(6), 629–635.

[15] Yan, H., Hongzhi, G., & Duan, J. (2014). Tour Route Multiobjective Optimization Design Based on the Tourist Satisfaction Discrete Dynamics in Nature and Society 2014: 603494.

[16] Zhu, C., Hu, J. Q., Wang, F., Xu, Y., & Cao, R. (2012). On the tour planning problem. *Annals of Operations Research*, *192*(1), 67–86. https://doi.org/10.1007/s10479-010-0763-5

[17] Xiao, Z., Li, S., Fan, Y., Liu, B., Zhang, B., & Li, B. (2017). Tourism Route Decision Support Based on Neural Net Buffer Analysis. Procedia Computer Science 107: 243–47.

[18] Wu, X., Guan, H., Yan, H., & Ma, J. (2017). A tour route planning model for tourism experience utility maximization. Advances in Mechanical Engineering 9: 1–8.

[19] Liao, Z., & Zheng, W. (2018). Using a heuristic algorithm to design a personalized day tour route in a time-dependent stochastic environment. Tourism Management 68: 284–300.

[20] Nedjati, Arman, Gokhan Izbirak, and Jamal Arkat. 2017. Bi-objective covering tour location routing problem with replenishment at intermediate depots: Formulation and meta-heuristics. Computers & Industrial Engineering 110: 191–206.

[21] Moscato, P. (1999). Memetic algorithms: A short introduction. In D. Corne, M. Dorigo, & F. Glover (Eds.), New Ideas in Optimization (pp. 219–234). McGrw-Hill

[22] Prins, C. (2004). A simple and effective evolutionary algorithm for the vehicle routing problem. *Computers & Operations Research*, *31*(12), 1985–2002. https://doi.org/10.1016/S0305-0548(03)00158-8

[23] Bouly, H., Dang, D. C., & Moukrim, A. (2010). A memetic algorithm for the team orienteering problem. *4or*, *8*(1), 49–70. https://doi.org/10.1007/s10288-008-0094-4

[24] Zhao, H., Xu, W. (Ato), & Jiang, R. (2015). The Memetic algorithm for the optimization of urban transit network. *Expert Systems with Applications*, *42*(7), 3760–3773. https://doi.org/10.1016/J.ESWA.2014.11.056

[25] Talbi, E.-G. (2009). *Metaheuristic: from design to implementation*. *John Wiley & Sons.*

Improving of Pre-Processing Image for Child Cleft Lip by Image Processing

Jutturong Charoenrit
Department of Computer Science
Faculty of Science, Khon Kaen University, Thailand
jutturong@kkumail.com

Wararat Songpan
Department of Computer Science
Faculty of Science, Khon Kaen University, Thailand
wararat@kku.ac.th

Apisak Pattanachak
Department of Computer Science
Faculty of Science, Khon Kaen University, Thailand
apisak@kku.ac.th

***Abstract*—This paper proposed improving of pre-processing image for child cleft lip which are very important to evaluate shape of lip after operation. This method performed with image enhancement by shades and adjust image intensity by adjust image intensity as watershed method, which are applied for edge lip detection. The problem can be solved by the image that has too brightness and noise remove from the shave area upper lip. The problem is factors that are effecting to lib detection image as threshold background and unable close mouth completely, image segmentation. Therefore, this paper proposed approach to solved improving the of pre-processing image for child cleft lip that are suitable before image processing of lip detection clearly. To evaluate treatment results with experts according to the principle of nasolabial from 5 person. To reduce conflicts in inconsistent results conclusion by measuring pixel diff of the white pixel that have a minimum distance. In addition, the minimum distance value both the pixel and white pixel difference was found and show the efficiency of image processing. Finally, this paper is the most suitable method is the Zero-crossing detection method in the results obtained from Pixel difference and Distance difference are the smallest values.**

Keywords— child cleft lip, image processing, lip detection, image segmentation, adjust image intensity

I. INTRODUCTION

Cleft lip and palate [1] is a type of facial anomaly which found abnormalities in the upper lip and oral cavity can occur in newborns. In this paper, these images of cleft lip and cleft palate children are brought from the medical examiner of the Srinakharin hospital and Tawanchai center. There are many factors that make these images cannot be processed to find the shape of lips. For example, images with many noises [2], images with too much brightness, pale lips do not cut against the background, and images of patients with moustache above their lips. Therefore, it is necessary to have a child patient picture preparation process before image processing such as adjusting the brightness for some point, adding beige shades to the intensity of the lips, processing the shape of lips with watershed method, classify the primary colors which are red, green, and blue (RGB) [3],[4] with HSV method [5], and method to find the clearest shape of lips. Then crop [6] only the upper part of lips, convert it into grey and to black and white (binary) [7], and reduce the noises of images. Change the background color into black which makes pixel to be continuous. After that, convert the background color into white to find the edge line of lips with edge detection processing [8] from these methods, Roberts, Prewitt, Sobel, Canny, LoG, and Zero-crossing detection to find the sum of pixel difference. Draw the edge line of lips then measure the overlap between the border of the images and the real lips, the pixel of non-overlapping will be white, the value of the pixel difference is 1 and the distance difference between the border of the images and the real lips. The most appropriate method gives out the least sum of pixel difference and distance difference to lead to the evaluation results after treatment for cleft lip and cleft palate children [1] that has been surgically treated symmetrical or not. This is to reduce conflicts from inconsistent evaluation results to find the conclusion after surgery. However, the problem is how to pre-processing image in which is still adjusted image from original image into other steps.

In this paper focused on improving the pre-processing image for child cleft lip. The method of increasing the intensity of the lips image by color adjusting method color and separate them into three colors, including red, green and blue which used the watershed method before filter them by HSV technique to separate each color. The step is organized into five sections. The section II reviews the literature and related research. The section III describes the pre-processing of our approach. The section IV explains the experimental results. Finally, the summarization of the study results is shown in Section V.

II. RELATED WORKS

A. Related theory

1) Cleft lip and cleft palate [1] is one of the most common congenital disabilities of a child which is a disability of the head and face, and also a condition that has many defects such as abnormal structure of mouth and face, problems with food absorption, slow development and growth, the occlusion of abnormal teeth and difficulty in speaking (speaking unclear, drone, hoarse, and otorrhea). Children with cleft lip and cleft palate require many complex maintenance and long time to heal that needs to be treated from birth to at least 20 years old.

2) There are many factors that cause the chance of cleft lip and cleft palate so it is unable to specify the exact cause but found that the incidence of disease caused by internal, external, or environmental factors below.

a) Internal factors such as congenital, it is found that this condition is associated with genetics approximately 12-20% of all patients. And it is genetically inherited by people who have had this disease with family members.

b) External factors or environmental factors such as maternal illness during pregnancy, malnutrition during pregnancy, smoking or drinking alcohol during pregnancy, lacking of folic acid during pregnancy, obesity during pregnancy, and continuously receiving medication or substances during pregnancy such as antiepileptic, steroid, acne medication that contains Accutane, Methotrexate chemotheraphy including drugs that used to treat cancer, arthritis, and psoriasis

3) RGB Image [3] also called True color image will determine the number of each pixel, red, green, and blue in m x n x 3 array data format. It is the combination of red, green, and blue, each configuration is 8 bit in the form of RGB image equals to 24 bit and will be in the same plane of that position.

Converting RGB to YCbCr [9]

$$Y = (0.299 \times R) + (0.587 \times G) + (0.114 \times B) \quad (1)$$

$$Cb = (-0.168736 \times R) - (0.331264 \times G) + (0.5 \times B) \quad (2)$$

$$Cr = (0.5 \times R) - (0.418688 \times G) - (0.081312 \times B) \quad (3)$$

4) Gray-scale Image [10] Gray-scale Image is matrix data that shows the color value of the gray color. The single or double matrix class will use the concentration of the image between [0,1]. 0 stands for black and 1 for white, for the matrix class of unit8 or unit16 will use the concentration of the image between [0,255] or [0,65535]

5) Binary image or black and white image is the image that contains only black and white color, and digitally means that there are only two status which are 0 and 1 (0 for black, and 1 for white). The conversion of grayscale images to binary images must specify the intensity to be referenced or the Threshold value which the user can arrange by themselves or use the algorithm to find the value.

6) HSV color system [11],[12] HSV color system or Hue Saturation Value is a color consideration by using Hue Saturation Value. Hue is the value of primary colors (red, green, and blue). Pragmatically, the value will be between 0 and 255, if the Hue equals to 0 it will be in red. And Hue increases, the color changes according to the color spectrum until it reaches 256, it will turn into red again. This can also be substituted in degrees, red is 0 degrees, green is 120 degrees, and blue is 240 degrees. Hue can be calculated from RGB system as shown below.

$$Red = red - \min(red, green, blue) \quad (4)$$

$$Green = green - \min(red, green, blue) \quad (5)$$

$$Blue = blue - \min(red, green, blue) \quad (6)$$

7) Threshold the gray-scale image will be reversed before converting to a binary image which will select the gray level. As you can see in the original image, when reversing value for every pixel, this is to set the gray value to greater or less which is a part of image segmentation that needs to be separated the object from the background. And can be written in equation below.

$$g(x,y) = \begin{cases} a, & \text{if } f(x,y) < T \\ b, & \text{if } f(x,y) \geq T \end{cases}$$

when

$g(x,y)$ is a position (x,y) that is converted to binary 0 and 1

$f(x,y)$ is a posotion (x,y) is a dark gray level

T is the new threshold value

8) Edge detection [12] is to find the perimeter caused by the immediate change of the brightness of the image. By looking at the size and the rate of change of size. Examples of ways to detect edges:

a) Edge Detection: Is to find the perimeter caused by the immediate change of the brightness of the image. By looking at the size and the rate of change of size. Examples of ways to detect edges.

b) Gradient descent method[13] will find the edge by finding the lowest and highest point in the form of the first derivative of the image where the edge is located above the Threshold value [3] so it may cause the border to look thick. Example of ways to find the edges of this group are Roberts, Prewitt, Sobel, and Canny.

III. RESEARCH METHODOLOGY

A. Propose Pre-Processing Image Method

Normally, the image processing that consists many steps and this paper proposed a new method for pre-processing image by add five steps that including with step 4 to 8 in order to improving the image recognition clearly. These steps as showed in Figure 1 which can explain the steps as follows.

Step 1-3 to import and crop the lip area.

Step 4 to increase the intensity of the color by defining all 3 shades (RGB).

Step 5 crop only the upper part of the lip to limit image process area.

Step 6 Remove the background with use watershed method to show the shape of the object from RGB.

Step 7 Separate all 3 primary colors (RGB) from the watershed process by HSV method.

Step 8 Choose the primary color that gives the shape of the lips the most.

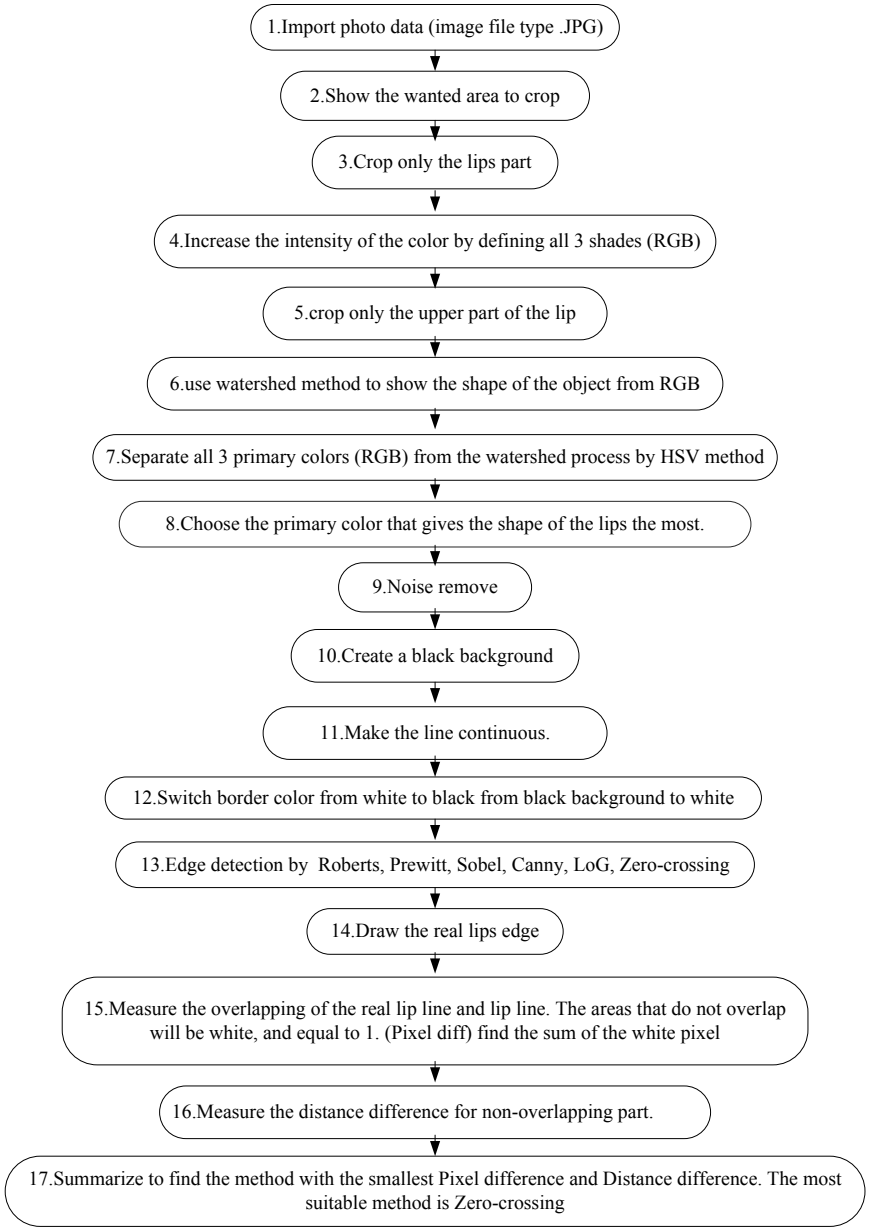


Fig 1. The improving of pre-processing image method for child cleft lip image detection

However, step 9 to 17 of child cleft lip images processing are the exist steps which described in [14] the edge detection and symmetry method for cleft lip and palate children using image processing research.

B. Data collection and System environment

The data is collected from cleft lip and cleft palate patients from Tawanchai Center, Faculty of medicine at Khon Kaen University.

MATLAB R2014a used for process under hardware that Intel® Core™ i5-3210M model, speed of CPU 2.50GHz,

RAM 12.0 GB and the operating system software that we used Microsoft Windows 10 Pro 64 bit.

IV. Experimental Results

The image used below are the 5 straight images of children with cleft lip and cleft palate after the surgery in RGB image. And each image consists of various factors such as too much brightness, too colorful background, patients with moustache above the lips, pale and unclear lips color.

(a) (b) (c) (d) (e)

Fig. 2. The sample of five pediatric patient's photos into .jpg file

The images used in the research are only the area of the patient's lips so it is necessary to crop the images by specifying the specific spot on the lips and increasing the intensity of the beige shade in the images. By Matlab command *Intensity of the image = imadjust (filename.jpg,[0.3 0.7],[]).*

(a) (b) (c) (d) (e)

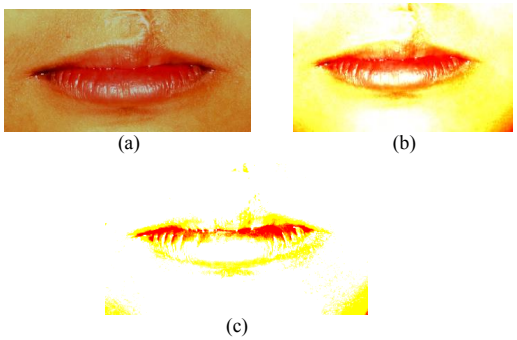
Fig. 3. Cropping the lip area with CascadeObjectDetector library

Have cropped only the lips area which is defined in the image 5 to use the images to process in the next procession.

To increase the shade level of the lips area by setting the shade level to *3* levels to find the image with the most suitable intensity shade level. The intensity of shade level can be classified as shown below. For example, from the Figure 3(a) we can adjust the shading intensity level *1* = the image that adjusts the color intensity * *(0.75)* as showed in Figure 4(a). Level *2* = the image that adjusts the color intensity * *(2.5)* as showed in Figure 4(b) and shading intensity level *3* = the image that adjusts the color intensity * *(7.5)* showed as Figure 4(c) respectively.

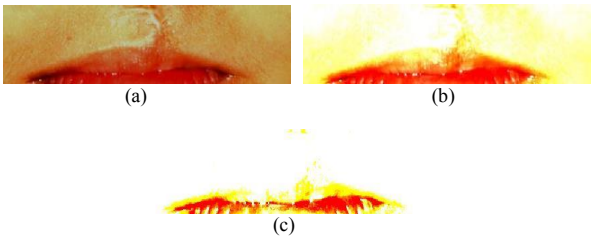
Next step, to choose only the upper lips of the image from the three levels of shading intensity, then select the most suitable image for watershed method.

In additions, in order to transform to RGB shade which we use the watershed method to display the shape of the object with primary colors (Red, Green, and Blue) as showed in Figure 6. Which Table I. showing suitable colors for the images to be used in edge detection process of the object.



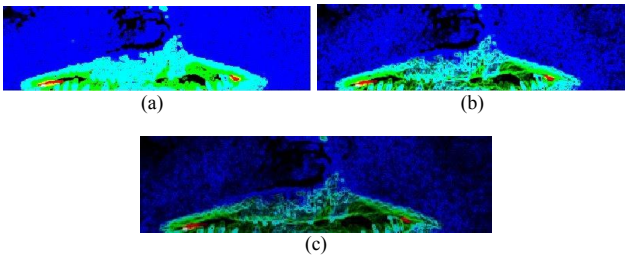
(a) (b) (c)

Fig. 4. Adjust the shading intensity (a) level 1, (b) level 2, (c) level 3



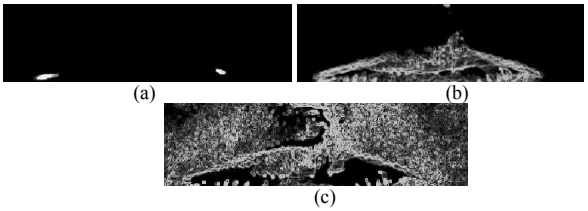
(a) (b) (c)

Fig. 5. Cropped only the upper part of lips



(a) (b) (c)

Fig. 6. Used the watershed method to show the lip shape from the three primary colors (RGB) a(red),b(green),c(blue)



(a) (b) (c)

Fig. 7. Used the HSV method to show the lip shape from The three primary colors (RGB) a(red),b(green),c(blue)

TABLE I. Suitable color to detect the shape of the object from watershed method

Image File	Colors selected from HSV
img1.JPG	Blue
img2.JPG	Green
img3.JPG	Green
img4.JPG	Green
img5.JPG	Red

From watershed method, pick the most suitable color and convert into binary image (black and white image) but there will be many noises in the image.

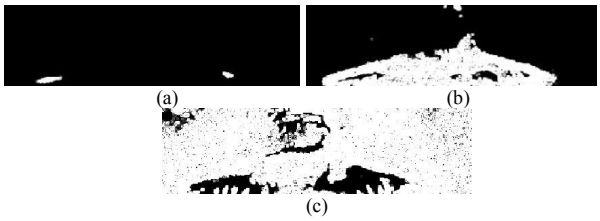


Fig. 8. Choose the most suitable color then convert to binary image red color (a) green color, (b) and blue color(c)

Remove the noises from the image from separating primary colors (RGB). Binary images have got many noises in the image so the noises need to be removed first to get the clearest edge image. The image on the right side is the image that has been removed the noises.

Fig. 9. Remove the noises from the image. The left image (a) is before removing (b)

From the Figure 9. the right image(b) is clearly image after removing the noise. And then the process of creating a black background image in order to obtain a clearer shape and edge of the object it illustrated in Figure 10. In additions, to make the line continuous to create the connecting point of the pixel as showed in Figure 11. Then, after make the line can transform image the background color from black color to white color that showed in Figure 12.

Fig. 10. Create a black background image. The left figure (a) is the image before creating the background. The right figure (b) is the creation of a black background.

Fig. 11. Make the line continuous. From the figure on the left (a) is before creating a continuation of the line, the figure on the right (b) is after making the line continuous.

Fig. 12. Change the background color to white(b) to obtain black (a) edge line of the object.

Therefore, to detect the lips edge by using Roberts, Prewitt, Sobel, Canny, Laplacian of Gaussian (LoG) and Zero-crossing detection method with showed in Figure 13.

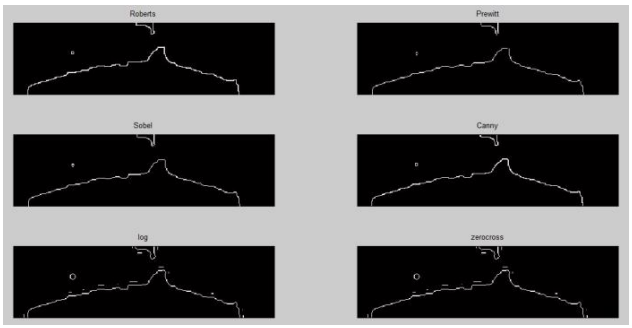


Fig. 13. Detect the edge of lips by using Roberts, Prewitt, Sobel, Canny, and Laplacian of Gaussian (LoG), Zero-crossing detection.

From the detected the lips in Figure 13. we can draw a lip edge to compare with the edge line of lips.

Fig. 14. Draw a lip edge to compare to the edge line of lips

Finally, the measurement of the overlapping of the lips to assess the effectiveness. The overlapping areas will be white lines, and the value equals to 1 (Pixel difference) measure the distance difference. Which the result of child cleft lip detection by image processing and illustrated in Figure 15.

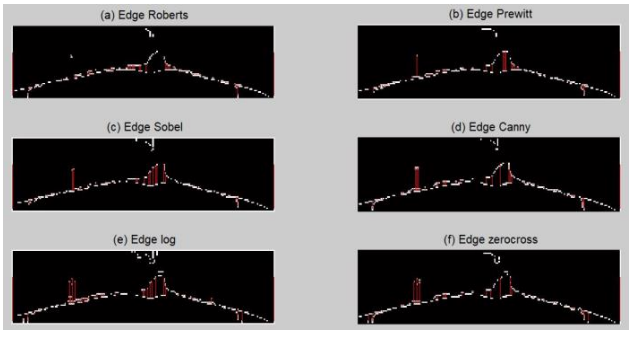


Fig. 15 Compare the lips to assess the effectiveness

The calculation for pixel diff the number of white pixel is a representation of the overlap the edges of the lips image. And Distance diff to measure vertically distance of the edge of the lips as showed in Table II.

TABLE II. SHOWS THE VALUE COMPARING BETWEEN PIXEL DIFFERENCE AND DISTANCE DIFFERENCE FROM THE METHOD OF ROBERTS, PREWITT, SOBEL, CANNY AND LAPLACIAN OF GAUSSIAN (LOG),ZERO-CROSSING DETECTION.

Methods	**Pixel diff**					
	Img1	**Img2**	**Img3**	**Img4**	**Img5**	**Avg.**
Roberts	343	542	834	593	1159	694.2
Prewitt	642	396	828	507	1008	676.2
Sobel	674	842	424	944	700	716.8
Canny	283	824	420	877	817	644.2
LoG	703	826	398	925	719	714.2
Zero-crossing	**274**	**803**	**396**	**867**	**733**	**614.6**
Methods	**Distance diff**					
	Img1	**Img2**	**Img3**	**Img4**	**Img5**	**Avg.**
Roberts	218	1437	890	1209	941	939.0
Prewitt	144	1169	789	1186	746	806.8
Sobel	228	1399	587	1261	753	845.6
Canny	175	1397	694	1096	903	853.0
LoG	273	1316	624	1134	800	829.4
Zero-crossing	**176**	**1368**	**609**	**1135**	**819**	**805.0**

V. CONCLUSION AND FUTURE WORK

This research has been done to improve the quality of the original in order to obtain a clearer edge line of lips image by increasing the intensity of beige shade and classify the shading level to 3 levels. Using watershed method to detect the shape of the object to get a suitable and clear image in order to detect the edge line of lips from Roberts, Prewitt, Sobel, Canny, Laplacian of Gaussian (LoG), and Zero-crossing detection method. The problems that occur are the image has the color that unclearly cut against the face color, the original image is too bright, pale lips, the face color does not cut against the lips color, too much noise, moustache above the lips, and the patient's lips are not completely closed. These methods impossible to separate the shape of the lips from the background of the image. Therefore, this research proposes guidelines for improving image quality to be suitable before the lip's shape detection process. By measuring the area of the overlapping line, the non-overlapping part will be white, and the value equals to 1 (Pixel difference) which the sum is all of white parts, and measuring the distance difference between the lip line to the image edge. The research presented proves that the Zero-crossing detection method gives out the least value for both pixel difference and distance. In the future work, can be apply these methods for the animal capture detection.

REFERENCES

[1] Krit Khwanngern. (2014). Comprehensive care cleft lip and cleft palate. Chiang Mai: Surgery Faculty of Medicine Chiang Mai University.

[2] M. Rakhshanfar and M. A. Amer, "Low-Frequency Image Noise Removal Using White Noise Filter," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 3948-3952.

[3] Suppakitti Sopasoap. (2017). Facial Space and Object Detection Technique Development on Eye Boundary Using Image Processing. Thesis for the Degree of Master of Engineernig Program in Electronics and Telecommunication Engineering Rajamangala University of Technology Thanyaburi.

[4] Suzuki-Sakamaki M., Amemiya K. Development of high signal-to-background ratio depth-resolved soft X-ray absorption spectroscopy by fluorescence energy selection. Japanese Journal of Applied Physics 2018; (57): 57.

[5] P. Charoensawan, S. Phongsuphap and I. Shimizu, "Comparison of Fabric Color Naming Using RGB and HSV Color Models," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 1-5.

[6] K. Thenmozhi and U. S. Reddy, "Image processing techniques for insect shape detection in field crops," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 699-704.

[7] D. Chowdhury, S. K. Das, S. Nandy, A. Chakraborty, R. Goswami and A. Chakraborty, "An Atomic Technique For Removal Of Gaussian Noise From A Noisy Gray Scale Image Using LowPass-Convoluted Gaussian Filter," 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), Kolkata, India, 2019, pp. 1-6.

[8] Y. Liu, S. Shang, F. Guo and H. Yi, "Research on the edge detection of the cashmere and wool fibre images," 2009 9th International Conference on Electronic Measurement & Instruments, Beijing, 2009, pp. 426-428.

[9] Y. Yang, P. Yuhua and L. Zhaoguang, "A Fast Algorithm for YCbCr to RGB Conversion," in IEEE Transactions on Consumer Electronics, vol. 53, no. 4, pp. 1490-1493.

[10] Sutiwat Supaluk. (2011). Manhole Cover Detection on Road Surface Image Using Image Processing Technique. Thesis for the Degree of Master of Engineernig Program in Computer Engineering Department of Computer Engineering Faculty of Engineering Chulalongkorn University.

[11] L. Feng, L. Xiaoyu and C. Yi, "An efficient detection method for rare colored capsule based on RGB and HSV color space," 2014 IEEE International Conference on Granular Computing (GrC), Noboribetsu, 2014.

[12] X. Zhang, J. Jiang, Z. Liang and C. Liu, "Skin color enhancement based on favorite skin color in HSV color space," in IEEE Transactions on Consumer Electronics, vol. 56, no. 3, pp. 1789-1793.

[13] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala and C. O. Aigbavboa, "A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 92-99.

[14] Nutthisara Choolikhit. (2018). Edge Detection and Symmetry Method for Cleft Lip and Palate Children Using Image Processing. Thesis for the Degree of Master of Science in Information Technology Faculty of Science Khon Kaen University .

Comparison of Face Classification with Single and Multi-model base on CNN

Sarin Watcharabutsarakham
Image Processing and Understanding Research Team
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
sarin.wat@nectec.or.th

Supphachoke Suntiwichaya[1], Chanchai Junlouchai[2]
Leveraging Technology Solutions Section
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
supphachoke.suntiwichaya@nectec.or.th[1], chanchai.junlouchai@nectec.or.th[2]

Apichon Kitvimorat
Image Processing and Understanding Research Team
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
apichon.kitvimonrat@nectec.or.th

***Abstract*— Since the coronavirus disease 2019 (COVID-19) outbreak has spread across the country, our research applies to remind the people to wear a face mask when we go outside because a facial image detection and classification method will be used to authentication and authorization. This paper has shown that our created models based on CNN can detect the face mask-wearing, glasses-wearing, and gender with comparison two models. We training model with mix public datasets such as WIDER FACE, AFW, and MAFA. Moreover, we use VGG-Face to pre-train the model for the advance detection rate.**

***Keywords*— convolutional neural network, CNN, pre-training, classification, face classification**

I. INTRODUCTION

The current image classification techniques are used in extensive applications including, security features, face recognition, face verification, traffic identification, medical diagnosis, and other fields. The idea of image classification can be solved by different approaches[12]. Over the past few years, the increasing security is growing attention toward authentication based on voice, fingerprint, face and others. Face recognition is very interesting because it is useful for a wide number of real-time applications, such as surveillance, security systems, and access control. There are many works that were done in recent years, with many different approaches that have been proposed for a facial image. One topic, recognition methods, is trying to know the person and re-identify them [2-8]. The research problem in verification or re-identification[3] proposed deep discriminative representation learning (DDRL) for the unconstrained person re-identification. Moreover, [6] except for face recognition, such methods are often incapable of handling the open-set scenario of identifying new persons not included in the current set of known persons. Next, research problems from the source of the image such as [4] use to matching facial images captured from different sensors or sources with NIR-VIS to improve face recognition and [14] have additional depth images in the training data captured using depth cameras such as Kinect. In particular, we extract visual features and depth features from the RGB images and depth images. The next topic, geometry-feature-based methods, is to try to identify the position and relation between a part of the face such as eyes, nose, mouth, and shape, or size of the regions [9-10]. Once, COVID-19 was spread, the public health experts recommended universal mask-wearing, and some cities ordered residents to wear them under penalty of fine or imprisonment. In this real situation, the application may not undergo facial processes because the analysis algorithm to the face part has lost information. Anyway, COVID related application of computer vision, this one on detecting whether or not a person is wearing a face mask [9]. The research problem about detecting an object on the face such as [9] detect faces with occlusions is a challenging task due to two main reasons: the absence of large datasets of masked faces, and the absence of facial cues from the masked regions and [10] proposed method based on active appearance model (AAM) to remove eyeglasses from face images.

The purpose of this work is to propose a facial image classification method which consists of glasses and mask over the faces. So, the method appears suitable to be also applied in combination with this approach. It will focus on some parts of the face, such as eyes and mouth. Face detection is a critical technology, due to being applied in many fields such as authentication and authorization. In order to allow go to various locations, But due to the flu situation, general people have to wear a mask when they go outside. In this manner, many missing facial can be primarily recovered and exist in technology, not enough. Therefore need to create a new model based on the original model that is already effective facial detection technology. In this paper, we propose a method for facial classification based on features extracted with convolutional neural networks (CNN) and using the advantage of a pre-trained model in similar works.

II. RELATED THEORY

Recently, Almost of research is using deep learning because it becomes the first order of the state of art recognition algorithms. Especially, convolutional neural networks (CNN) have shown great potential in computation tasks. CNN use in the tasks of object recognition, tracking, classification, and face recognition. It has shown an excellent capability to solve complex problems and image classification tasks, which are impossible for human computation. The CNN creates a new image classification models which are much faster and more accurate than ever before, and they were applied in several objects [12-15]. Current CNN-based face detection methods divided into two categories. In the first category, CNN is used as a feature extractor in a traditional face detection framework to improve performance[17]. The second category, CNN, refers to the focus point on the face method, regards face detection as a particular case of generic object detection, and solves it using

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

CNN-based object detection algorithm relying on identifying faces in the image [16].

VGG-16 network, proposed by K. Simonyan and A. Zisserman from the University of Oxford, is a convolutional neural network architecture. It is compounded with 16 layers. Each layer consists of convolutional layers maximum pooling, or max pooling layers, activation layers, and connected layers with fully. VGG-16 is convolution network for classification and detection. Next, The VGG-Face CNN descriptors are implemented by the University of Oxford-based on the VGG-Very-Deep-16 CNN architecture that is trained on over 2 million celebrity images as described in [5]. Other technologies were applied in the face images such as Facenet, OpenFace, and DeepFace. Even, DeepFace was developed by Facebook for face recognition models as a good alternative to VGG-Face.

III. METHODOLOGY

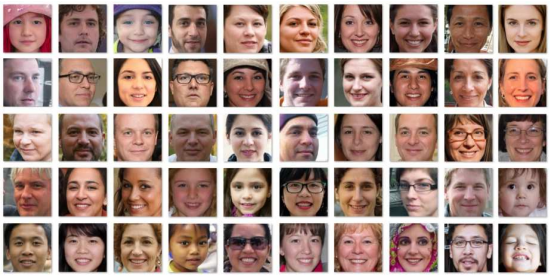
The algorithms used in this research based on the CNN architecture. We applied VGG-Face, the models are trained better and are able to identify different levels of image representation. The model will give the characteristic of face image such as glasses-wearing, a mask-wearing, or gender.

TABLE I. TRAINING DATASET

Group	Data Types	#no. Images
1	Glasses wearing	8,900
2	No Glasses wearing	8,900
3	Face Mask	2,200
4	No Mask wearing	2,200
5	Man	27,000
6	Woman	27,000

A. Data Collection

The public datasets used in the paper are WIDER FACE [19], Annotated Face in-the-Wild (AFW) [18], and MAFA [9]. To demonstrate their proposed method achieves state-of-the-art results. We mixed the dataset from MAFA, WIDER FACE, and AFW for training. Moreover, Deep learning needs to be trained with a huge training data set to achieve satisfactory performance. The public datasets usually contain limited images. Training dataset, the positive images are generally fewer than negative images. So, data augmentation is better to boost the performance and a widely used compensate in deep learning [20]. So, augmented data will be making many images to achieve data balancing. Data augmentation is widely used



a) The face data

b) The face mask data

Fig. 1. Example of Training dataset

TABLE II. PROPOSED CNN ARCHITECTURE

Layer type	Parameters
Input Layer	224x224 RGB image
Convolution	#64 224x224
Convolution	#64 224x224
Max Pooling	#64 2x2
Convolution	#128 112x112
Convolution	#128 112x112
Max Pooling	#128 2x2
Convolution	#256 56x56
Convolution	#256 56x56
Convolution	#256 56x56
Max Pooling	#256 2x2
Convolution	#512 28x28
Convolution	#512 28x28
Convolution	#512 28x28
Max Pooling	#512 2x2
Convolution	#512 14x14
Convolution	#512 14x14
Convolution	#512 14x14
Max Pooling	#512 2x2
Avg. Pooling	#512 1x1
Flatten	#512 1x1
Dense (Relu)	#4096 1x1
Dense (Relu)	#4096 1x1

to compensate for deep learning. Deep learning needs to be trained with a huge training data set to achieve satisfactory performance. We use the data augmentation to boost the

performance when training the deep network. We divided the image training set into six sets in our experimental detail in Table I.

The training set, we crop the input images into 224 × 224 pixels and horizontally flip them around the y-axis. To achieve data balancing, we sample the same number of positive images and negative images in each classes when start the training process.

B. Network architecture and Training

The overall framework is shown in Fig. 1. The framework is composed of two steps of neural networks. Firstly, pre-trained, this model start with the smaller networks converged and then used as initializations for the larger and deeper networks. The goal is to find the parameters of the network that minimize the average prediction loss value after the softmax layer. We applied VGG-16 and used VGG-Face structure model in pre-training step. Secondly, our layer, the classifier will have 2 convolution layers, 1 max pooling layers, 1 flattening layer and finally an output layer with Adam optimizer. The 33,605,442 parameters are trained in our proposed model. Our implementation is based on the python program with the NVIDIA CuDNN libraries to accelerate training. All our experiments were carried on NVIDIA 2070 RTX GPUs with 6GB of onboard memory.

IV. Experiments and Results

To allow for a direct comparison to previous work, while our CNNs are trained on the mixed dataset in the section before. We design two experimental groups to measure the accuracy rate, whether one model or multi-model is better for face classification with our model structure and unbalanced data. To obtain all the training data's average responses, we put all training data into the fine-turned VGG-face. Setting X= [x1, x2, x3, . . . , xn] denotes the image of training data, n is the number of training data. This method for fine-tuning the pre-trained VGG-face.

Fig. 2. Hierarachy Classification

A. First Experimental

The experiment is to try to classify the face with a hierarchy structure, as shown in Fig 2. The dataset is consists of 4 groups of face images.

-The image of 10,000 faces with a face mask or glasses and 10,000 regular faces.

-The image of 8,900 faces with glasses and 8,900 faces without glasses.

-The image of 2,200 faces with a face mask and 2,200 faces without a mask.

-The image of 27000 man faces and 27000 woman faces. We start with the pre-train of VGG-face and add three of the last layers to parameters training. This classifier will have one flattening layer, one max pooling layer, two convolution layers, and finally, an output layer with Adam optimizer. In the following, we improved the last three layers in the proposed models for increased accuracy. The input image will classify gender with first and compute the confidence value in $Score_i$. Next, the face image will classify mask-wearing and compute the confidence value in $Score_j$. Last, the face image will classify glasses-wearing and compute the confidence value in $Score_k$. Finally, the accuracy rate of testing is shown in Table III.

Condition for group classification
If found face in the image
1) If the value of $Score_i < 0.5$ and the value of $Score_j < 0.5$ and the value of $Score_k < 0.5$
classified in a confused group
2) If the value of man > 0.9
classified in a man group
or the value of woman > 0.9
classified in a woman group
or classified in a none group
3) If value of mask > 0.8
classified in a mask group
4) If value of Glass > 0.8
classified in a glass group

Fig. 3. Pseudo code

When we use the 4 models to predict the face, we select the condition and threshold by the value of data training. First, If the confidence value of gender, mask, and glasses was less than 0.5, the images were rejected into the group. Other criteria to classify groups by the confidence value are shown in Fig 4. To evaluation, we classified the face images into 6 classes as Table 1 (glasses-wearing, mask-wearing, man, woman). We use 4 trained models to classify like a hierarchy structure. In the first experiment, we trained four models with 96,000 face images to classify into four groups. The evaluation is shown that the models get an average 94.99% accuracy rate.

TABLE III. Result of Models

	Accuracy Rate (%)	
Model	**Training**	**Testing**
M1	99.44	98.58
M2	98.54	91.12
M3	97.59	96.13
M4	97.18	94.12

M1 is a trained model for classification of the face image that wears a face mask or glasses and a normal face. M2 is a trained model for face image classification to a face mask wearing and non-face mask-wearing. M3 is a trained model for face image classification to glasses and no glasses. M4 is a trained model for gender classification by face images. The accuracy rate of the experiment is shown in Table III.

B. *Second Experimemtal*

The experiment uses a trained model to classify the face into four classes, such as glasses-wearing (S1), mask-wearing (S2), man (S3), and woman (S4). The dataset uses the same as section A. We use to train with the pre-train of VGG-Face. The accuracy rate of training is 85.5%.

.

TABLE IV. CONFUSION METRIC OF RESULT

Actual	Predicted			
	S1	S2	S3	S4
S1	0.75	0.18	0.05	0.02
S2	0.01	0.90	0.01	0.08
S3	0	0	0.95	0.05
S4	0	0.10	0.10	0.80

We use our proposed architecture, explained in section A, for classification face images by only single model. In the second experiment, we trained a model with 6,000 face images to classify into four groups with balance data in every group. The evaluation is shown that the model gets an average 85.5% accuracy rate. We use the ReLu function to solve the vanishing gradient problem because it is suitable for balance data. This function does not have an asymptotic upper and lower bound. Thus, the earliest layer can receive the last layers' errors to adjust all weights between layers. By contrast, a traditional activation function like sigmoid is restricted between 0 and 1, so the errors become small for the first hidden layer. This scenario will lead to a poorly trained neural network. The experimental results showed that the model trained with many images would give an advantage. Although we use the image augmentation to make data balancing, the real face images are the best for training.

V. CONCLUSION

The comparison of single and multi-model are shown that the dataset in this experiment ware cross the group in some case so the multi-model will be loss performance to predict in the face image by our proposed model architecture was shown that the face images could give information for deep learning network to predict the objective setting. The result of this paper shows that the multi-model get high accuracy, but on the other side that we have to conscious is the space in memory of the computer. If we use a single model to classify, it uses a resource less than the multi-model. Fig 5 shows a comparison of the accuracy rate in each group. Moreover, face image was applied in many areas because pre-train VGG-Face helps the suitable tuning parameters by million face images. This advantage model structure can be helpful for researchers to create a new model of about-face images. Notably, our research uses to detect the face in the coronavirus disease 2019 (COVID-19) outbreak event to remind the people to wear a face mask when going outside or inside the public places. The compare results were shown in Fig. 4. We use a multi-model in section A and a single model in section B to classify face into four groups as a normal face, glasses-wearing, mask-wearing, and gender. The accuracy rate for classifies normal face (C1) is 75% with a single-model and 98% with multi-models. The accuracy rate for classifies the glasses-wearing group (C2) is 90% with a single-model and 91% with multi-models. The accuracy rate for classify mask-wearing group (C3) is 95% with a single-model and 96% with multi-models. The accuracy rate for classifying gender (C4) is 80% with a single-model and 94% with multi-models. In future work, the facial will classify a particular person because deep learning makes amazing in recognition applications.

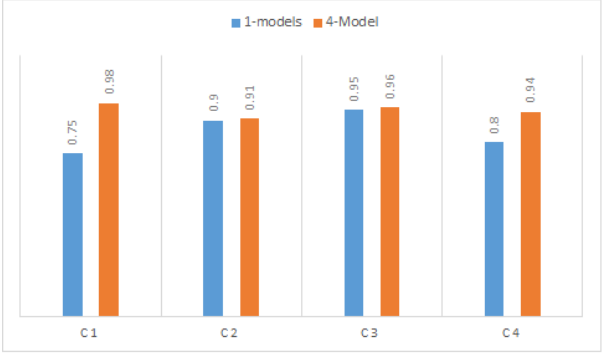


Fig. 4. The result of prediction with single and multi-model

ACKNOWLEDGMENT

We are thankful for public data distributors, which all use in this paper. Moreover, we thank our team in Artificial Intelligent Research Team which provided insight and expertise that greatly assisted the research, although the infrastructure may not be enough with all of the resources for this paper.

REFERENCES

[1] X. Mengyu, T. Zhenmin, Y. Zhenmin, Y. Lingxiang, L. Lingxiang and X. Jingsong ,"Deep Learning for Person Reidentification Using Support Vector Machines," Adv. Multim. 2017, doi:10.1155/2017/9874345.

[2] G. Artur , K. Marcin and, P. Norbert," Face re-identification in thermal infrared spectrum based on ThermalFaceNet neural network," International Microwave and Radar Conference (MIKON), May 14-17, 2018, doi: 10.23919/MIKON.2018.8405170

[3] J. Yu, D. Ko, H. Moon and M. Jeon, "Deep Discriminative Representation Learning for Face Verification and Person Re-Identification on Unconstrained Condition," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 1658-166z2., doi: 10.1109/ICIP.2018.8451494

[4] C. Peng, N. Wang, J. Li and X. Gao, "Re-Ranking High-Dimensional Deep Local Representation for NIR-VIS Face Recognition," in IEEE Transactions on Image Processing, vol. 28, no. 9, pp. 4553-4565, Sep, 2019, doi: 10.1109/TIP.2019.2912360

[5] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition," British Machine Vision Conference,2015.

[6] V. Madhawa, S. Imesha,K. Pasindu,V. Jayan and P. Indika, " Open Set Person Re-identification Framework on Closed Set Re-Id Systems," International Conference on Signal and Image Processing (ICSIP)," AUG 4-6, 2017, doi: 10.1109/SIPROCESS.2017.8124507

[7] S. Kanchanapreechakorn and W. Kusakunniran, "Robust human re-identification using mean shape analysis of face images," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, 2017, pp. 901-905., doi: 10.1109/TENCON.2017.8227986

[8] G. Agus and W. H. Dwi, "Key Frame Extraction with Face Biometric Features in Multi-shot Human Re-identification System," International Conference on Advanced Computer Science and

information Systems (ICACSIS), OCT 12-13, 2019, doi: 10.1109/ICACSIS47736.2019.8979799

[9] S. Ge, J. Li, Q. Ye and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-4, doi: 10.1109/CVPR.2017.53

[10] Y. Wang, J. Jang, L. Tsai and K. Fan, "Improvement of Face Recognition by Eyeglass Removal," 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, 2010, pp. 228-23, doi: 10.1109/IIHMSP.2010.64

[11] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220

[12] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 2319-2323, doi: 10.1109/RTEICT42901.2018.9012507.

[13] X. Xu, W. Li, and D. Xu, "Distance Metric Learning Using Privileged Information for Face Verification and Person Re-Identification," IEEE Transactions on Neural Networks and Learning Systems, vol. 26 , 2015, pp. 3150 - 3162.

[14] Y. Chen, and A. Baskurt, Person re-identification in images with deep learning, 2018.

[15] C.W. Ngo, and et al.,"Deep Learning for Person Reidentification Using Support Vector Machines," Advances in Multimedia, 2017, doi: doi.org/10.1155/2017/9874345.

[16] A. Jourabloo, M. Ye, X. Liu and L. Ren, "Pose-Invariant Face Alignment with a Single CNN," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 3219-3228, doi: 10.1109/ICCV.2017.347.

[17] M. Nakada, H. Wang and D. Terzopoulos, "AcFR: Active Face Recognition Using Convolutional Neural Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 35-40, doi: 10.1109/CVPRW.2017.11.

[18] X. Zhu and D. Ramanan. "Face detection, pose estimation,and landmark localization in the wild," In IEEE Conference on Computer Vision and Pattern Recognition, pages 2879-2886, June 2012.

[19] Y. Ping L. Chen, C. Loy and, X. Tang,"WIDER FACE: A Face Detection Benchmark," 2015, CoRR abs/1511.06523

[20] Mei Wang and Weihong Deng, "Deep Face Recognition: A Survey",School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China,2019.

Model proposed to cost reduction in printing

Noppon Mingmuang, Chaleedol Inyasri, Wirote Jongchanachavawat
Faculty of Engineering and
Industrial Technology
Phetchaburi Rajabhat University
Phetchaburi, Thailand
jongwirote@gmail.com

***Abstract*—** This paper illustrates the benefit of model proposed in order to reduce cost and improve toner replenishment process. The proposed design printers will be reduced from 133 devices to 80 devices (the existing devices will be 40 and the new devices will be 40) to cover all area services. The proposed results have proven in reducing the cost from 112,688 Baht/month to 91,328 Baht/month or decreasing cost 21,360 Baht/month (decreasing 19%). The proposed re-design toner replenishment process to be better. In addition, the toner replenishment task will be transferred from IT to suppliers. Finally, this work proved to improve the process of inbound fax from 2 steps to 1step and improve the process of outbound fax from 3 steps to 1step. The indirect benefit of the proposed design was the new model of printer will have more security by smart card or key-code and employee could print from anywhere with the proposed printer.

Keywords—cost reduction, printing, model proposed

1. Introduction

In today all the organizations aim to find ways to reduce costs in the business, production or business processes.

Reducing production cost can significantly improve performance and competitiveness for a manufacturer and it is a common action for manufacturer[1]-[2].

Reducing production costs were studied with information systems for decision making[3].

Therefore the studies are in the pooling principle to maximize the use of available resources[4].Like a car pool, you use one resource (one car) to do the same job that would otherwise require two resources (two cars).This reduces your cost to work but may also cause you some inconvenience.

There are many studies on the improving new models of printers. The printers can communicate with each other. So a program has been developed to support these activities, to enable users to share printers efficiently[5]-[9]. Also some researchers study the pooling of human resources for the effective works[10], for instance "The studies on how to effectively reduce the cost of using the printer in the office to be lower cost" [11],[12].

The cost of printer is one of the items that considered to be the highest cost in the administration tasks.

As a result, this study has been proposed a model to reduce cost in printing for the cost-effective in the university.

2. Methodology

The research methodology flowchart is shown in Figure 1, It can be divided into five steps:

1) Survey and Interview
2) Data Analysis
3) Proposed Design
4) Proposed Printer Layout
5) Proposed Result

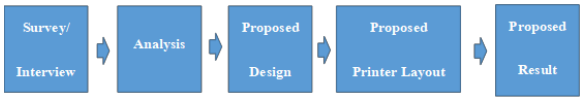


Figure 1 Methodology flowchart.

This five steps method involves survey/interview, data analysis, design, implementation, and result.

Survey/Interview

Explore the printers that are available in all areas of the university. Interview relevant users in each area to collect printer usage information.

Data Analysis

Analyze the collected data from each area of the university for the current cost of printing in each printer category in order to prepare for designing the suitable model in each area.

Proposed Design

Design and allocate suitable printers in each area by the condition on performance of printer must better than before and the cost must be lower than the current cost.

Proposed Printer Layout

Implement printer pooling as the proposed design by selecting the appropriated devices to serve their capabilities in each area.

Proposed Result

Calculate simulated expenses that will be incurred by using data from collection in each area. Calculate the simulated results that will occur with the current results in order to present the simulated results to the university administrators for acknowledgment.

2.1 Survey/Interview

Researcher will survey & interview 150staffs obtained by a purposive sampling covering 5 Buildings at Krirk University.

2.2Analysis

Fact Finding

After analyzing the issues, we have encountered major problems. In terms of the cost as the following.

1. The ratio of machines to employees was 1:1.13 as in Table I, which means the rates of the machines and the number of employees were in close proportion. Almost one machine per one employee considered cost management in the matter of the printer was ineffective.
2. Most printers were older than 5.33 years as in Table II. Usually, machines over the age of 3 years tend to have maintenance problems and more expensive to maintain.
3. There were 92 different models in 13 brands in use as in Table III, which make managing ink in each model problematic and oblige spare ink at a cost. That showed the inefficiency in the cost management.

Table I Employee device ratio

Employee	Qty. Printer	Employee Device Ratio
150 persons	133	1.13

Table II The asset age in Printers.

Category	Age
MFP	N/A
Copier	N/A
AIO	4.17
Printer	5.83
Fax	5.75
Scanner	4.17
Total	5.33

Table III 13 Brands and 92 models of all printers.

Product	Brand	No. of Model	Number
MFP	Fuji Xerox	1	2
	Brother	1	1
	HP	2	2
Copier	Canon	4	4
	Sharp	1	1
	Minolta	1	1
AIO	Brother	2	3
	Canon	5	6
	Epson	6	7
	HP	12	15
	Lexmark	1	1
	Samsung	1	1
Printer	HP	23	39
	Samsung	3	5
	Epson	3	8
	Canon	6	13
	Kyocera	1	1
	Fuji Xerox	1	1
Fax	Sharp	7	8
	Brother	2	2
	Panasonic	4	4
Scanner	HP	1	1
	Epson	2	3
Copy-print	Gestetner	1	2
	Ricoh	1	2
Total	13	92	133

We classified 7 types of printers among 133 printers as in Fig.2, which found that most printers ranking from one to three were Printer, All in One (AIO) and Fax.

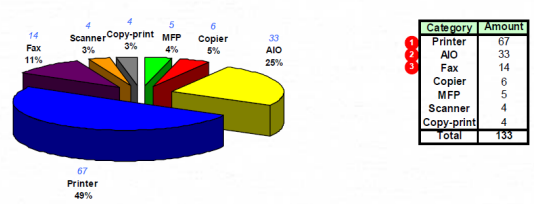


Figure 2 Category of printers

We collected the data of the number of printed paper in each room of all 5 buildings at Krirk University and we assumed the cost calculation as the following.

1.Waste paper about 2%

2. The toner and the hardware cost based on a market price.

3. The printers without paper recording and the toner recording assumed using 1 tank/18 month.

4. An average paper price equal to 88 Baht/ream.

5. The printers without the installation date, we use the launch date.

6. All printers were 5 years of depreciation.

Table IV The detail of Monthly cost per building & room

Location	Floor	Room	Unit	TDV	Direct Cost				Indirect Cost			Total Cost
					Consumable	Hardware	S&M	Total	Paper	Electricity	Total	
Building1	1	Graduation Evaluation	3	241	238	0	250	488	43	33	76	565
Building1	1	Registration	4	2,326	1,534	512	375	2,421	419	114	533	2,953
Building1	1	Office of Krirk University Library	3	900	980	0	125	1,105	162	65	227	1,332
Building1	1	Document/Book	7	122,225	9,982	9,982	9,982	29,945	22,001	472	22,473	52,418
Building1	2	Project Phd.	4	4,067	3,898	513	0	4,411	732	32	764	5,175
Building1	2	Administration	5	1,088	1,578	75	0	1,653	196	120	316	1,969
Building1	2	Accounting/Financial	1	444	467	0	0	467	80	30	110	577
Building1	2	Journal	1	19	27	0	0	27	3	14	17	44
Building1	2	Office of Education Quality Assurance	2	448	969	100	125	1,194	81	23	104	1,298
Building1	2	Office of Academic	1	267	280	0	125	405	48	16	64	469
Building1	3	Computer Laboratory	9	2,423	2,243	664	0	2,907	436	110	546	3,453
Building1	3	Lecturer Computer	2	1,167	914	0	250	1,164	210	61	271	1,435
Building1	3	Office of IT Center	3	410	435	0	125	560	74	37	111	671
Building1	3	IT Head Center	1	667	613	0	125	738	120	35	155	893
Building1	4	Office of Information Technology and Management	1	500	605	0	0	605	90	30	120	725
Building1	4	Lecturer Master Degree	1	500	658	0	125	783	90	9	99	882
Building2	4	Lecturer IT	1	33	83	0	0	83	6	14	20	103
Building2	1	Admission	2	214	136	0	125	261	39	41	79	341
Building2	2	Public Relations	2	442	2,093	432	125	3,550	79	32	142	3,692
Building2	3	Project Political Communication College	3	267	210	0	0	210	48	30	78	287
Building2	4	Project Political Communication Saturday Sunday	2	2,573	2,833	198	125	3,845	463	83	546	3,592
Building2	5	Lecturer Art	3	438	1,428	0	0	1,428	79	24	103	1,531

Location	Floor	Room	Unit	TDV	Direct Cost				Indirect Cost			Total Cost
					Consumable	Hardware	S&M	Total	Paper	Electricity	Total	
Building3	1	Cafeteria	1	152	521	0	0	521	27	14	41	562
Building3	2	Office of Management Project course	3	388	702	162	0	864	70	141	211	1,075
Building3	3	Office of Student Affairs	2	60	95	60	0	155	11	28	38	193
Building3	3	Office of Academic Affairs	7	2,895	2,677	760	125	3,562	595	80	585	4,147
Building3	3	Support Center/Service	1	60	107	0	0	107	11	14	25	131
Building3	4	Sport	3	120	200	0	0	200	22	28	49	249
Building4	2	Living	1	367	59	0	0	59	66	62	128	187
Building4	2	Admin	1	81	104	0	125	229	15	9	24	253
Building4	2	Dean of Communication Art	2	882	746	123	0	869	159	35	194	1,063
Building4	2	Head of Communication Art	1	183	180	265	0	445	33	14	47	492
Building4	3	Lecturer Art	5	1,949	1,895	0	0	1,895	351	154	505	2,400
Building4	3	Secretary Art	4	960	1,276	0	0	1,276	173	53	226	1,502
Building4	4	Lecturer Politics	2	458	450	0	0	450	83	95	178	628
Building4	4	Lecturer General/Chinese	4	592	613	0	0	613	107	137	244	857
Building5	1	Krirk University Language Institute	1	1,217	1,488	215	0	1,703	219	120	339	2,042
Building5	1	Lecturer	1	7	2	0	0	2	1	14	15	17
Building5	1	Research and Development Center	2	202	281	0	125	406	36	32	68	474
Building5	3	Director	4	1,952	2,549	0	125	2,674	351	55	406	3,080
Building5	4	Dean of Law	2	667	821	0	0	821	120	90	210	1,032
Building5	4	Lecturer law	2	741	685	0	125	810	133	9	143	953
Building5	4	Lecturer business	5	1,721	2,704	96	250	3,050	310	90	400	3,450
Building5	4	Lecturer Economics	3	286	208	0	125	333	51	130	182	514
Building5	4	Meeting Room	2	535	672	0	0	672	96	10	106	778
Building5	4	Lecturer Financial/Bank/Logistics	1	14	66	50	0	116	2	5	8	123
Building5	4	Lecturer Human Resource	1	770	511	0	125	636	139	20	159	795
Building5	4	Lecturer Marketing	2	333	327	0	125	452	60	9	70	521
Building5	4	Lecturer Accounting	3	427	459	72	125	655	77	64	141	797
Total			133	159,586	53,480	14,187	13,357	81,024	28,726	2,938	31,664	112,688

From table IV shows that the monthly document cost was about 112,688 Baht. The cost per employee was 751 Baht/user/month.

The number of documents classified by the category of the printer as in Fig.3.

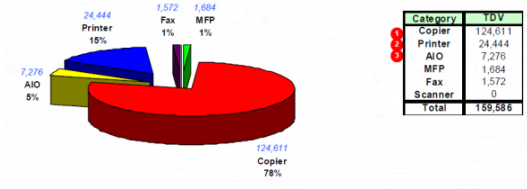
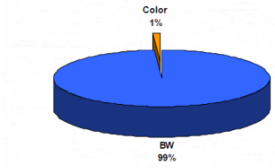


Figure 3 Classify total document volume by the printer categories.

From Fig.3 shows the most document volume of top 3 rank were the Copier, the Printer and the All in One (AIO).

In addition, the total document volume (TDV) classified by color as in Fig. 4.



Category	TDV
BW	157,851
Color	1,735
Total	159,586

Figure 4 Classified color of documents printing

From Fig.4 we found that most of the cost of printings was in black and white (99%). The other cost was in color (1%).

Then, a direct cost and an indirect cost could be distributed as in figure 5. The direct cost was 72% and the indirect cost was 28%. The direct cost included a hardware, a consumable product, and service & maintenance. The indirect cost consisted of the paper and an electricity.

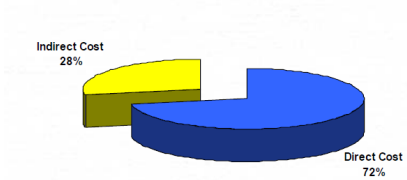


Figure 5 Direct Cost & Indirect Cost

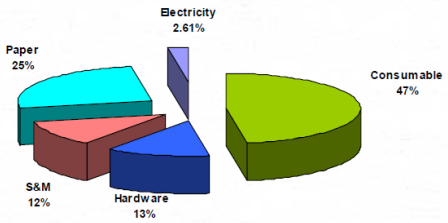


Figure 6 Classified Cost by Category

The Cost could be allocated according to category. With the top 3 expenses were a consumable product cost (47%), a paper cost (25%) and a hardware cost (13%) as shown in Fig.6.

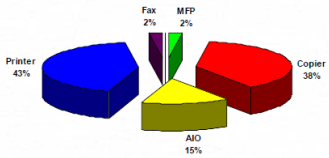


Figure 7 The direct cost 81,024 Baht/month classified by printer category

The direct expense was 81,024 baht per month according to the category of printer shown in Figure 7. The top3 were the printer (43%), the copier (38%) and the AIO (15%) respectively

Therefore, when taking the direct expenses in 3 major cost categories, namely, the hardware cost, the consumable product cost and the service & maintenance cost, they are shown as in Fig. 8, 9, and 10 respectively.

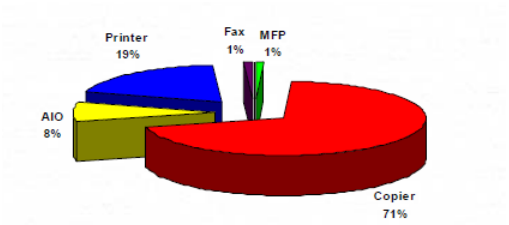


Figure 8 Classified direct cost by Hardware cost

From Fig.8 showed that the cost of hardware for the top 3 were the copier(71%), the printer(19%) and the AIO(8%).

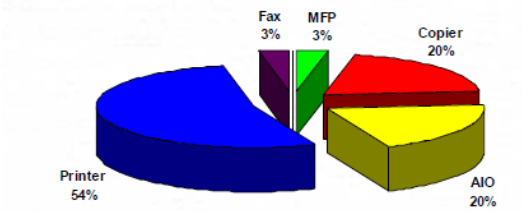


Figure 9 Classified direct cost by the consumable product cost

From Fig.9 showed that the cost of the consumable product cost for the top 3 were the printer(54%), the AIO(20%) and the copier(20%).

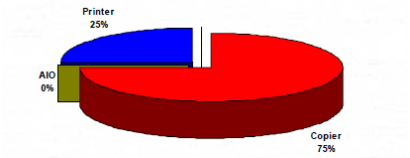


Figure 10 Classified direct cost by service & maintenance cost

From Fig.10 showed that the cost of service and maintenance for the top 2 were the copier(75%) and the printer(25%).

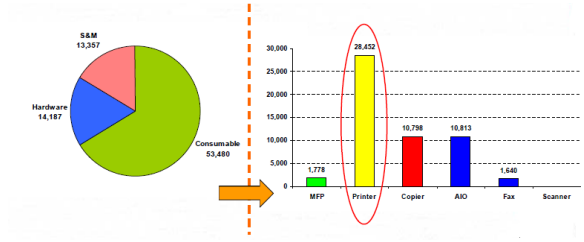


Figure 11 Classified the consumable product cost by the printer category

From Fig.11 showed that the cost of the consumable product. The top 3 were the printer (28,452 Baht/month), the AIO (10,813 Baht/month) and the copier(10,798 Baht/month) respectively.

Therefore, we can calculate the cost per page by TDV, direct cost and cost per in Table V

Table V Cost per page classified by printer category

Category	Unit	TDV	Direct Cost	Cost per Page	
		Total	Total	BW	Color
MFP	5	1,684	1,902	1.11	1.95
Copier	6	124,611	30,761	0.25	
AIO	33	7,276	11,972	1.32	3.56
Printer	67	24,444	34,583	1.17	10.71
Fax	14	1,572	1,806	1.15	
Scaner	4	0	0	N/A	N/A
Copy-Print	4	0	0	N/A	N/A
Total	133	159,586	81,024	0.45	6.09
If Excluded Copier		34,976		1.19	6.09

From Table V, the cost per page was 0.45 Baht (BW: Black &White) and 6.09 Baht (Color). In case of without the copier, the cost per page was 1.19 Baht (BW) and 6.09 Baht (Color). All of the cost per page excluded the indirect cost but their cost was enough to calculate the fundamental cost per page.

2.3 Proposed Design

Based on our finding, we proposed a design by condition as the following.

1. Price per page (Market price) were 0.45 Baht (BW) and 4 Baht (Color) after negotiating with suppliers. This price can reduce the cost of printing in the university.
2. Reducing the resistance from employee, we will use the existing devices mixed with the new device and still cover all area as the beginning.
3. Setup the printer policy and announce to all employee that if the existing devices could not print, then the employee need to use the printer pooling at the closest area without purchasing a new devices or a new toners.

Table VI The existing & The new printer devices

The current printer devices	The new printer devices	
133	80	
	The current devices	The new devices
	40	40

From Table VI the proposed design would reduce the existing devices from 133 to 80 devices by using the current 40 devices and the new 40 printers. We design the current and proposed devices in 5 buildings as the Table VII.

Table VII The existing devices & proposed devices

No.	Building	Floor	The existing	Printer Category	Proposed	Printer Category
1	1	1	17	AIO=1, Printer=6, Copy-print=4, Copier =4, Scanner=2	13	AIO=1,MFP=2, Printer=2, Copy-print=4, Copier=3, Scanner=1
2	1	2	14	AIO=5, Printer=6, Fax=2, Scanner=1	9	AIO=3, MFP=5, Printer=1
3	1	3	15	AIO=8, MFP=1, Printer=6	6	MFP=4, Printer=2
4	1	4	3	MFP=2, Printer=1	3	AIO=1, MFP=2
5	2	1	2	Printer=1, Fax=1	2	Printer=1, Fax=1
6	2	2	3	AIO=1, Printer=1, Scanner=1	2	MFP=1, AIO=1
7	2	3	3	AIO=1, Printer=1, Fax=1	3	AIO=1, Printer=1, Fax=1
8	2	4	7	AIO=1, Printer=3, Fax=2, Copier=1	4	AIO=1, MFP=1, Printer=2
9	2	5	3	AIO=1, Printer=2	3	AIO=1, Printer=2
10	3	1	1	AIO=1	1	AIO=1
11	3	2	3	AIO=1, Printer=1, Fax=1	2	Printer=1, Fax=1
12	3	3	10	AIO=4, Printer=4, Fax=2	6	AIO=3, Printer=1, MFP=1, Fax=1
13	3	4	3	AIO=3	2	AIO=2
14	4	2	5	AIO=1, Printer=2, Fax=1, Copier=1	5	AIO=1, Printer=2, Fax=1, Copier=1
15	4	3	9	AIO=1, Printer=6, Fax=1, MFP=1	3	Printer=1, MFP=1, Fax=1
16	4	4	7	Printer=7	2	Printer=2
17	5	1	4	AIO=2, Printer=1, Fax=1	4	AIO=1, Printer=1, MFP=1, Fax=1
18	5	3	4	AIO=2, Printer=1, Fax=1	3	AIO=1, MFP=2
19	5	4	21	AIO=6, Printer=11, MFP=3, Fax=1	13	AIO=4, Printer=3, MFP=6

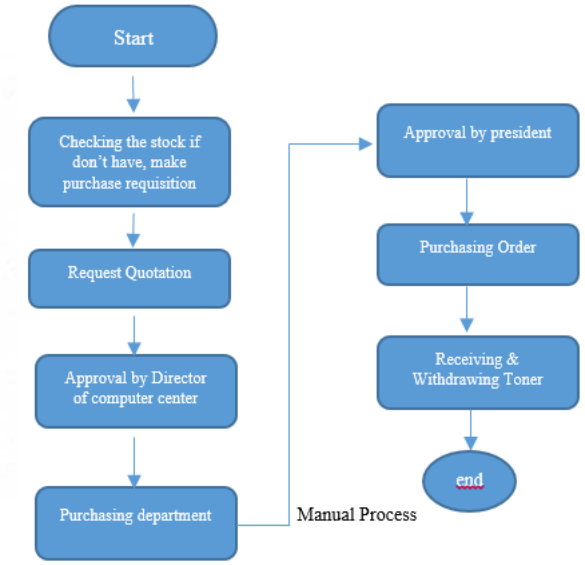


Figure 12 The current Toner Replenishment process

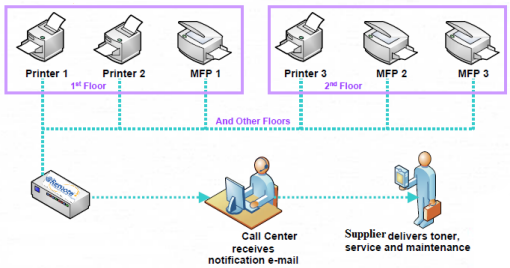


Figure 13 The new toner replenishment process

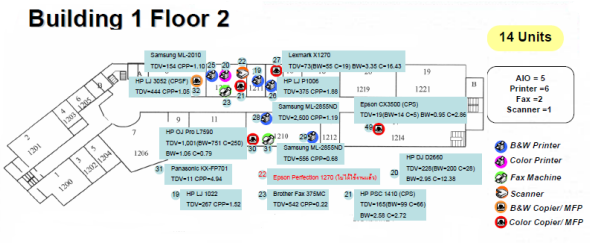
From Fig. 12 & 13, our new design can reduce process time of toner replenishment.

Proposed Printer Layout

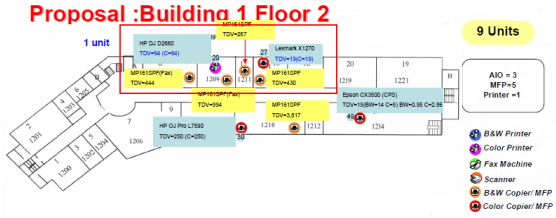
We used the printer of 2 model because their specifications have enough capability as the following.

1. Ricoh MP161SP for BW printing.
2. Ricoh CL4000DN for BW and color printing.

We will install the driver for all area and setup one server for control the printer networking. We can monitor all printers' activity and we will use @remote application from Ricoh to implement.



a) The existing devices layout



b) The new devices layout

Figure 14 Devices layout implementation at building 1 floor 2

From Fig. 14a and 14b display the one example devices layout implementation at building 1 floor 2.

3.Proposed Result

In this section, The proposed result of this work shows in cost reduction in Fig.14.

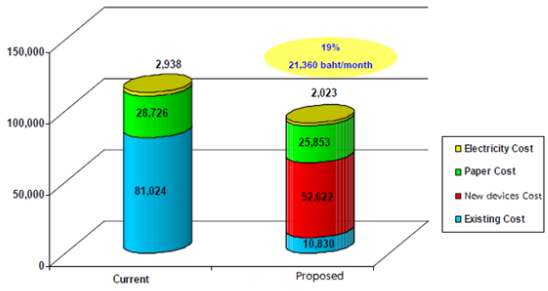


Figure 15 The cost of printing before and proposed

From Fig.15, the current cost of printing were classified in 3 items.

Current cost

1.The existing cost was 81,024 Baht/month
2.The paper cost was 28,726 Baht/month
3.The electricity cost was 2,938 Baht/month
Total Cost was 112,688 Baht/month

Proposed cost

1.The existing cost will be 10,830 Baht/month
2.The new devices cost will be 52,622 Baht/month
3.The paper cost will be 25,853 Baht/month
4.The electricity cost will be 2,023 Baht/month
Total Cost will be 91,328 Baht/month

The proposed result can reduce the cost 21,360 Baht/month or decreasing 19% per month.

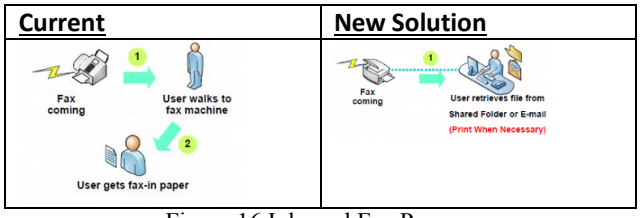


Figure 16 Inbound Fax Process

From Fig.16, reducing from 2 steps to only 1 step and in addition, reduce lost document, reduce printing cost and receive soft copy document.

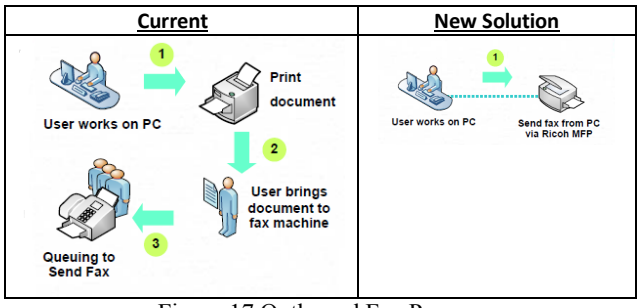


Figure 17 Outbound Fax Process

From Fig.17 , reducing from 3 steps to only 1 step. In addition, the printing cost reduction because no need to print document out, as a result, increasing user's satisfaction because the speed of service is quicker.

4.Conclusion/Future Research

This proposed conceptual model for printer pooling adoption in Krirk University proved the proposed 3 aspects as the following.

1.Reducing the cost of printing about 21,360 Baht/month (Decreasing 19%)

2. Improving the toner replenishment process from 7 days to 0 day and toner replenishment task will be transferred to supplier.

3. Improve performance of inbound & outbound fax per time as the following.

- Inbound fax per time has improved from 2 steps to 1 step. The performance of improvement is 50%.
- Outbound fax per time has improved from 3steps to 1 step. The performance of improvement is 67%.

The indirect benefit of proposed design is the new model of printer will have more security by smart card or key-code and employee could print in anywhere with the proposed printer.

Future Research: Implementing the printer pooling proposed and comparing the printer cost before and after implementation. Also comparing IT service level before and after implementation by survey users' satisfaction.

Acknowledgement

This work had the financial support by Krirk University in Thailand. Ricoh gave the specification of printer and consulting.

References

1.Porter, M.E., 1998, **Competitive Advantage: Creating and Sustaining Superior Performance.** Free Press, New York

2.M.A. Hannan, et al. 2020, **Waste collection route optimization model for linking cost saving and emission reduction to achieve sustainable development goals,** International Journal of Sustainable Cities and Society, Pre-proof.

3.Xiaojie Sun, et al., 2019, **Manufacturer encroachment with production cost reduction under a symmetric information.** International Journal of Transportation Research Part E, 128:191-211

4.Kyle Cattani and Glen M. Schmidt, 2015. **The Pooling Principle.** Informs Transactions on Education. Vol.5, No.2, pp.17-24

5.Alan Hlava., **Print Model that allows direct job submission to physical printer objects wile preserving job defaulting, validation and scheduling,** US Patent No. US6,614,549B1, Sep.2, 2003.

6.Waller, et al.,2020, **Document Pooling Mechanism**, US Patent.20200133602A1, Apr.30, 2020

7.Nathan Taylor,2008, **" Cutting printer costs" ,** IT Technical Writer, Kyocera

8.Shibusawa, et al.,2000, **Printer managing apparatus, printer system and printer setting method,** US Patent No. 6,088,120, Jul.11, 2000

9.Waller, et al.,2020, **Document Pooling Mechanism**, US Patent.20200133602A1, Apr.30, 2020

10.Wallace, Whitt. 2004, **Resource pooling and staffing in call centers with skill-based routing.**Working paper available at http://www.columbia.edu/~ww2040/recent.html

11.Chris Poole, 2018, **Major Costs in Printing and How to Reduce Them,** Printer Logic

12.I.Jack Gerard, " What Is Printer Pooling?" , https://smallbusiness.chron.com/printer-pooling-40366.html

.

COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective

Abir Abdullha
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
abir15-7773@diu.edu.bd

Sheikh Abujar
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sheikh.cse@diu.edu.bd

***Abstract*—Most of the countries are now affected by COVID19, COVID-19 is now the name of the biggest problem in the world. Bangladesh is also affected by COVID-19. The whole country is facing this virus as the biggest problem. So try to analyze the data day by day to understand the situation. We also try to use some model, algorithm, logic, analysis to find the solution to this current situation. We are also using some machine learning algorithms to predict the future situation. Machine learning supervised are Linear Regression Model and k-nearest neighbors (KNN) Algorithms. There are different types of data sets and algorithms. We have tried to explain these well.**

***Index Terms*—COVID-19, Data Visualization, Supervised Algorithms, Regression Model, k-nearest neighbors (KNN) Algorithm, Prediction and analysis**

I. Introduction

COVID-19 is now the name of the biggest panic. Nowadays on TV, newspaper, online, social media and every people think of this virus in their mouths and this virus makes people panic, why not, all the countries of the world become helpless today with this virus which name is COVID 19 caused by Nobel coronavirus. The whole world is stunned by this epidemic virus. The big countries are struggling today because of the coronavirus. The biggest force in the coronavirus is the human transmission. The most common symptoms of coronavirus are fever, tiredness, dry cough. But day after day the patient rushes to death. Attempts to make a vixen for it are still ongoing. Although its outbreak was seen later in Bangladesh, Bangladesh is in danger in a very bad way. Bangladesh is in the same way as the whole world has been infected by the virus. In Bangladesh, as in other countries, the economy, education, health, and many more sector are all affected. The whole of Bangladesh is now locked down. People are no longer able to get out of the house for fear of the virus. The government is trying to take effective measures as much as possible. But human negligence is a big risk because of this virus and because of human transmission, this virus is only a few moments away to take the form of an epidemic. Nobel coronavirus hit Bangladesh on March 7, 2020, First case was confirmed on this day. Due to people coming from abroad, viruses come to the country. They brought the virus from other countries to their bodies. If they maintain the rules of 14 days hope it would not have been so epidemic then. After the 18th of March, It is assumed that the local transmission starts after 11 days. Now over 14000 people are affected by the Nobel coronavirus and more than two hounded people died by the virus. It is increasing day by day.

There are various ways in Bangladesh to get rid of this outbreak from all sectors. Artificial intelligence can be one of the most helpful in this situation. There is a lot of sector in AI (Artificial Intelligence) like big data, deep learning, machine leering, computer vision these can help in this situation a lot. The world tries to rid of this virus, everyone should try in his position as we try to like this in this epidemic situation. Inspired from these made a small attempt this time to fight against the coronavirus.

There will be data analysis and try to predict the future condition of Bangladesh in the coronavirus and it can help to understand the people about this situation who are not taken this epidemic seriously. It can be analyzed and understand the future situation. There some part of our research work as data visualization, these learn from data hope our algorithms will give us a good prediction, description of these algorithms, implement these algorithms lastly and elaborate expectation Work, how it can be improved and our future work.

II. Background and Motivation

Coronavirus has put powerful countries around the world at risk. If there is no awareness in Bangladesh too, the result could be very bad. Bangladesh has to face in a bad situation like other countries. The situation in Bangladesh is getting worse like other countries. The way a strong country works with the help of this Artificial Intelligence, everyone should also work to resist the lack. China, the first to be infected with the virus, has been able to reduce its epidemic with the help of artificial intelligence. With the help of Artificial Intelligence
Big data, Prediction genome sequencing, make faster diagnoses, carry out scanner analyses or, more occasionally, handle maintenance and delivery robot these sectors help a lot in these situations. That's why most of the countries are develop and try to get help from Artificial intelligence. All over the world working with AI so that it can be helpful for this epidemic situation. The sector of artificial intelligence is a great tool during this epidemic. We have to fight with the coronavirus to keep pace with the times. Artificial intelligence is a great tool for us to fight in this situation, so we have to emphasize this.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

For our research work, there have seen a lot of research work, theory, news and information.COVID-19 is being worked on every intention." AI and control of the Covid-19 coronavirus" there is a lot of information they given how China works with AI, how we can fight the coronavirus with Artificial Intelligence [3]. Their main contribution was how Artificial Intelligence contributed to search for a cure, the contribution of AI this epidemic situation, observer and predictor of the evolution of the pandemic, assist healthcare personnel, tool for population control, and evaluation of its use in the aftermath of a crisis. These sectors can be a helpful epidemic in every situation.

”Blockchain and AI-based Solutions to Combat Coronavirus (COVID-19) - like Epidemics: A Survey" this paper they used a blockchain and Artificial Intelligence to fight against the epidemic [4]. Blockchain detection of the detection outbreak, protecting user services, and outbreak tracking. On Other hand, Artificial intelligence provides a solution for symptoms and treatments. They also survey the latest research. So totally it's a combination of Blockchain and Artificial Intelligence.

”Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts” [7] Who work with Big Data, epidemic outbreak, Artificial Intelligence, and Deep Learning? The main focus of the research paper was preventing severe effects. They also work with big data and Artificial Intelligence then they identify the fight against the Coronavirus and solutions.

For research purpose need to understand some algorithms paper how they work, how they implement their algorithms because to work with these algorithms." Comparing K Nearest Neighbour and Linear Regression is there reason select one over the other” is one of them [8], though they are comparing these two algorithms they implement these algorithms in a very good way. They make an experiment performance of linear regression and K Nearest Neighbour. They worked on both balanced and unbalanced data set.

”Suitability of KNN Regression in the Development of Interaction Based Software Fault Prediction Models” it is also a prediction model, they using prediction with the help of KNN (KNearest Neighbour) Regression [10]. They predicted the model Development research also. A combination of matrices improve their model and work significantly with data mining. They used the classification task in data mining. Accurate fault prediction is an indispensable step with their KNN (K-Nearest Neighbour) regression model.

"Linear Regression Analysis Study" and "A study on Multiple Linear Regression" these two papers in the about regression model which are used in our research work. "Linear Regression Analysis Study" is about the statically procedure of find out dependent variable from independent variable [12]. They try to use the technique of modeling where the dependent variable predicts one or more independent variables. They also try to get the help of static's because it is based on the statics technique. They explain the basic concept of linear regression and also how can calculation SPSS and excel with the help of the regression model." A study on Multiple Linear Regression” is based on the relationship among variables and result relation. The main focus of their research work to know about the dependent variable and independent variable. They tried to make the relation of the dependent variable and independent variable. They worked with a multi linear model and analysis. They verified the data assumption with analysis.

III. Research Methodology

There is two part of our research methodology. These are Types of Data, Data visualization. The first section is to discuss the data. How to collect data, there are many types of data, discussion of these data, Data visualization is part to show the design of data. In this part, it is about to understand the total data. Trying research work makes clear so that it can be understood very well. Discussion about our total research work to make clear.

A. Types of Data

Data is the most important for research work. In this research work have worked with data. This research work is based on data. This research work contains various types of data. These data are collected from newspapers and online websites, every data is updating. Data is continuously adding day by day. Data is increasing day by day. Data collection will continue until this epidemic ends. Then research work will be able to analyze the total data. But now there are enough data to make predictions and to learn the machine. Data has been collected from many places. But some newspapers and websites have done a lot of help for data collection. In the case of newspapers, The Daily Star, a lot of data has been collected [15]. Data has been collected from Wikipedia and IEDCR (Institute of Epidemiology, Disease Control, and
Research) websites [18]. Some websites have been created in Bangladesh for COVID-19, got help from those websites for data collection. There are two types of data. These are for Data visualization. Another part where the data had trained, used algorithm and it shows us prediction and accuracy.

There are fifty categories of data set. It is containing huge data and it is increasing day by day. Data categories are the date, The number of days coronavirus has been in Bangladesh, total test cases, total death, total positive cases, total recovery cases, The number of people dying every day, The number of people positive cases every day, The number of people tests cases every day, The number of people recovery every day. Daily data is being selected according to date. There are some more data categories where all divisions have been selected in Bangladesh. Number of home quarantines by division, number of completed home quarantine cases, number of death cases in Division, number of male and female death cases, number of death according to age, total home quarantine cases, number of cases according to ages, daily percentage of male and female cases these are the categories of this research work. There are trained data and some data are for the test, train test method has been used. After training this data set, predict our values. Prediction

can be made in the condition of death, cases as anything to prediction. As data is increasing day by day so that train data test data are also increasing. So its accuracy and prediction will be better.

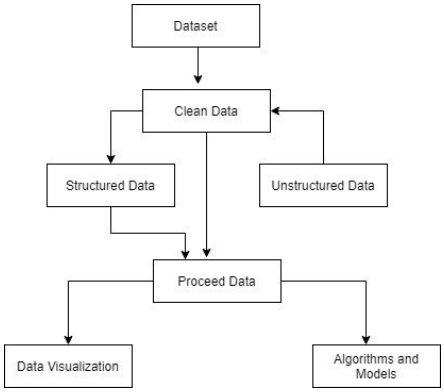


Fig. 1. Process of Data Collection

This is the total process of data collection, how the data was collected, and the method that is still using for increasing data collection. The whole process of data collection through graphs is going to be an idea.

B. Applied Machine Learning algorithm

Applied machine learning is a system of learning where we used for a specific problem. The relation between the data set and algorithms are used in the applied machine learning algorithm. For a specific problem solution with applied machine learning, need to fit the data and choose the correct algorithms to fit and solution for the algorithms. If failed to choose the algorithms or it was not fit for the solution then it will not show the correct solution for the problem. Applied algorithms have to be chosen for the work. Applied algorithms mostly use for the mapping of input and output knowledge.

C. Data visualization

This section is about visual data, find out the relationship of one data with another. Data visualization is graphical or pictorial from where the information can be understood. It helps any form of an innovative way of presenting large and complex information [20]. The virus is affected us day by day. How the virus is growing across the country. Day by day how people are leaning towards danger, the day there were some few cases but now the COVID-19 has reached epidemic proportions. People need to analyze so they can understand their direction. Death cases will visualize first.

D. COVID-19 Cases

At first, How COVID-19 total cases are increasing, how much COVID-19 cases are being tested every day, Day to day COVID19 test cases rate, how much COVID-19 is caught every day, increase and decrease of virus cases can see by the visualization. It is very useful to visualize this matter to visualize it. So at first the total test cases of COVID-19 will discuss because The most important thing to test for COVID19 is that without testing you will not realize that people are infected with the virus, if someone death by affected of virus we can find out it by testing. At present, 48 labs in Bangladesh are tested for the virus. When someone thinks that he has been infected with the virus, he calls the emergency number and comes and takes the sample. He was tested to find out if the COVID-19 was positive or negative. Bangladesh is increasing the number of its lab to test COVID-19.

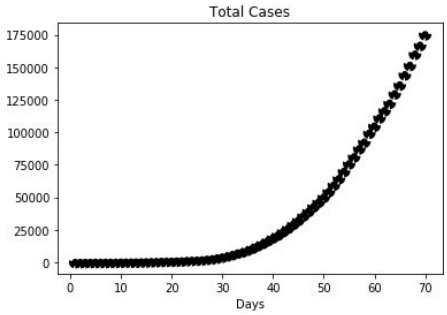


Fig. 2. New Test Cases according to days

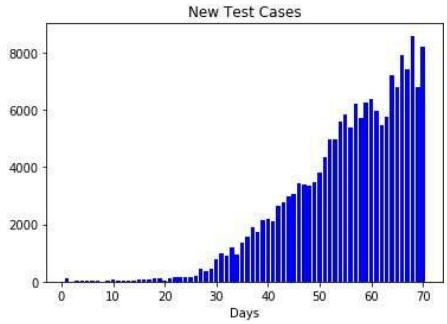


Fig. 3. New Test Cases according to days

In this direction, the discussion is about how many tests are done in our country every day. The more tests that are done, the better because it will let us know how much virus is spreading in the country. It will be better to give the effort to continue the test. The number of Tests is increasing day by day which is a good sign. .Analysis of last day, Cases of tested in Bangladesh is on the rise. It is hoped that this amount will continue to grow. Other countries are insisting on the test day by day.

Total positive cases are being visualized here. How bad it is can see here. It grows very fast at night. After 30 days, it grows very bad. It grows in an upward direction. Look at the last 10 days cases, its number is moving towards a serious epidemic. If it continues to grow in this way, the situation of Bangladesh will become like that of an affected European country. As another diagram, it shows are everyday positive cases of COVID19. How COVID-19 is growing has seen. Analysis of the last few days and it's ratio is increasing. After 60 days, its number is increasing in a very bad way. The most positive one is caught in

these few days. The amount of positive Cases does not decrease but increases.

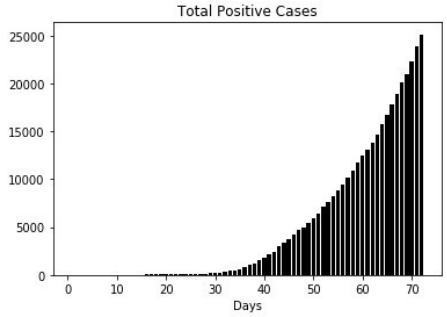


Fig. 4. Total Positive Cases according to days

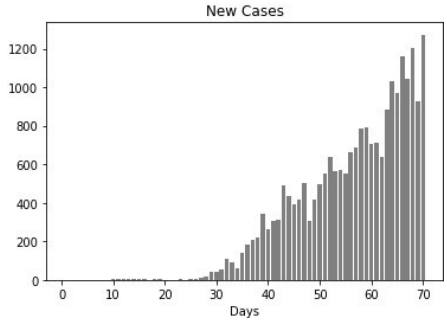


Fig. 5. Infected Positive Cases according to days

E. Death Cases

Visualization of COVID-19 Death Cases. At first, will see how Death cases grow day by day. After that visualize will be seen at the number of deaths in Dhaka Division. Male and female death cases will look for visualizing. Each death case is taken daily.

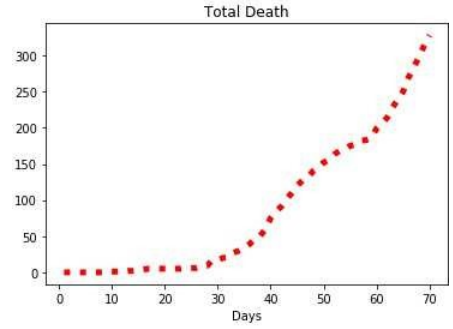


Fig. 6. Total cases of Death

This is the diagram of total death cases. You can just be increasing day by day. The situation can understand how fast COVID-19 death cases are on the rise. Looking at the day by day death cases, it is clear that the COVID-19 epidemic is getting worse.

This is in the death cases diagram of Dhaka according to date. All division in Bangladesh can find out every death cases

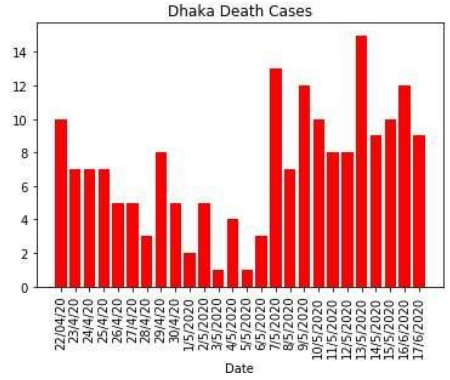


Fig. 7. Death cases of Dhaka

Diagram. Death cases data of all the divisions are in the data set but Dhaka Death Cases is more so Dhaka division is showing. This is the Dhaka division where the highest number died in one day. By data, visualization can understand this. Although the number has decreased in the middle, now it is seen that the number of Death cases is increasing.

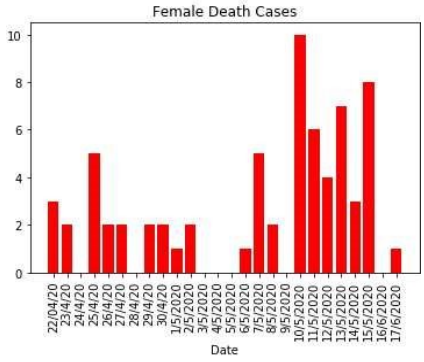


Fig. 8. Death cases of Female

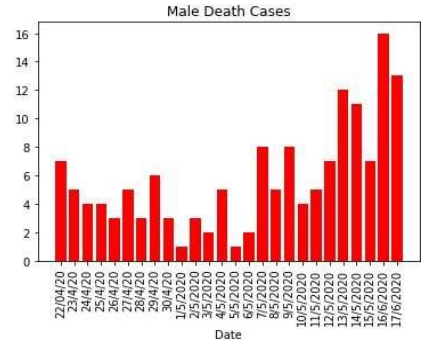


Fig. 9. Death cases of Male

The number of male death cases every day by looking at male and female cases. Some days it has happened that there is no female death case but almost every day, the number of male death cases is increasing. It is also a matter of concern that along with the death cases in Male, female death cases are Almost at the same rate.

F. Recovery Cases

This is the most important data because the country needs to understand that situation is getting better and worse. Recovery cases are those that have been affected by the virus and they have recovered and returned to normal. Every country is fighting against the virus. Affected patients are being tested to recover in various ways. The more the method works, the more the patient is being treated. The world will know which vaccine and which medicine works best, recovery can find out to analysis. The medicine that is working best in the recovery case is the medicine that is taking the patient for the better. It is the medicine of the next patient. This research work can clear many aspects through Recovery case Analysis. For the recovery of virus patients in Bangladesh, like the rest of the world. Efforts are being made in Bangladesh to heal the affected people of Bangladesh.

This is the graph of total recovery cases. After analysis, people can understand that recovery cases are increasing day by day. Here it is seen that the recovery case has suddenly increased. At first, the recovery cases were not known all over Bangladesh. Some lab of the capital city Dhaka was working on recovery cases. But now Recovery Cases are being upgraded and recovery cases are being collected through various labs in the country. Recovery Cases are being collected and analyzed all over Bangladesh.

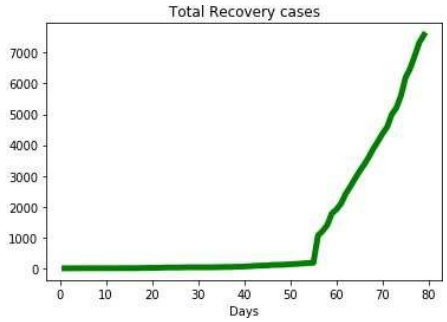


Fig. 10. Total cases of Recovery

As the discussion is about recovery cases that the number of recovery cases is increasing day by day which is giving the expected results. The recovery cases are looking daily for better analysis. This will make it easier to understand.

As analysis is seeing here that after 55 days the recovery Cases come out because recovery cases data has been collected from all over Bangladesh. At present, the number of recovery cases is increasing day by day as expected, as a result of which the country is seeing the light of hope.

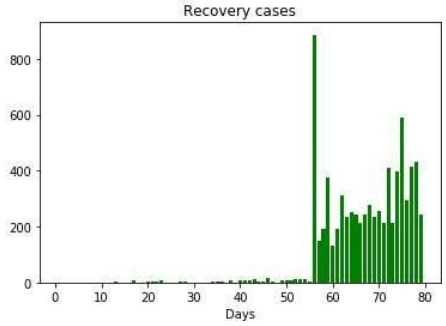


Fig. 11. Recovery cases according to days

EXPERIMENTAL RESULTS AND DISCUSSION:

Data are collecting constantly and are making predictions with the help of data. There are two algorithms for prediction. These are linear models and regression algorithms. As a result, by making predictions understand can be made about the condition of COVID-19.algorithmic accuracy come up with to test to find out who's making better predictions. This is the algorithm has used:

1. Linear Regression
2. KNN algorithm (k-Nearest Neighbors)

Regression analysis is a statistical analysis where there is a relationship between the dependent and two or more independent variables. There are many more regression algorithms but the most common regression in linear regression. The use of linear models and linear regression is more common in regression analysis. Linear regression has also been done in that series. Linear regression depends on the variable and at the same time uses the many independent variables. Independent variables are changing day by day. As a

Result, the prediction was able to update. So the independent variable is updating day by day. Form of equation of linear regression, then it will be y = MX + b is the basic linear regression equation Where, y= dependent variable m= slope b= intercept x= Independent variable

There is a relation to the true underlying parameters b, m, and data points is a create linear model.

Table 1: Classification of Linear Regression

Accuracy	98
coef	-8.37539048e+01,1.64593200e-01, 5.72713246e-01, 1.84456815e-03 -4.51650622e+01, 4.07358921e-02, 7.49409402e+01, 2.04531199e-02, 2.11251502e+00, 6.94709988e03, 5.19940817e-02, 7.57235023e-05, -2.46266926e-01, 1.83436442e-03
intercept	1.19408154e+03, -3.49402157e+00, 6.50109901e+02, -9.71838317e+02, -2.10490880e+01, 1.30112113e-01, -5.64815874e-01

A prediction analysis based on a linear regression algorithm. There is a prediction from 90 days, through linear regression get these predictions.

Table 2: Prediction Result on 90 days

Cases	Prediction Result
Days	90
Total test cases	400000
Total Cases	60290
Total Deaths	794
Total Recovery	12930
New test cases	13739
New Cases	2925
New Deaths	35
New Recovery cases	690
New Death cases of Dhaka	25

By analyzing 70 days of data, understand what future position will be. The new cases are the positive COVID-19 prediction on the 90th days. New test cases, new deaths, new recovery cases, new death cases are the prediction of the 90th day of all individual cases. And total cases are the prediction of total 90th days analysis. This is a prediction got from the linear regression algorithm. The algorithm learned from the data of 70 days, there are train and test method to find out the accuracy and prediction. Where the machine can automatically train and test its own and give the prediction.

KNN algorithm (k-Nearest Neighbors) is also a popular algorithm. KNN can use both classification and regression methods. The KNN regression model works well in the field of regression analysis and is used. So the KNN model was used and this algorithm works well. In KNN regression the output is trying to value the object. The KNN is the average value. It calculates the average numeric target of this algorithm. KNN regression and Classification has the same distance function.

Table 3: Classifier Of KNN

Accuracy	99
coef	-9.92581997e+01, 1.70176759e-01, 2.94443056e-01, 1.94052634e-03, -4.62080615e+01, 4.13116146e-02, 7.50054669e+01, 1.96511953e-02, 2.03522008e+00, 6.73069614e-03, 2.03522008e+00, 6.73069614e03, -1.09756824e+00, 1.89719860e03
intercept	1.46373915e+03, 1.36244356e+00, 6.73540695e+02, -9.59386003e+02, -1.56905385e+01, 2.79803914e-01, 1.28583620e+01

This is the prediction of 90 days of KNN ((k-Nearest Neighbors). This is the result of prediction by using KNN. This is the same prediction as to the linear regression prediction result.

Table 4: Prediction Result on 90 days

Cases	Prediction Result
Days	90
Total test cases	400000
Total Cases	56351
Total Deaths	763
Total Recovery	11875
New test cases	12602

New Cases	2559
New Deaths	36
New Recovery cases	520
New Death cases of Dhaka	25

Train and test method is used to train the data. All this data has been separated through the machine. After learning the data, It has automatically separated the train data and the test data. Dhaka City data have extracted it differently, so it came to the same prediction with Linear Regression and k-Nearest Neighbors. Its accuracy figured out its accuracy which will help us understand the functionality of the total model.

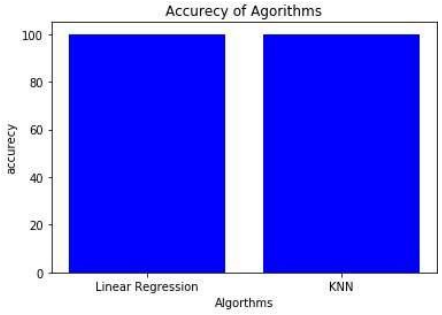


Fig. 12. Accuracy of Algorithms

There is a diagram of accuracy in which algorithms are used for research work. We used linear regression and the KNN algorithm, this is the diagram of the accuracy we found from these algorithms.

IV. Conclusion and Future Work:

COVID-19 is the epidemic panic all over the world. We should work for it from where we are. All caution and all sectors we work to reduce its impact.AI is also most help the full sector in this pandemic situation. .As long as this virus stays in Bangladesh, research will continue. This project may be a better model in the future. This model can use with the help of this epidemic in different countries. This analysis process and model can be used in the future. Or the algorithm that is not giving good predictions, need to work on the algorithm so that the algorithm gives more good predictions. More models can try to create using algorithms. We always work on learning from the past. Therefore, all the work of COVID-19 will be useful for our future. The work will work for us in the future as it will work for us now. Believe that this research work, analysis, and prediction model will help this epidemic situation.

References

[1] H. Li, S.-M. Liu, X.-H. Yu, S.-L. Tang, and C.-K. Tang, “Coronavirus disease 2019 (COVID-19): current status and future perspectives,” International Journal of Antimicrobial Agents, p. 105951, 2020

[2] M. Eisenstein, “Infection forecasts powered by big data,” Nature, vol. 555, no. 7695, 2018.

[3] Council of Europe, ”AI and control of Covid-19 coronavirus",2020.

[4] Dinh C. Nguyen, Ming Ding, Senior Member, IEEE, Pubudu N. Pathirana, Senior Member, IEEE, Aruna Seneviratne, Senior Member, IEEE, "Blockchain and AI-based Solutions to Combat Coronavirus (COVID-19)- like Epidemics: A Survey", SARS-CoV-2, 14 April 2020.

[5] WWW.COE.INT

[6] D. A. Keim, “Information visualization and visual data mining,” IEEE Trans. Vis. Comput. Graphics, vol. 8, no. 1, pp. 1–8, Jan. 2002.

[7] Quoc-Viet Pham, Dinh C. Nguyen, Thien Huynh-The, Won-Joo Hwang, and Pubudu N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-ofthe-Arts", IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, April 2020.

[8] Arto Haara, Annika Kangas, "COMPARING K NEAREST NEIGHBOURS METHODS AND LINEAR REGRESSION—IS THERE REASON TO SELECT ONE OVER THE OTHER?", Vol. 4, Issue 1, pp. 50–65,ISSN 1946-7664.MCFNS 2012.

[9] M. Callon, J. P. Courtial, and F. Laville, “Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry,” Scientometrics, vol. 22, no. 1, pp. 155–205, 1991.

[10] Fehrmann, L., A. Lehtonen, C. Kleinn, and E. Tomppo. 2008. Comparison of linear and mixed-effect regression models and a k-nearest neighbor approach for estimation of single-tree biomass. Can. J. For. Res. 38: 1–9.

[11] GK Uyanık, N Guler, "A study on multiple linear regression analysis-¨ sis”Procedia - Social and Behavioral Sciences 106 (234 – 240),2013

[12] Khushbu Kumari, Suniti Yadav, "Linear regression analysis study", Curriculum in Cardiology, IP: 14.139.45.244], Wednesday, May 23, 2018.

[13] Altman, N.S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Journal of the American Statistical Association 46: 175–185

[14] N. Crokidakis, “Data analysis and modeling of the evolution of COVID19 in Brazil,” arXiv preprint ar X iv:2003.12150, 2020.

[15] https://www.thedailystar.net

[16] Z.Xu, D.Yu, Y.Kao,andC.-T.Lin, "Thestructureandcitationlandscape ofIEEETransactionsonFuzzySystems(1994-2015), "IEEETrans.Fuzzy Syst., to be published, DOI: 10.1109/TFUZZ.2017.2672732.

[17] https://www.wikipedia.org

[18] https://iedcr.gov.bd

[19] J. Kim, H. Kim, and J. Diesner, “The impact of name ambiguity on properties of coauthorship networks,” J. Inf. Sci. Theory Pract., vol. 2, no. 2, pp. 6–15, 2014.

[20] Matthew N. O. Sadiku1, Adebowale E. Shadare2, Sarhan M. Musa3, and Cajetan M. Akujuobi4, "DATA VISUALIZATION "International Journal of Engineering Research And Advanced Technology(IJERAT), ISSN: 24546135, [Volume. 02 Issue.12, December– 2016]

[21] Haara, A., M. Maltamo, and T. Tokola. 1997. The K nearest-neighbor method for estimating basal-area diameter distribution. Scand. J. For. Res. 12: 200– 208.

[22] Magnussen, S., E. Tomppo, and R.E. McRoberts. 2010. A model-assisted k-nearest neighbor approach to remove extrapolation bias. Scand. J. For. Res. 25: 174– 184.

[23] E. Yan and Y. Ding, “Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coward networks relate to each other,” J. Amer. Soc. Inf. Sci. Technol., vol. 63, no. 7, pp. 1313–1326, 2012.

[24] [24] Public Health Blockchain Consortium: PHBC. [Online]. Available: https://www.phbconsortium.org/.

[25] Q. Liu, T. Liu, Z. Liu, Y. Wang, Y. Jin, and W. Wen, “Security analysis and enhancement of model compressed deep learning systems under adversarial attacks,” in 2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC), 2018, pp. 721–726.

[26] J. L. V. Sancho, J. C. Dominguez, and B. E. M. Ochoa, “An approach to the taxonomy of data visualization,” Revista Latina de Communication Social, vol. 69, 2014, pp. 486-507.

[27] D Karadag, Y Koc, M Turan, M Ozturk - Journal of Hazardous Materials, "A comparative study of linear and non-linear regression analysis for ammonium exchange by clinoptilolite zeolite”, Elsevier,2007

[28] Panchenko D. 18.443 Statistics for Applications, Section 14, Simple Linear Regression. Massachusetts Institute of Technology: MIT OpenCourseWare; 2006.
[29] [29] Freedman DA. Statistical Models: Theory and Practice. Cambridge, USA: Cambridge University Press; 2009.
[30] Z.Xu, D.Yu, Y.Kao,andC.-T.Lin, "Thestructureandcitationlandscape ofIEEETransactionsonFuzzySystems(1994-2015), "IEEETrans.Fuzzy Syst., to be published, DOI: 10.1109/TFUZZ.2017.2672732.
[31] X. Li et al., "Advanced aggregate computation for large data visualization," Proceedings of IEEE Symposium on Large Data Analysis and Visualization, 2015, pp. 137,138.
[32] Elazar JP. Multiple Regression in Behavioral Research: Explanation and Prediction. 2nd ed. New York: Holt, Rinehart, and Winston; 1982
[33] S. Grainger, F. Mao, and W. Buytaert, "Environmental data visualization for non-scientific contexts: Literature review and design framework," Environmental Modelling and Software, vol. 85, 2016, pp. 299-318.
[34] P. Gonczol, P. Katsikouli, L. Herskind, and N. Dragoni, "Blockchain implementation and use cases for supply chains-a survey," Ieee Access, vol. 8, pp. 11856–11871, 2020.
[35] Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of 2019 novel coronavirus at a global level. [Online]. Available: https://www.rti.org/publication/usetwittersocialmedia-activity-proxy-human-mobility-predictspatiotemporalspread
[36] S. Reddy, J. Fox, and M. P. Purohit, "Artificial intelligence-enabled healthcare delivery," Journal of the Royal Society of Medicine, vol. 112, no. 1, pp. 22–28, 2019.
[37] https://www.worldometers.info/coronavirus/country/bangladesh/
[38] W. Liu, P. T.-W. Yen, and S. A. Cheong, "Coronavirus disease 2019 (covid19) outbreak in china, spatial-temporal dataset," arXiv preprint ar X iv:2003.11716, 2020.
[39] [54] Q. Shen, T. Wu, H. Yang, Y. Wu, H. Qu, and W. Cui, "NameClarifier: A visual analytics system for author name disambiguation," IEEE Trans. Vis. Comput. Graphics, vol. 23, no. 1, pp. 141–150, Jan. 2017.
[40] https://www.who.int
[41] C. Chen. The Citespace Manual. Accessed: Mar. 24, 2018. [Online]. Available: http://cluster.ischool.drexel.edu/ cchen/citespace/
[42] A. Strokach, D. Becerra, C. Corbi-Verge, A. Perez-Riba, and P. M. Kim, "Fast and flexible design of novel proteins using graph neural networks," BioRxiv, 2020.

Direction of Arrival Identification Using MUSIC Method and NLMS Beamforming

Raungrong Suleesathira
Department of Electronic and Telecommunication Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi, Thung Khru, Bangkok, 10140, THAILAND
raungrong.sul@kmutt.ac.th

Abstract—**This paper provides the capability of the direction of arrival (DOA) identification to determine which the estimated DOA belongs to the desired signal and to undesired signals. One of the well known subspace-based methods for finding directions is MUSIC (MUltiple Signal Classification). The separation of signal and noise subspaces is the crucial step to give the precise estimation. The skewness coefficient is proposed to reinforce the conventional MUSIC method for the subspace division without knowing the number of source signals. The normalized least mean square (NLMS) beamforming is used to compute the weight vector so that it directs the mainbeam towards the desired user. The angle of the mainbeam is identified to be the DOA of the desired signal which makes the rest estimated DOAs belong to interference signals. The application of the DOA identification is shown to be advantageous to the null broadening beamforming. The simulation results confirm the effectiveness of the proposed method in the case of limited snapshots.**

Keywords-adaptive beamforming; DOA estimation; DOA identification; null broadening beamforming

I. INTRODUCTION

Direction of Arrival (DOA) estimation is an important issue in enhancing the quality of reception in wireless communication systems. Estimating the directions of signals impinging on the antenna array can accept signals from a certain direction, while rejecting signals that are declared as interferences. The importance of DOA estimation is not limited to communications. It can be utilized for target localization and tracking in radar system for commercial as well as for military applications.

DOA estimation can be performed using either parametric approaches or spectral approaches. Searching the values of the parameters related to the DOA such as maximum likelihood method [1] is the concept of the parametric approaches. The technique of the spectral approaches is to make the locations of sources appear in a spectral form. Spectral estimation techniques are subdivided into the non-parametric and subspace-based algorithms. Non-parametric algorithms include Capon method [1] whereas under subspace-based algorithm, MUSIC method [1, 2] has a better resolution. Capon methods failed mostly in cases when the sources are closely located or the incoming signals are highly correlated. However, such drawbacks can be solved by the MUSIC algorithm with spatial smoothing preprocessing.

There are some factors affecting the estimation results like number of array elements, number of snapshots, signal to noise ratio (SNR) and coherence of the signal source and closely spaced sources. Especially for super-resolution DOA estimation methods such as MUSIC, it requires to know the number of signals for dividing the signal and noise subspaces. However, in low SNR and low sample scenario, difficulties arise in the division. Akaiake Information Criterion (AIC) and Minimum Description Length (MDL) [3], are effective methods to estimate the number of signal sources which is the point at AIC or MDL minimum. This is an extremely computationally intensive process due to a search procedure requirement.

Various methods were proposed to enhance the DOA estimation in the case of unknown number of signal sources. In [4], it needs two steps: an algorithm using the linear prediction or Pisarenko methods in conjunction with adaptive signal parameter estimation and classification technique (ASPECT). After removing spurious peaks, the DOA estimation and the number of signals sources can be determined at the same time. According to the results, the probability of resolution is zero which is sensible to SNR below 10 dB. Similarly to [4], the change becomes using alternating projection and weighted subspace fitting in conjunction with ASPECT [5]. The probability of resolution for correlated signal is enhanced. An algorithm based on particle swarm optimization is proposed in [6] to minimize the fitness function for DOA estimation without knowing the number of signal sources. The solution of the unconstrained optimization required a recursive algorithm. In [7], a new optimization problem is proposed which is independent on subspace decomposition so that the number of sources is not required. Instead, the solution requirement is to know the look direction and the array correlation matrix.

Many attempts have been made on DOA estimation without identifying which directions belong to the desired signal and interference signals. Additionally, various beamforming algorithms require the prior knowledge of the directions of the sources. This requirement is mostly done by assumption. None of researches has been conducted for DOA

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

identification. In this paper, a procedure to identify which signal the estimated DOA belongs to is presented. It begins by using MUSIC method for DOA estimation. To generate the noise subspace, the skewness coefficients are applied to augment MUSIC algorithm. The normalized least mean square (NLMS) beamforming is adaptive to identify the DOA of the desired signal. Once the DOA of the desired signal is decided, the rest estimated DOA is definitely belongs to interference signals. To illustrate the use of DOA identification, null broadening beamformers are applicative.

The paper is organized as follows. Section II introduces the signal model. In section III, MUSIC method is explained. Section IV provides the normalized least mean square beamforming. Benefit of DOA identification to null broadening beamformers is highlighted in section V which consists of the covariance matrix taper (CMT) and projection and diagonal loading (PDL) approaches. Section VI presents the simulation results to validate the effectiveness of the proposed technique. The concluding remarks are given in section VII.

II. SIGNAL MODEL

In cellular system, a base station (BS) mounted with an antenna array serves to multiple single-antenna mobile stations (MS) as shown in Fig. 1. Since the intra-cell users (red line) are allocated to difference channel resources, therefore they do not interfere to each other. Due to the frequency reuse, the inter-cell users located at nearby cells can interfere to the considered BS (blue line). Let the intra-cell user and the inter-cell users considered as the desired signal $s_d(n)$ and interference signal $s_i(n)$, respectively. Consequently, the uplink signal arriving to the BS consists of the desired signal and interference signals under an additive zero-mean white Gaussian noise. Note that a narrowband propagation model is used, i.e., the signal envelopes do not change from one antenna element to another. In the far field, the received signal at time n by an L-elements uniform linear array (ULA) as depicted in Fig. 2 is modeled as [8, 9]

$$\mathbf{x}(n) = s_d(n)\mathbf{a}(\theta_d) + \sum_{i=1}^{I} s_i(n)\mathbf{a}(\theta_i) + \mathbf{z}(n) \qquad (1)$$

where $\mathbf{x}(n) \in \mathbb{C}^{L\times 1}$ in the presence of the Gaussian noise $\mathbf{z}(n) \in \mathbb{C}^{L\times 1}$. The vectors $\mathbf{a}(\theta_d)$ and $\mathbf{a}(\theta_i)$ represent the steering vectors of the desired signal related to the DOA θ_d and of the interference signal related to the DOA θ_i $i = 1,2,\ldots,I$, respectively. The steering vector as a function of DOA θ is denoted as $\mathbf{a}(\theta) = [1 \quad e^{-j2\pi\frac{d}{\lambda}\sin\theta} \quad \ldots \quad e^{-j2\pi(L-1)\frac{d}{\lambda}\sin\theta}]^T$ where $\theta \in [-90°, 90°]$, d is the distance between the elements and λ is the wavelength. The superscript T denotes transposition. Eq. (1) can be rewritten as

$$\mathbf{x}(n) = \mathbf{A}(\Theta)\mathbf{s}(n) + \mathbf{z}(n) \qquad (2)$$

where $\mathbf{s}(n) = [s_d(n) \quad s_1(n) \quad \ldots \quad s_I(n)]^T$ is a source vector and $\mathbf{A}(\Theta) = [\mathbf{a}(\theta_d) \quad \mathbf{a}(\theta_1) \quad \ldots \quad \mathbf{a}(\theta_I)]$ is a steering matrix of the angle set $\Theta = \{\theta_d, \theta_1, \ldots, \theta_I\}$. Note that the available degrees of freedom are generally one less than the number of sensors, i.e., $(I+1) < L$.

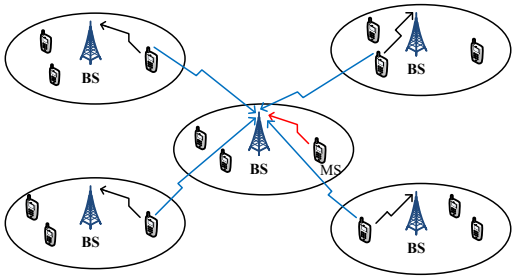


Figure 1. Cellular system scenario with the desired signal (red line) as the intra-cell user and interference signals as the inter-cell users (blue lines)

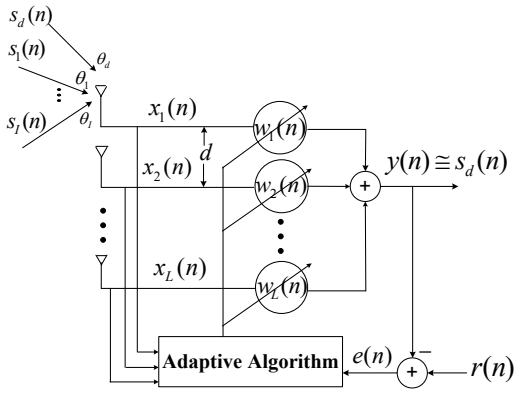


Figure 2. A ULA of L antenna elements receiving $I+1$ directional sources and a generic diagram of an adaptive beamforming

The covariance matrix of the received signal can be defined as

$$\mathbf{R}_{\mathbf{x}} = E[\mathbf{x}(n)\mathbf{x}^H(n)] \qquad (3)$$

where $E[\bullet]$ denotes the expectation and the superscript H denotes conjugate transposition. The source signals are uncorrelated and the noise process is independent of the source signals. The noise is assumed to be temporally and spatially uncorrelated and independent. Substituting Eq. (2) into Eq. (3), then Eq. (3) becomes

$$\mathbf{R}_{\mathbf{x}} = \mathbf{A}(\Theta)\mathbf{R}_{\mathbf{s}}\mathbf{A}^H(\Theta) + \sigma^2\mathbf{I} \qquad (4)$$

where $\mathbf{R}_{\mathbf{s}} = E[\mathbf{s}(n)\mathbf{s}^H(n)]$ is an $L \times L$ source covariance matrix, σ^2 is the noise power and $\mathbf{I}$ is an identity matrix of L dimension. In practice, the covariance matrix $\mathbf{R}_{\mathbf{x}}$ is unknown. The sample covariance matrix is used instead when there are N snapshots as

$$\hat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{x}(n)\mathbf{x}^H(n). \qquad (5)$$

Denoting $\mathbf{w}(n) \in \mathbb{C}^{L\times 1}$ as the weight vector, the output of an adaptive beamformer can be written as

$$y(n) = \mathbf{w}^H(n)\mathbf{x}(n). \quad (6)$$

The aim is to produce the output signal $y(n)$ approximate the desired signal $s_d(n)$ as shown in Fig. 2.

III. MUSIC Algorithm Augmented with Skewness Coefficient

MUtiple SIgnal Classification (MUSIC) method is a subspace-based approach for DOA estimation. According to Eq. (1), the received signal is separable to the signal and noise subspace. The separation is done by eigen-decomposition of the sample covariance matrix as

$$\hat{\mathbf{R}}_{\mathbf{x}} = \sum_{l=1}^{L} \lambda_l \mathbf{e}_l \mathbf{e}_l^H . \quad (7)$$

The eigenvalues $\{\lambda\}_{l=1}^{L}$ have the following relation:

$\lambda_1 > \ldots > \lambda_{(I+1)} > \lambda_{(I+2)} = \ldots = \lambda_L = \sigma^2$. Eigenvalue λ_l corresponds to the eigenvector $\mathbf{e}_l$. The noise subspace can be constructed by $\mathbf{E}_n = [\mathbf{e}_{(I+2)} \quad \mathbf{e}_{(I+3)} \quad \ldots \quad \mathbf{e}_L]$. The number of columns of $\mathbf{E}_n$ is equal to $r = L-(I+1)$ with each column being the eigenvectors corresponding to the $L-(I+1)$ smallest eigenvalues. Once the noise subspace is determined, a search for directions is made by looking for the steering vectors that are orthogonal to the noise subspace. It leads to the MUSIC spatial spectrum given as

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)} . \quad (8)$$

After searching the largest peaks of $P_{\text{MUSIC}}(\theta)$ over all θ, the DOA estimates are the points where $P_{\text{MUSIC}}(\theta)$ take the peaks.

To separate the signal and noise subspace properly, the priori information of the number of signal sources is needed. One way of distinguishing between the signal and noise eigenvalues is to determine the number of small eigenvalues that are equal. Since the sample covariance matrix of the received signal is formed using finite samples in practice, the smallest eigenvalues are not exactly equal. Note that when the number of snapshots is limited in Eq. (5), the performance of the subspace-based DOA estimation method such as MUSIC deteriorates due to the distortion of noise subspace. On the other word, the error of covariance matrix induced by a few N snapshots can cause the DOA estimation bias.

Without knowing the number of sources, in this paper, the skewness coefficient of eigenvalues for selecting the number of eigenvectors r is proposed to determine the noise subspace $\mathbf{E}_n$ and support the case of limited snapshots. The skewness coefficient for each eigenvalue $skew(\lambda_l)$ for M simulation can be calculated by [10]

$$skew(\lambda_l) = \frac{\frac{1}{M}\sum_{m=1}^{M}(\lambda_{l,m} - \bar{\lambda}_l)^3}{\left[\frac{1}{M-1}\sum_{m=1}^{M}(\lambda_{l,m} - \bar{\lambda}_l)^2\right]^{3/2}} \quad (9)$$

where $\lambda_{l,m}$ is the m^{th} observation for the l^{th} eigenvalue and $\bar{\lambda}_l = \frac{1}{M}\sum_{m=1}^{M}\lambda_{l,m}$. If $skew(\lambda_c)$, $c \in \{1,\ldots,L\}$, is the maximum, then select $r = L-(c-1)$. The procedure of MUSIC method involving the skewness coefficients can be summarized as follows:

1. Calculate the sample covariance matrix Eq. (5) and find its eigen-decomposition Eq. (7).
2. Simulate the received signals M times and calculate the skewness coefficient Eq. (9) for each eigenvalue to select the optimum value of r.
3. Construct the noise subspace $\mathbf{E}_n$ spanned by the eigenvectors corresponding to the r smallest eigenvalues.
4. Plot the MUSIC spatial spectrum Eq. (8).
5. Search the positions of the large peaks as the estimated signal directions.

IV. Normalized Least Mean Square Beamforming

For the time-varying signal propagation environment, an adaptive algorithm is needed to track a moving MS. Fig. 2 shows a generic diagram of an adaptive beamforming. The received signal vector $\mathbf{x}(n)$ and the error signal $e(n)$ are fed into an adaptive algorithm which controls the weight vector according to some criteria for termination. The output signal is substracted from an available reference signal $r(n)$ to generate an error signal as $e(n) = r(n) - y(n)$. The mean squared error (MSE) for a given $\mathbf{w}$ is given as [11, 12]

$$\begin{aligned} E\left[|e(n)|^2\right] &= E\left[|r(n) - y(n)|^2\right] \\ &= E\left[|r(n) - \mathbf{w}^H(n)\mathbf{x}(n)|^2\right] \\ &= E\left[\{r(n) - \mathbf{w}^H(n)\mathbf{x}(n)\}\{r(n) - \mathbf{w}^H(n)\mathbf{x}(n)\}^*\right] \\ &= \mathbf{w}^H\mathbf{R}_{\mathbf{x}}\mathbf{w} - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + p_0^2 \end{aligned} \quad (10)$$

where $\mathbf{p} = E[\mathbf{x}(n)r^*(n)]$ is the $L\times 1$ cross correlation vector between the received signal vector and the complex conjugate of the reference signal and $p_0 = E\left[|r(n)|^2\right]$ is the power of the reference signal. By the gradient vector of Eq. (10) with respective to complex conjugate of $\mathbf{w}$, the optimum solution is the well known Wiener-Hoff equation as

$$2\frac{\partial E\left[|e(n)|^2\right]}{\partial \mathbf{w}^*}\bigg|_{\mathbf{w}_{\text{MSE}}} = 0 \quad (11.1)$$

$$2\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{MSE}} - 2\mathbf{p} = 0 \tag{11.2}$$

$$\mathbf{w}_{\mathrm{MSE}} = \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p} \tag{11.3}$$

where $(\bullet)^{-1}$ denotes the inverse operation. From Eq. (11), a priori knowledge of $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{p}$ and the computation of inversion are needed. This requires high computation especially in cases involving real-time processing, massive array and nonstationary scenario. Therefore, to overcome these difficulties, some adaptive methodology is introduced such as least mean square (LMS) algorithm based on the steepest-descent method. The recursive relation for updating the weight vector is

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{2}\nabla E\left[\,|\,e(n)\,|^2\,\right] \tag{12}$$

where the gradient vector of $E\left[\,|\,e(n)\,|^2\,\right]$ with respect to the complex conjugate of $\mathbf{w}$ is

$$\nabla E\left[\,|\,e(n)\,|^2\,\right] = 2\mathbf{R}_{\mathbf{x}}\mathbf{w}(n) - 2\mathbf{p}. \tag{13}$$

Replacing $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{p}$ by the instantaneous estimates given as $\mathbf{R}_{\mathbf{x}} = \mathbf{x}(n)\mathbf{x}^{H}(n)$ and $\mathbf{p} = \mathbf{x}(n)r^{*}(n)$ into Eq. (13), the LMS algorithm is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e^{*}(n)\mathbf{x}(n) \tag{14}$$

where μ is a scalar constant which controls the rate of convergence. Since only the received signal vector present in Eq. (14), the computational complexity is reduced. In addition, employing the LMS algorithm requires a reference signal which can be generated in a number of ways depending on the applications. In digital mobile communications, the intra-cell user transmits a predefined training sequence known to the BS as a reference signal.

A normalized version of the LMS algorithm is the change of the update rule by substituting the step-size constant μ in Eq. (14) as $\dfrac{\mu}{\mathbf{x}^{H}(n)\mathbf{x}(n)}$. This time-varying step size gives a faster convergence. Thus, the normalized LMS (NLMS) update rule is obtained as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\frac{e^{*}(n)\mathbf{x}(n)}{\mathbf{x}^{H}(n)\mathbf{x}(n) + \varepsilon}. \tag{15}$$

Adding a regularization term ε is to prevent divisions by very small value. After convergence of the weight vector $\mathbf{w}$, calculate the array pattern given as

$$B(\theta) = 20\log_{10}|\,\mathbf{w}^{H}\mathbf{a}(\theta)\,|\,. \tag{16}$$

where $B(\theta)$ is the average power of the response that has the maximum gain from the direction θ_d. As a result, we can use the NLMS beamformer to identify which the DOA estimate belongs to the desired signal and the remaining DOAs belong to undesired signals. This DOA identification could be benefit to some beamforming approaches.

V. Application of DOA Identification to Null Broadening Beamforming

When an angular spread exists around the DOA of interference signals due to multipath propagation or rapidly motion, the interference cancellation performance by using the traditional beamformers can severely degraded as the output SINR (signal to interference plus noise ratio) might be low. It is then desired to create a suppressed angular sector in the beampattern. Forming broad nulls at the DOA of interferences is an effective means to reject the interference signal not only from a specific direction but also from a specific spatial region. As a result, null broadening allows interference signals move in a certain area. Several null widening techniques have been proposed in literature. Based on matrix tapers [13], a covariance matrix taper (CMT) and the projection and diagonal loading (PDL) are used to demonstrate the application of the proposed DOA identification.

A. The CMT Approach

The tapered sample covariance matrix $\mathbf{R}_{\mathrm{CMT}}$ can be determined by the Hadamard product $\circ$ between the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{x}}$ and the taper matrix $\mathbf{T}$ as

$$\mathbf{R}_{\mathrm{CMT}} = \hat{\mathbf{R}}_{\mathbf{x}} \circ \mathbf{T}. \tag{17}$$

The element of a^{th} row and b^{th} column of $\mathbf{T}$ is given as

$$T_{ab} = \frac{\sin((a-b)\Delta)}{(a-b)\Delta} \tag{18}$$

where Δ denotes the width of the nulls. Nulls can be broaded by using the optimal weight vector of the minimum variance distortion response (MVDR) beamformer as [13]

$$\mathbf{w}_{\mathrm{CMT}} = \frac{\hat{\mathbf{R}}_{\mathrm{CMT}}^{-1}\mathbf{a}(\theta_d)}{\mathbf{a}(\theta_d)\hat{\mathbf{R}}_{\mathrm{CMT}}^{-1}\mathbf{a}(\theta_d)}. \tag{19}$$

To avoid calculation of the matrix inversion in Eq. (19), the iteration version of MVDR beamforming is presented. The weight adjustment is updated as [8, 14]

$$\mathbf{w}_{k+1} = \mathbf{w}_k - c(\mathbf{R}_{\mathrm{CMT}}\mathbf{w}_k - \mathbf{a}(\theta_d)). \tag{20}$$

The constant c is set to $c = \dfrac{2}{(\lambda_{\min} + \lambda_{\max})}$ where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $\hat{\mathbf{R}}_{\mathbf{x}}$, respectively.

In CMT approach, it only requires the prior information of θ_d without knowing θ_i. Then, the good performance depends on accurately known the steering vector of the desired signal, $\mathbf{a}(\theta_d)$. Although the computational complexity is very low, the null depth of the CMT is shallow.

B. The PDL Approach

To increase the null depth, the CMT and projection are combined. For PDL approach, prior knowledge of θ_i is needed besides θ_d. It begins with constructing the correlation matrix of I interference signals as

$$\mathbf{Z}_I = \sum_{i=1}^{I} \mathbf{a}(\theta_i)\mathbf{a}^H(\theta_i). \quad (21)$$

Taper the correlation matrix of interference signals as

$$\mathbf{R}_I = \mathbf{Z}_I \circ \mathbf{T}. \quad (22)$$

Eigen-decompose $\mathbf{R}_I$ yields

$$\mathbf{R}_I = \sum_{l=1}^{L} \upsilon_l \mathbf{v}_l \mathbf{v}_l^H. \quad (23)$$

By using P eigenvectors belonging to P largest eigenvalues, the projection matrix is constructed as

$$\mathbf{T}_p = \sum_{l=1}^{P} \upsilon_l \mathbf{v}_l \mathbf{v}_l^H. \quad (24)$$

Project $\mathbf{T}_p$ onto the sample covariance matrix as

$$\mathbf{R}_{\text{PDL}} = \mathbf{T}_p \hat{\mathbf{R}}_{\mathbf{x}} \mathbf{T}_p^H. \quad (25)$$

Like Eq. (19), the optimal weight vector for the PDL approach is expressed as [13]

$$\mathbf{w}_{\text{PDL}} = \frac{\hat{\mathbf{R}}_{\text{PDL}}^{-1}\mathbf{a}(\theta_d)}{\mathbf{a}(\theta_d)\hat{\mathbf{R}}_{\text{PDL}}^{-1}\mathbf{a}(\theta_d)}. \quad (26)$$

Like Eq. (20), the weight vector is updated as [8, 14]

$$\mathbf{w}_{k+1} = \mathbf{w}_k - c(\mathbf{R}_{\text{PDL}}\mathbf{w}_k - \mathbf{a}(\theta_d)). \quad (27)$$

VI. Simulation Results

A 12-element ULA is used and inter element spacing of $d = \lambda/2$ is created. Two interference signals are from directions of $\theta_1 = -40°$ and $\theta_2 = 40°$. Assume the width of broad null is set to $\Delta = 8°$. The DOA of the desired signal is taken at $\theta = 0°$. All signal sources have the same power in the additive white complex Guassian noise with zero mean. The input SNR is considered to be 0 dB while the INR (interference and noise ratio) is set to 3 dB. The number of snapshot is limited to $N = 30$ in order to demonstrate the low sample support.

Figure 3 illustrates the result of $skew(\lambda_l)$ for $l = 1,\ldots,12$. Each skew coefficient is the average for 1000 simulations. It indicates that $skew(\lambda_4)$ is maximum, then $c = 4$. This enables us to construct the noise subspace $\mathbf{E}_n$ spanned by the eigenvectors corresponding to the nine smallest eigenvalues. Once the noise subspace is obtained, the DOA estimate is done via the MUSIC spatial spectrum. According to Fig. 4, the DOA estimate is precise where three distinct peaks are located at $-40°, 0°, 40°$ exactly the DOAs of the source signals. Since there is no mechanism embedded in the MUSIC algorithm to identify which DOA belongs to the desired signal. In this paper, the NLMS beamforming is used to form the main beam at the desired signal direction. Due to the weight adjustment in Eq. (14) depending on the received signal vector, the number of iterations is equal to the number of snapshots. The constant is taken to $\mu = 0.1$. In Fig. 5, the array pattern places the main beam at $0°$. Therefore, DOA identification is achieved by setting the angle $0°$ belongs to the desired signal DOA and certainly, the remaining angles at $-40°, 40°$ belongs to the interference signal DOAs.

The benefit of this DOA identification is shown by applying to two types of null broadening beamformers: the CMT approach and the PDL approach. Without iteration using Eqs (19) and (26), the mainbeams are deformed and high rising sidelobes occur as shown in Fig. 6. As mentioned previously, the null depth of the PDL approach (red) is deeper that the CMT approach (blue). The array pattern of the PDL approach widens the two nulls from $-36°$ to $-44°$ and from $36°$ to $44°$ corresponding to $\Delta = 8°$. The results of the iterative null broadening beamforming using Eqs (20) and (27) are shown in Fig. 7. The PDL approach (red) not only has accurate null positions but also has good performance on the mainbeam and low sidelobes. For CMT approach, the use of iterative beamforming gives the array pattern similar to the result of without the weight adjustment.

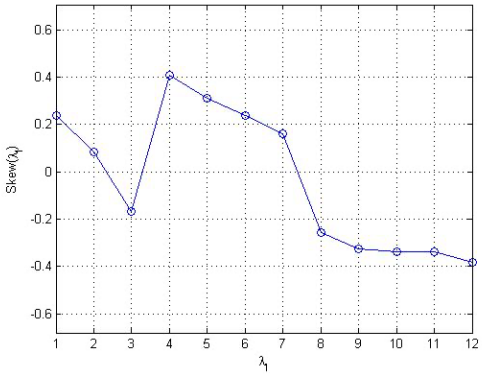


Figure 3. Skewness coefficients

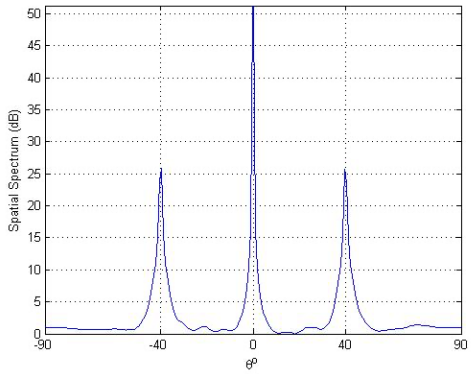


Figure 4. MUSIC spatial spectrum

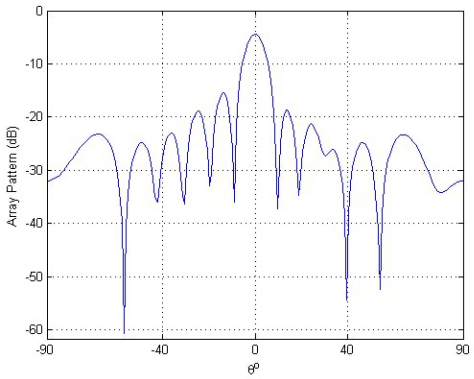


Figure 5. NLMS beamforming

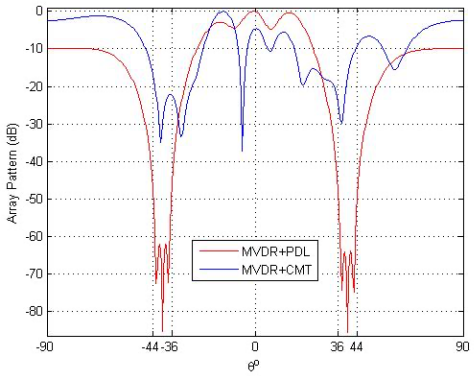


Figure 6. Null broadening beamforming

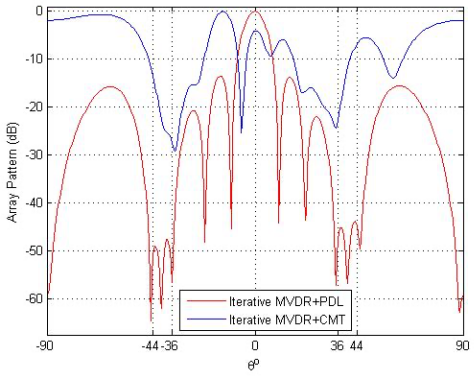


Figure 7. Iterative null broadening beamforming

VII. CONCLUSIONS

The DOA information is normally assumed to be known for several applications. Thereby, the DOA estimation and identification is proposed by using the MUSIC method and NLMS adaptive beamforming. The MUSIC algorithm is used to estimate DOAs. Unlike the conventional MUSIC procedure, the skewness coefficients are dedicated to determine the dimension of the noise subspace with unknown the number of signal sources. The use of NLMS beamforming can converge to form the mainbeam at the desired user direction. This enables us to identify the DOA of the interference signals. Two types of null broadening MVDR beamformings which their computation requires to know which signal the DOA belongs to are shown as the application of this proposed DOA identification.

REFERENCES

[1] J. Foutz, A. Spanias, and M. K. Banavar, "Narrowband direction of arrival estimation for antenna arrays," Morgan & Claypool, 2008.

[2] M. Devendra, and K. Manjunathachari, "DOA Estimation of System using MUSIC method," *International Conference on Signal Processing and Communication Engineering*, pp. 309-313, Jan. 2015

[3] M. Wax, and T. kailath, "Detection of signals by information theoretical criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 33, pp. 387-392, Apr. 1985.

[4] Q. Chong-ying, Z. Yong-shun, H. Ying, and C. Xi-hong, "An algorithm on high resolution DOA estimation with unknown number of signal sources," *4th International Conference on Microwave and Milimeter Wave Technology*, pp. 227-230, Aug. 2004

[5] C. Park, J. Choi, J. Yang, and S. Nah, "Direction of Arrival Estimation using Weighted Subspace Fitting with Unknown Number of Signal Source*," 11th Internationa Conference on Advanced Communication Technology*, pp. 2295-2298, Feb. 2009.

[6] M. Taha, and D. Alnadi, "DOA Estimation for unknown number of signals using particle swarm optimization," *IEEE international Conference on Computer, Communication and Control Technology*, pp. 167-171, Apr. 2015.

[7] Z. Ying, and B. P. Ng, "Music-like DOA estimation without estimating the number of sources," *IEEE Transactions on Signal Processing*, Vol. 58, No. 3, pp. 1668-1676, 2010.

[8] R. Suleesathira, "Projected iterative MVDR beamforming for null broadening and first sidelobe suppression in the presence of DOA mismatch," *10th International Workshop on Computer Sicence and Engineering*, pp. 95-102, Feb. 2020.

[9] R. Suleesathira, "Robust LCSS Beamformer against DOA mismatch," *11th International Conference on Advances on Information Technology*, No. 4, pp1-7. July, 2020.

[10] N. Alharbi, and H. Hassani, "A new approach for selecting the number of the eigenvalues in singular spectrum analysis," *Journal of the Franklin Institute*, Vol. 353, pp. 1-16, 2016.

[11] H. Zeng, Z. Ahmad, J. Zhou, Q. Wang, and Y. Wang, "DOA estimation algorithm based on adaptive filtering in spatial domain," *China Communications*, Vol. 13, pp. 49-58, Dec. 2016.

[12] G. M. Zilli, C. A. Pitz, E. L. O. Batista, R. Seara, and W. Zhu, "An adaptive approach for the joint antenna selection and beamforming optimization," *IEEE Access*, Vol. 7, pp. 99017-99030, July 2019.

[13] W. Li, Y. Zhao, Q. Ye, and B. Yang, "Adaptive antenna null broadening beamforming against array calibration error based on adaptive variable diagonal loading," *International Journal of Antennas and Propagation*, Article ID 3265236, 9 pages, 2017.

[14] B. Ors and R. Suleesathira, "First and Second Order Iterative Null Broadening Beamforming," *3rd International Conference on Imaging, Signal Processing and Communication*, pp.48-51, Jul. 2019.

A Structured Transformer Neural Machine Translation on Abstractive Text Summarization for Bangla seq2seq Learning

Abu Kaisar Mohammad Masum
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
mohammad15-6759@diu.edu.bd

Sharun Akter Khushbu
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
khusbhu15-6083@diu.edu.bd

Mumenunnessa Keya
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
mumenunnessa15-10100@diu.edu.bd

Sheikh Abujar
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sheikh.cse@diu.edu.bd

Syed Akhter Hossain
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
aktarhossain@daffodilvarsity.edu.bd

***Abstract* — Neural Machine Translation is a dominant sequence of text summary representations addressing neural summarization throughout the functionality of original code summarization. Transformer articulated an attention mechanism in which encoder decoders frequently modeling over language pairs between source of input sequences to longest n-grams dependencies. MT tends to learn sequence-to-sequence patterns that have an effective outperform of learning code representation. By this simple mechanism inflict a self-attention architecture where it concatenate with Multiple heads of word vectors to accept their Long-range paraphrasing. Encoding layer profound with LSTM network and decoding layer also consist of the same layer with an opponent function while it paraphrases the code token by positional encoding then the model generates a superior performance of text summarization. Due to the study of abstractive summarization we process text in Bangla input from valuable sources after training the model the pair of Bangla text-to-Bangla summary visualize a momentous margin that will also employs next future work.**

Index terms* — *MT, Bangla-to-Bangla, NMT, Transformer, Abstractive Text summarization, LSTM.

I. INTRODUCTION

The text of summarization meaning in which the transformation process is conversion of languages that has come on account of the endeavours of Neural Networks. 20th century neural networks introduced by deep neural networks as a subfield [1] that has a power of transfer languages in Machine translation. Henceforth, MT generates sequence-to-sequence [2] approach with NMT model. Recently this study underwent state-of-the-art performance expansion [3]. The concept of statistical Machine translation how fixed the effective output had formerly been discussed in few research but now NMT generate an text summarization based approach according to the lack of complete pipeline in large volume of documents. Loss of the translation quality is the SMT boundaries where MT community guide the NMT model.

Due to the attention mechanism analysis of sentence structure into token of words has been possible with NMT where networks build two LSTM encoder and decoder networks. Typically, LSTM encoder and decoder formulate a contrastive pattern delivered a model in which encoder processes input source in Bangla sentences and another one decoder decodes the prediction of vector receiving token words to generate Bengali language summary which is our output. Without loss of transformation quality Natural language processing produces privilege in abstractive summarization of different languages that consist of core architecture modulated with RNN also encoding happens by the Convolutional Neural Network (CNN).

NMT [4] has the flexibility to train corpus into vector output in an end-to-end process that translates the longest sequence from input in Bangla language. Language translation models familiar with the couple of encoder decoder layers by the NMT where it generates chain system output using statistical reorder of the sentences and each word translation connects to the next word that is concentrated with NMT sentence based translation. According to this approach, it gives assurance of maximum translation where it builds by the single neural network of machine translation [5]-[7]. NMT leads faster than SMT to define translation models where there is no need for other different model existence in the phrase based translation language.

Typically, longer sentences are difficult to transfer into the context of fixed vectors nevertheless NMT drives the State-of-the-art representation very effectively in different language pairs. According to the study of review we have found some limitations about long sentences [4] where encoding and decoding with attention is needed also we take initiative to do it for Bangla sentences. In literature Attention mechanisms help to translate in which model jointly learns for Bangla-English, English-Bangla [8] English-Germany [9], English-French [10] and so many pairs that have shown a different flow of attention mechanism [29], [28]. NMT [28], [24], [25] has developed the community couple with self-attention [24] and multi head attention based in order to simulate the length of word connectivity.

This is a sequence-to-sequence [11] text summarization learning process by NMT in which the machine reduces the words by generating a summary from available document type data in Bangla language. For long range dependencies NMT [12] [13] approaches conduct with an experiment and

evaluates the Bangla sentences document to summary document.

The demonstration of abstractive summarization based on NMT is as follows. Section II, that shows the relevant work of MT regarding Bangla and other different languages also brief discussion over NMT based on other studies. In Section III that we discuss our experimental methodology in order to section IV is the most expected result and the discussion part shows the output. In Section V lastly we conclude the study of our language translation process.

II. Literature Review

The text summarization with MT exploration along with languages conversion was represented in 1991 [14]. Until now Bangla-English language and other language pairs have been getting much seeking attention. MT started when research got placed using rule based approaches existing with predefined rules for Bangla-English language to check grammar [15], or structural sentence statement using flow of rule based model over training [16], [17]. Apart from that MT also successfully SMT approaches covered by text sentences to word or phrasal models in Bangla-English text summarization approach [18].

In [19], the authors developed an abstractive summarization in sequence-to-sequence rule set as a pre-trained transformer with a single encoder or decoder on daily mail dataset. With pre-trained BERT/bidirectional or attention mechanism compare impressive result between pre-trained transformer LM and non pre-trained transformer LM summarizing the 3000 texts where transformer LM score 13.1 ROUGE2 and pre-trained encoder-decoder models score 2.3 ROUGE-2.

In [20], the authors proposed multi document text summarization architecture on WikiSum dataset by the transformer solution which is conveyed with a flat transformer for summarizing 3000 sentences into token. In summary, quality rating represented the help of AMT to generate a hierarchical transformer proposed paragraph position (PP), Multi-head pooling(MP), Global transformer layer (GT). In [21], the author introduced a sequence-to-sequence NMT which is visualization attention based text summarization by the transforming token using GPT-2 tool and BERT model. In [22], the authors made an abstractive text summarization method that got optimum results from constructed two side contrastive attention mechanisms on Chinese-English language with effective transformer baseline. In [23], the authors declined abstract summarization on the basis of 4 types of documents by the TLM using LSTM seq2seq propagation in which ROUGH reported at 0.24 reached with 95% confidence. In [24], other authors reported a NMT with self-attention transformer depending on the multiple language German-English pairs from which BLEU score is 3 point using fixed encoding decoding pattern.

Jinpeng et al [25] proposed Multi-head attention with the seq2seq model for automatic summarization which considers the previous words when creating new words to avoid unnecessary word repetition. Song Xu et al [26] employed a transformer base an assistive system over the mechanism of Abstractive summarization where the importance of each sound source is identified using a Self-Attention layer. By calculating the score of each source word they add it to the module. The Multi-Attention learning framework is proposed to summarize neural networks and return estimates to direct observation where the weight of attention is considered as the solvent information [27]. In [28], the authors proposed Transformer based Attention Mechanism, which the score of the state-of-art BLEU is 28.4 for English to German translation and for English to French is 41.0. Minh-Thang et al [29] research the global approach of attention mechanism that always attends the source word and local attention [28] that acts as a subset of the source word where the BLEU point is 5.0 over non-attentional systems. The authors report the current state-of-art attention architecture of WMT'15 English to German languages, in which the BLEU point is 25.9 and the improved BLEU point is 1.0 backed by NMT & n-gram ranker.

In [30], the authors propose a probabilistic and optimization view of the attention mechanism and assert both upper-lower and/or doubly-normalized attention mechanisms. Fredrik et al [31] has shown that the Transformer model performs very well in the RNN based attention model which is related to machine translation and also compares it with LSTM-based encoder decoder with attention on abstractive summarization. Puruso Muhammad et al [32] focused on abstractive text summarization based on local attention (i.e. subset of input terms) which give higher ROUGE-2 in LSTM model. In [33], the authors report on how to advance the state of art by improving the neural source code by shortening it. The abstracts of the term were used in the context of the file of the subroutines in the study with the help of the Attention mechanism system.

III. Experimental Methodology

In the recent trend Transformer has a large community addressing with encoder decoder that has transformed a self-attention mechanism [28]. Self-attention generates longer sequence translation where input Bangla text to decode the text summarization. NMT requires a necessary path to process data onto a transfer model sequence that shows the below needs of the architecture over maximizing the longest match sequence.

A. Training Parameters Preprocessing

Data preprocessing can transmute raw data into obligate and efficient formats. Some steps are followed to pre-process datasets. We first used construction in the description of the summary and text. Excluding constructions, full form has been used. After cleaning the construction, cleaned the texts, that is, deleted the unwanted characters [34], encoded in appropriate format, tokenized [35], PAD sequence and removed the stop word. Removed redundant components from the texts using regular expression.

B. NMT Architectures

1) Hierarchical seq2seq Learning:

NMT occurs to outperform response by the transformer when it modulates the attention network in which contextually processes sequence-to-sequence of sentences [4] [15] [41]. According to the attention procedure encoder

encoded with the original sentence and the attention model of the right portion decoder decodes the sentence with an expected output which is known as target output. In order to the encoder maintain sequence by the layer where the hidden vector produces a multi head of word. This summarization approach connected as our previous work is Bengali text abstractive summarization in order to Bengali text converted into Bengali summarization through the RNN model for encoding and bidirectional LSTM model for decoding (Abu Kaiser Mohammad Masum, 2019). To the best of study we focused the Bengali text for better enhancement on behalf of applied with transformer model to get maximum length from the previous work.

2)Transformer–Self Attention based Network

Self-attention based neural network transformer defined with a model by encoding decoding with Layer. Each layer model trains position wise feed-forward neural network for joint the input vector and jumps into the next two word vector. According to the sequence labeling that has no fitting with CNN and LSTM networks in RNNs fitted for the sequence such as sequence mapping, modeling and speech identification but not fitted for the longest sequence any time RNN deteriorates the translation quality. Above all difficulties for parallelization [15] [36] are solved in one way using a transformer where feed-forward network elaborate position wise. Figure-1 the transformer method conducting a self-attention network in which contains 6 numbers of encoders and decoders. Particularly encoder set up with two sub layers of Multi head Attention for mapping the word head by the attention of Softmax function [15], [20].

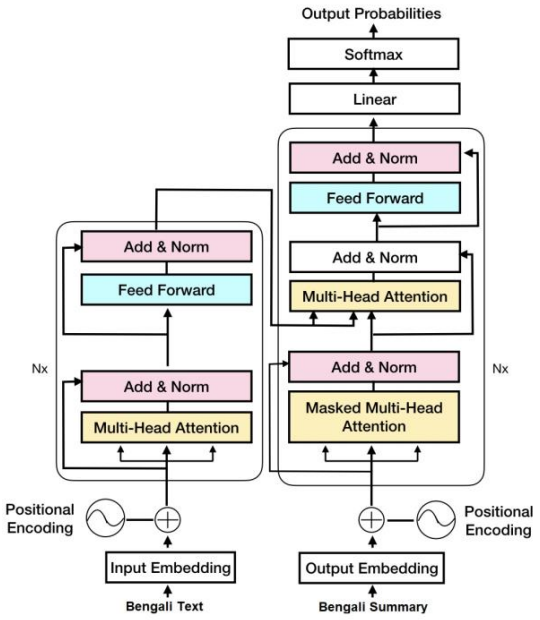


Figure_1: NMT Transformer model self-attention longest sequence referred as Vaswani et al. [28]

Self-Attention : Self-attention level counts as a first appearance of the encoder input, at the same time it focuses on the other word of the sequence although it defines only processes of fixed words. Moreover to search for information that modelling by Multi-head attention level which information is related to input sequence. Added with each input word is received such as vector of size 512 accordingly first self-attention layer. Generally, such the vectors of the input word are called " word embedding ". In order to scale dot-product having the capability of faster and space sequence as well as "Additive Attention" efficiency too [28] [35] [20]. Throughout the transformer model attracting for the self-attention. Nevertheless, "additive attention" makes the outperforms over the point-product attention on behalf of scaling subtraction to any longer sequences [28] [35]. Long- distance dependencies learn easily through the self focus level that surpass other existing models.

Multi-Head Attention : Multi-head attention have used to 8 parallel attention levels or heads that we used to train our model. Multi-head attention allows the model to participate collectively from sub-spaces of different delegations based on different residences [28]. In place of focusing on a single dimension that the transformer level inflicts attention to every noise emerging on which type of source of sequence existency [35] [20].

Position-wise Feed-Forward Networks : In attention mechanism model each layer mapping with different layers where the encoding layer fully relates to the feed-forward network with the opponent connected by the decoder. Feed-Forward Networks [15] [20] consists of a deeper transformation with Relu activation function where it carries the transformation to the next sub words. In Figure-2 Feed-Forward Networks reproduce the maximum weight between the two add and norm sublayers and this network continues by the each sub layers with new identification. In a sense Position-wise feed forward produces the maximum match of a word from the source of Bangla sentence. Inner layer set dimensionality as of 2048 on behalf of the base model where input and output both sequences contained 512.

Positional Encoding : "Positional Encoding or Embedding that has an ability to extrapolate the model. Accordingly, neural networks both in encoding and decoding produce identical results or words and embedding with the sequence of longer lengths during the process of input embedding. In the model section positional encoding saturated at the quadrant section in the decoder and encoder [15] [20]. Sinusoidal function may find the maximum path of input sequences.

C. Experimental Settings

We used the Neural Machine Translation workbox to train the method [35] [37]. The NMT workbox uses a variety of neural network systems. Before the loaded dataset we fixed some tools like- initialized maximum sentence length with 300, for preprocessing use maximum number of samples 2000, batch size 32 which is used to estimate the error gradient before the model is updated, buffer size 20,000 which means how long a computer takes to perform the task. Transformer base neural networks have been used in the training process with word embedding dimension 256, hidden unit 512, number of layer 2 [35], multi-head attention size 8 (number of multi-head attention in both encoder & decoder layer is 4), epochs 100 and the model moreover uses dropout mask 0.1 and 0.3 gradually. This

passes through the transformer encoder decoder layer. We also use a sample transformer layer in NN with vocabulary size 2238, number of layer 4, hidden unit 512, number of head 4 and dropout mask 0.3, optimizer Adam [38] and learning rate 0.9. The improvement set is utilized in optimizing the criteria of the representation instructed utilizing the instruction set. The test result was computed with the use of Bilingual Evaluation Understudy (BLEU) [28] [29] [39].

IV. Results & Discussion

A. Results

Dataset Refinement : Experiment procedure happened by our self made Bangla texts in which declare - Original Input to Original summary and Response Summary. Total data contained in the dataset is 1026. The machine will adopt summaries with the original input from the dataset and determine the Response summaries. For the parameters we use the attention based transformer model. From the Facebook posts we have collected different categories data and summarized them for our research purpose from that data [34].

Scaled dot-product Attention : That has used in the Transformer where query, key and value vectors are explained gradually. The softmax function distributes transpose function the weight of attention which works axis wise from the source location to the encoder and decoder. Attention based transformers carry self-attention, multi-head accomplishment and concatenate it simultaneously [22].

Masking More Attention Weights for Deriving the Attention mechanism : Regarding the transformer model, a mask [22] is used to focus attention scores (substituted with 1e-9) before multiplying the standard values by the matrix. In multi-head attention, padding tokens are clearly ignored with their masks. This is related to the parameter of padding-mask. Because of self-attention, in multi-head attention blocks used in decoders, the attention-mask is used to force predictions to appear in tokens in previous positions, so that the model can use them spontaneously during the guesswork. Masking all pad tokens in the sequence batch ensures that the model will not use padding as input and it indicates where the value of the pad is 0: it outputs 1 and otherwise 0. Look-ahead masks are used to mask future tokens in a sequence. In other words, the mask indicates that no entries should be used.

In Table-1, we show the result of the BLEU which is an algorithm for appreciating the attribute of text that translates from one natural language to another through machine translation and the score of BLEU always remains between 0 & 1. The text gives an indication of how logical the reference is to the text and a more accurate summary if there is a value near 1. BLEU accurately calculates metrics using n-gram.

	BLEU			
	1-Gram	**2-Gram**	**3-Gram**	**4-Gram**
Score	0.42	0.48	0.48	0.47
	0.42	0.43	0.44	0.45

Table_1: Performance summary in terms of BLEU aggregate and 1-4 [33]

Neural machine translation (NMT) from Bengali to Bengali summarization achieves best BLEU [3] score from 4-th gram which is 0.48. From the table we can see that we've achieved superior performance using NMT emerging access. Summarization from the transformer base model and the summaries we have created are very much connected to the evaluation.

For text summarizing we have used Bengali [33] text data. Text summarization is a process of shortening a long piece of text. The purpose is to create a compatible and affluent summary with the key points described in the record. Abstractive text summarization means creating a summary from text with its main idea. Here (from table-2), original input is the text from which by taking the main idea we generate the summary. When training the general transformer architecture in very long lessons, the model must learn that it individuals with self-attention positions should focus on closer sounds rather than distant sounds encoding. After MT, from the long text and summary it'll automatically generate the output which is response summary. In Table-2: we provide some examples on Bengali [33] [36] abstractive text summarization.

INPUT	দেরিতে হবে কিন্তু ঠিকই হবে তুমি যা চাও তাই হবে, মনে রেখো তোমার সময়টা খারাপ তোমার জীবনটা নয়। শুধু অপেক্ষা কর সময় সবকিছু ফিরিয়ে দিবে।
MACHINE SUMMARY	ধর্যের ফল মিষ্টি হয়
INPUT	তুমি কুকুর ভালবাসো , তুমি গান ভালোবাসো, তুমি কবিতা পড়, রাত বিরাতে মুগ্ধ হয়ে চাঁদ দেখো, তুমি এই শর্টকাটের যুগে চিঠি পড়, রাস্তায় হাঁটতে ভালোলাগে, তুমি মুখচোরা ? তাহলে তুমি আবেগী গাধা আর বিশ্বাস কর তোমার কপালে কঠিন দুঃখ আছে।
MACHINE SUMMARY	ছোটখাটো জিনিস এর মধ্যে ভালোবাসা থাকে

Table_2: Sample of Summary

B. Discussion

The NMT that has a need for speech translation, paraphrasing, language modeling extracted by the input text. Transformer with encoding and decoding is a method of auto tuning in which Transformer + Attention is evaluated over BLEU score insight of 0.48. BLEU score measure with its weight behaviour. N-grams weight shows a sequence of word length by its own parallelization. Transformer + encoder input sequence specifies the word length by the number of head its call "word embedding" using a padding mask requires maximum output to look ahead to the pointer of words. It takes attention + inputs produce vectors of word in the same way but also a contrastive way as decoding. Transformer + decoder fetch the multi head attention applying with positional encoding in which maximize the Attention + output such as input to text, text to original summary and finally it converts into response summary by this 8 parallel attention hidden layers where decoder divide into each 2 layer with multi head attention setting d-model value is 128 which covered 512 unit in each layer. Dimensionality of learning rate reduces the head cost by the matches with single-head attention. Sparcical Categorical Cross Entropy productive method for loss function that loss function estimates the value is 0.217 with the learning rate and 100 epoch. Sparcical Categorical Cross Entropy has

specialization in which is useful for longest sequence vector connection. After that model visualization loss rate is 0.217 in figure_2 described and plotted a curve by learning rate and train step. 4 grams weighted word sequence 0.25 predict best output where BLEU score is 0.48 that means our NMT model tends to paraphrase longer sequences. N-grams as the grams increase the number of 2,3,4 the added word also copied from the input sentences. Large chunks of sentences abstract summarization great for model proficiency.

INPUT	ফেব্রুয়ারি মাস জুড়ে এই সেই কতো ডে কাপল গুলার জন্য।তা আমগো জন্য একটা "বিরিয়ানি ডে" ও তো রাখতে পারতেন সাথে? খালি ডাবল গো কথা ভাবলেই চলবো? আমাদের মতোন বিরিয়ানি লাভার দের কথাটা কেউ মাথায় রাখলোনাহ।হায় আফসুস!
MACHINE SUMMARY	বিরিয়ানি লাভারদের কথা কেউ মাথায় রাখলো নাহ।
INPUT	দোষ না করেও যে স্যরি বলে, তাকে আগলে রাখুন! মানুষ টা চায় আপনি তার কাছেই থাকুন! দোষ না করেও স্যরি বলা ভীষণ কঠিন কাজ! যে এই কাজ করতে পারে তাকে আগলে রাখাটা আপনার দায়িত্ব! মানুষটাকে হারিয়ে ফেললে আফসোস এ পুড়বেন! নিশ্চিত ভাবেই!
MACHINE SUMMARY	দোষ না করেও যে স্যরি বলে, তাকে ধরে রাখুন।

Table_3: Original input to output summary

Batch Size	RNN Size	Probability	Learning Rate	EPOCHS
2	256	0.75	0.001	70

Table (1) comparison of [34]

Batch Size	Units	D-model	Learning Rate	EPOCHS
32	512	128	0.0014	100

Table (2) transform model acceptance

INPUT	আমি এটুকুই চাই এই অবহেলিত জনগোষ্ঠী যেন আর অবহেলার শিকার না হয়। - প্রধানমন্ত্রী শেখ হাসিনা পর্যায়ক্রমে দেশের আটটি বিভাগীয় শহরে বৃহৎ পরিসরে পিতামাতা ও অভিভাবকহীন নিউনরো-ডেভেলপমেন্ট প্রতিবন্ধি মেয়েদের জন্য পরিচর্যা কেন্দ্র স্থাপন করা হবে। এসব কেন্দ্রে তাদের শিক্ষা, প্রশিক্ষণ, চিকিৎসা, খেলাধুলাসহ সব সুবিধা অন্তর্ভুক্ত থাকবে।
MACHINE SUMMARY	প্রতিবন্ধি মেয়েদের জন্য পরিচর্যা কেন্দ্র স্থাপন করা হবে
INPUT	অনেক অনেক দিন পর তোর হাটের আলতো ছোঁয়া টা আজ আবার পেলাম, অনেক ইচ্ছা ছিল তোর হাতটা ধরে সূর্য উঠার ঠিক আগ মুহূর্তে আর একবার এই ব্যস্ত শহরের অলিগলি রাস্তা গুলোতে ঘুরবো, কিন্তু ভাবিনি সেই দিনটা হবে!!! অনেক ধন্যবাদ "নীলাম্বরী" এই সকাল টা আমার করে দেওয়ার জন্য।
MACHINE SUMMARY	জাতি হিসাবে আমরা বড়ই অসভ্য

Table (3) comparative output summary of [34]

Table_4: Abstractive text summarization before [34] and after adding structured transformer model

In authors [34] that had been done an approach of prioritize in Bangla published a sequence to sequence RNN study where employed by the Bengali summarization by the method of RNN. Regarding the summarization approach technically shows the same way summary of any text using RNN but not with the BLEU score. In order to this abstractive method with attention newly added BLEU score such as accurate and efficient strong summary representation. Attention based approach that shows the machine optimization over compiling process the reason behind is attention mechanism more sufficient to cover up larger corpus of Bangla language rather BLEU is the monolingual and bilingual groups are created as normalized scores and more specific. BLEU score near to 0.41-0.47 that has been necessary to analysis word gap concern with proper summary. In Table (3)[34] summary what have produce by the input tagged words and then output where sometimes occur wrong output summary or in the noise are getting then RNN supposed to a unclear summary. However, thus comparatively transformer model learned Bengali approach in which estimate summarization by the attention mechanism. In table-4 that we visualize comparison between our previous approach and transformer model approach parameter where learning rate, epochs and batch size produce different output. So far Transformer model has been defended like a sequential technique where each of part paraphrasing the word for making sentence and this has been a new technique with transformer model which had employed effectively for summary correctness. Thus, the originality of the proposal is the abstractive summarization claimed as a new technique based on the Bangla Language where in Bangla paragraph have produced one line summary and tropically, this invention explore original Bangla language uses.

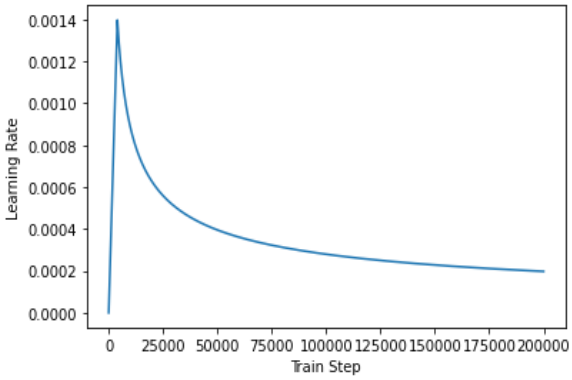


Figure_2: Transformer model learning point scale.

V. CONCLUSION & FUTURE WORK

A structured seq2seq process attention mechanism in order to explicit the Bangla language to Bangla text summarization. While input file analyzing the number of files was 1026 Bangla summary such as input as NMT. In this section we concluded that transformers are an effective solution for language modeling. Longer sequence tuned by removing noisy translation paraphrasing the next word using multi-head attention, positional encoding, Feed-forward neural network. Thereby this conceptual based method maximizes the output after 100 epoch cover and auto tuned by the 4-grams weighted word length on 0.48 BLEU score. The word length overfitting loss calculated as 0.217 which

one is a low loss function rate that filters the unnecessary response. Encoder-Decoder removes the lexical ambiguity, producing an outperform of adjacency between token input. Future studies define that NMT techniques will remain more in which tend to humans in loop generation where process different languages. Will explore long document summarization by dynamically taking input into delivery as response summary.

VI. ACKNOWLEDGMENT

DIU_NLP & Machine learning Research Lab give support to accomplish our research. We are thankful to give us support together with facilities and guidance besides we are grateful to our Department of Computer Science & Engineering.

REFERENCES

[1] A. Waibel, A. N. Jain, A. E. McNair, H. Saito, A. Hauptmann, and J. Tebelskis, "Janus: A speech-to-speech translation system using connectionist and symbolic processing strategies", in Proc. of the ICASSP, 1991, pp. 793–796.

[2] O. Kuchaiev, B. Ginsburg, I. Gitman, V. Lavrukhin, J. Li, H. Nguyen, C. Case, and P. Micikevicius, "Mixed-precision training for nlp and speech recognition with open seq2seq", 2018

[3] S. Jean, O. Firat, K. Cho, R. Memisevic, and Y. Bengio, "Montreal neural machine translation systems for wmt-15," in Proc. of the 10th WSMT. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 134–140. [Online]. Available: http://aclweb.org/anthology/W15-3014.

[4] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[5] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in Proc. of the EMNLP, 2013, pp. 1700–1709.

[6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Advances in neural information processing systems, 2014, pp. 3104–3112.

[7] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," arXiv preprint arXiv:1409.1259, 2014.

[8] J. Francisca, M. M. Mia, and S. M. Rahman, "Adapting rule based machine translation from english to bangla,"IndianJournalofComputer Science and Engineering (IJCSE), vol. 2, no. 3, pp. 334–342, 2011.

[9] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, "On using very large target vocabulary for neural machine translation", arXiv preprint arXiv:1412.2007, 2014.

[10] M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba, "Addressing the rare word problem in neural machine translation", arXiv preprint arXiv:1410.8206, 2014.

[11] Brenden M. Lake, "Compositional generalization through meta sequence-to-sequence learning", NIPS, 2017.

[12] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey,MaximKrikun,YuanCao,QinGao,KlausMacherey, etal, "Google's Neural Machine translation system: Bridging the gap between human and machine translation", arXiv preprint arXiv:1609.08144, 2016.

[13] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio, " A structured self-attentive sentence embedding", arXiv preprint arXiv:1703.03130, 2017.

[14] S. Naskar, S. Bandyopadhyay, "Use of machine translation in india: Current status," AAMT Journal, pp. 25–31, 2005.

[15] S.Dasgupta,A.Wasif,andS.Azam,"An Optimal Way Of Machine Translation from english to bengali", 2004, pp. 648–653.

[16] G. K. Saha, "The e2b machine translation: a new approach to hlt," Ubiquity, vol. 2005, no. August, pp. 1–1, 2005.

[17] K. M. Salm, A. Salam, M. Khan, and T. Nishino, "Example based english-bengali machine translation using wordnet," 2009.

[18] T. Banerjee, A. Kunchukuttan, and P. Bhattacharyya, "Multilingual indian language translation system at wat 2018: Many-to-one phrasebased smt," in Proc. of the 5th Workshop on Asian Translation, 2018.

[19] Urvashi Khandelwal, Kevin Clark, Dan Jurafsky, Lukasz Kaiser, "Sample Efficient Text Summarization Using a Single Pre-Trained Transformer", arXiv:1905.08836v1 [cs.CL], 2019.

[20] Yang Liu, Mirella Lapata, "Hierarchical Transformers for Multi-Document Summarization", arXiv:1905.13164v1 [cs.CL], 2019.

[21] Jesse Vig, "A Multi-scale Visualization of Attention in the Transformer Model", arXiv:1906.05714v1 [cs.HC] , 2019.

[22] Xiangyu Duan,Hongfei Yu,Mingming Yin, Min Zhang, Weihua Luo, Yue Zhang, "Contrastive Attention Mechanism for Abstractive Sentence Summarization", arXiv:1910.13114v2 [cs.CL], 2019.

[23] Sandeep Subramanian, Raymond Li, Jonathan Pilault, Christopher Pal, "On Extractive and Abstractive Neural Document Summarization with Transformer Language Models", arXiv:1909.03186v2 [cs.CL], 2020.

[24] Alessandro Raganato,Yves Scherrer, Jorg Tiedemann, "Fixed Encoder, Self-Attention patterns Transformer-Based Machine Translation", arXiv:2002.10260v2 [cs.CL], 2020.

[25] Jinpeng Li, Chuang Zhang, Xiaojun Chen, Yanan Cao, Pengcheng Liao and Peng Zhang, "Abstractive Text Summarization with Multi-Head Attention", IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019.

[26] Song Xu, HaoranLi, Peng Yuan,Youzheng Wu, Xiaodong He and Bowen Zhou, "Self-Attention Guided Copy Mechanism for Abstractive Summarization", Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1355–1362 July 5 - 10, 2020. c 2020 Association for Computational Linguistics.

[27] Piji Li, Lidong Bing, Zhongyu Wei, Wai Lam, "Salience Estimation with Multi-Attention Learning for AbstractiveText Summarization." arXiv:2004.03589v1 [cs.CL] 7 Apr 2020.

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need", 31st Conference on Neural Information Processing Systems (NIPS 2017),2017.
[29] Minh-Thang Luong, Hieu Pham, Christopher D.Manning, "Effective Approaches to Attention-based Neural Machine Translation", Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1412–1421, Lisbon, Portugal, 17-21 September 2015.
[30] "DOUBLY-NORMALIZED ATTENTION", Under review as a conference paper at ICLR 2020.
[31] Fredrik Jonsson, "Evaluation of the Transformer Model for Abstractive Text Summarization", DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING, SECOND CYCLE, 2019.
[32] Puruso Muhammad Hanunggul, Suyanto Suyanto, "The Impact of Local Attention in LSTM for Abstractive Text Summarization", june 15, 2020(ISRITI).
[33] Shakib Haque, Alexander LeClair, Lingfei Wu, Collin McMillan, "Improved Automatic Summarization of Subroutines via Attention File Context", MSR '20, October 5–6, 2020, Seoul, Republic of Korea ©2020 Association for Computing Machinery, https://doi.org/10.1145/3379597.3387449 , arXiv:2004.04881v1 [cs.SE] 10 Apr 2020.
[34] Md Ashraful Islam Talukder, Sheikh Abujar, Abu Kaisar Mohammad Masum, Fahad Faisal, Syed Akhter Hossain, "Bengali abstractive text summarization using sequence to sequence RNNs", 10th ICCCNT 2019 July 6-8, 2019 , IIT - Kanpur, Kanpur, India.
[35] Md. Arid Hasan, Firoj Alam, Shammur Absar Chowdhury, Naira Khan, "Neural Machine Translation for the Bangla-English Language Pair", 2019 22nd International Conference of Computer and Information Technology (ICCIT), 18-20 December, 2019.
[36] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multi task learners", OpenAI Blog, vol. 1, p. 8, 2019.
[37] Guillaume Kleiny, Yoon Kim, Yuntian Deng, Jean Senellarty, Alexander M. Rush, "OpenNMT: Open-Source Toolkit for Neural Machine Translation", arXiv:1701.02810v2 [cs.CL] 6 Mar 2017.
[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[39] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," in Proc. of the 40th ACL, ser. ACL '02. Stroudsburg, PA, USA: ACL, 2002, pp. 311–318. [Online]. Available: https://doi.org/10.3115/1073083.1073135
[40] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

Developed Credit Card Fraud Detection Alert Systems via Notification of LINE Application

Narumol Chumuang[1],Sansanee Hiranchan[2], Mahasak Ketcham[3], Worawut Yimyam[4], Patiyuth Pramkeaw[5], Sakchai Tangwannawit[6]

[1]Dep. of Digital Media Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand.
[2,3,6]Dep. of Management Information System, King Mongkut's University of Technology North Bangkok, Thailand.
[4]Dep. of Business Information Management, Phetchaburi Rajabhat University, Thailand.
[5]Media Technology Program, King Mongkut's University of Technology Thonburi, Thailand.

lecho20@hotmail.com1, s6207021910051@kmutnb.ac.th2, mahasak.k@it.kmutnb.ac.th3,
worawut_yimyam@hotmail.com4, patiyuth.pra@kmutt.ac.th 5,sakchai.t@it.kmutnb.ac.th6

***Abstract*— As nowadays, prevention of fraud is another important issue, researcher have initiated the idea of applying suspicious frauds in credit cards to line application. The objectives of this research are: 1) for developing the suspected credit card fraud via API LINE Notify. 2) Measure the accuracy of the developed system in the notification to prevent suspicious fraud credit card. The measurement method is comprised of five steps which are: 1) Analysis of work systems is a study and analysis of problems to determine needs. 2) System design is the process of designing research tools. 3) Developing a system is the process of developing research tools. 4) A test of the tools is executed 5) Summary of results, discussion results, and suggestions. The measurement results of efficiency, accuracy, and completeness of the data were in a very good level, equal to 86.67%. The results of the measurement of efficiency to the conditions set are very good, equal to 80.00 %. The results of the measurement on time very good, equal to 86.67%. In conclusion, the developed system accomplishes all research goals.**

Keywords— Credit card fraud, API, Line Notify

I. INTRODUCTION

Currently, the risk of the credit card business has changed according to user behavior and found that in Thailand and ASEAN, credit card fraud is 0.35-0.5% of the value of spending [1]. Credit card fraud has many channels both from people close to the merchant and online identity theft which users must be more careful. The nature of fraud is common, the criminals will collect ATM Slip from ATMs with quite a large amount to use to find important information in financial transactions such as date of birth [2]. ID card number by using different methods, such as falsely claiming to be civil servants, requesting civil registration information from administrative officials or search for the account number in ten digits and then try to transfer via online banking in order to know the name of the account owner after receiving the victim's information fraudulent identity cards are fake, using the name of the victim as the owner of the card, but attaching a picture of the criminal. and bring the said card to request to open an account deposit and new ATM cards of the same bank, but different branches as well as requesting to enable online banking services for all victims' deposit accounts to transfer all funds to the newly opened deposit account and using an ATM card to withdraw money [3]. Fraud Detection/Fraud Prevention is another important issue that will help banks to have virtual eyes to monitor suspicious activities that occur in the process of fraud or the intention that may present a risk of fraud in the future. But banks still don't have systems to detect those behaviors causing bank personnel to use behavioral analysis themselves or analyze their behavior before which if the behavior of corruption has changed will make the bank officials unable to verify which can cause damage, and according to a survey by ABA Banking Journal, 1 in 5 of the banks surveyed want to invest in developing a system to detect fraud continuously in the years to come [4]. The researcher foresaw that this problem occurred and created the idea to develop a credit card fraud monitoring system.

LINE application provides notification service from official accounts via LINE Notify, which can connect to web services with a variety of services and receive group notifications. Applications require a Line account and then can apply for LINE Notify. LINE is the most important marketing method in the digital age because it can easily reach users via mobile phones, both operating systems can use IOS and android - including personal computers which for Thailand has grown rapidly to become a major market with more than 33 million users because LINE has the advantage of free to use, easy to use. There are variety of features allowing LINE to grow faster than Facebook and Twitter in just 13 months of launch, reaching 50 million users worldwide while Facebook and Twitter took up to 36 months after launch to have the same number of users and currently has more than 490 million LINE users worldwide [5].

Therefore, based on the problems and benefits from the study above, we has the idea to use LINE Notify technology together with the concept that focuses on credit card fraud prevention by analyzing the risk behavior of credit card fraud, sending data from the API LINE Notify through the LINE group that was created to check for the fraudulent transactions of the card fraud prevention officers of Krung-Thai Credit Card Public Company Limited. Popular social media is widely used and provides quick and easy communication, allowing staff to take care of customers' credit cards to receive notifications quickly via mobile devices or personal computers to prevent the risk of credit card fraud immediately.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

II. LITERATURE REVIEW

A. Rule-based

It is a pre-defined rule consisting of rules for working in the IF-THEN format. Users can add facts so that rules can summarize the problems and apply those rules to process queries in the most efficient way[6]. The condition will be determined precisely because it creates a fixed condition or rule in order to check the data to match the specified conditions, therefore it is necessary to have the conditions to be covered if the conditions are not covered or have other conditions from the set conditions, the data will be inaccurate.

B. LINE Notify

LINE is an application for popular communication due to various capabilities and can work on many devices whether a smartphone, tablet, or even on a computer for the outstanding ability that makes LINE different from other communication applications, it is an image, cartoon character, an emotion called sticker. Its advantage is that it helps to reduce the amount of typing and helps to create newness in conversation as well. The services provided by LINE in the form of APIs for developers can be used to further develop projects that need to send notifications to our groups or personal accounts [5].

C. Type of credit card fraud

1) Lost / Stolen. The nature of the fraud found the information on the front and back of the card has been recorded by the cashier staff and used to make e-commerce transactions.

2) Counterfeit. Common corruption characteristics the information on the magnetic stripe is copied through skimmer to create fake cards and bring the card to use abroad.

3) E-Commerce Fraud. The type of fraud that customers receive e-mail phishing. Fraudulent credit card information updated and it was brought to a transaction (fake website) or a customer received an e-mail trick to fill in credit card information in exchange for a reward and was taken to a transaction (click bait website).

4) Account Take Over (OTA). Common corruption characteristics the fraudsters contact to change the mobile number in the system to their own in order to receive the OTP code to shop online.

5) Fake QR there are two common types of fraud as Malicious QR It is a QR that embeds malware in a QR. Which when scanning the QR will infect the mobile device with malware. Fake QR It is a fake QR code where fraudsters create a fake QR code to replace the merchant's QR code. Causing the payment to be transferred to the fraud instead of the merchant's account. [1].

6) Copy data from the card's magnetic stripe by a skimmer machine installed at the ATM. Common corruption characteristics criminals often install a skimmer machine at the card slot of the ATM to copy the information from the card and equipped with a numeric keypad to save the password that the victim pressed or may install a tiny camera to secretly look at the password.

7) Forgery of credit card application documents common corruption characteristics criminals may forge or use the victim's personal documents, such as a stolen copy of their ID card. Then apply it to apply for a credit card or notify change of address, change card by informing financial institutions to send documents and newly issued card directly to criminals when receiving a credit card, it will be spent on behalf of the company [2].

8) Stealing information from ATM slip, the common fraud criminals will keep the ATM slip at ATMs with relatively large balances to find important information in financial transactions such as date of birth. ID card number by using different methods, for example, impersonating a government official to request information from the civil registry from administrative officials. or find the account number to complete 10 digits and take it to try online bank transfer to know the account holder's name. Once the victim's information was obtained criminals will forge a fake government ID card using the name of the victim, owning the card, but attach a picture of the criminals. Then bring the said card to open a deposit account and make a new ATM card of the same bank but at different branches. As well as requesting the activation of online banking services with all victims' deposit accounts To transfer all funds to the newly opened deposit account Then use an ATM card to withdraw money.

D. Related work

J. Chaisuwan and N. Chumuang [7] presenta model for customer behavior prediction system based on credit risk of 1,100 corporate clients to forecast and decide by management in the credit decision of each customer on different conditions from the customer. Differentiate customer behavior with decision tree techniques And it was found that the organization was predictable and the credit decision was accurate or up to 85.82%.

C. Chang and H. Lin [8] on botnets and phishing threats with machine learning techniques or machine learning schemes with artificial neural network.

S. Kaewkorn, et al., [9], proposed the influence of the constituent factors in the use of the LINE application within the organization on the decision to use the LINE app within the organization of private employees in the district. Bangkok Through the survey research techniques, the behavior factors of using the LINE application within the organization, the decision to use the LINE application within the organization is a tool for data collection.

N. Balasupramanian [10] presented an online banking is the one most common service availed by almost all banking customers in the current era. Every second the banking organization, generate enormous amount of valuable data from their customers and their transactions. These valuable data need to be saved and analysed effectively using big data analytic techniques so as to get the necessary insights for the banking organizations.

B. Omair and A. Alturki [11] review on fraud detection methods in credit card transactions using techniques 1. Hidden Markov Model 2. Artificial Neural Network 3. Convolutional Neural Network (CNN) 4. Decision Tree 5. Rule based method.

III. METHODOLOGY

This paper is an applied research which our process is based on the following steps.

A. System analysis

The study and problem analysis process to determine the needs by studying and analyzing the problem, the researcher can get an idea of the problems arising from the prevention of damage from the fraudulent theft of credit card transactions, delays and not yet a system. Which alerts the staff to prevent credit card fraud in real-time to prevent damage from credit card fraud immediately.

Therefore, the researcher had the concept of using LINE Notify technology to work in conjunction with querying the behavior that is likely to be corrupted from the database by creating a connector to the database and creating conditions using rule based techniques to capture and forward the information to the connector with LINE Notify, it can be alerted to any potential or suspected action via LINE application causing staff to prevent credit card fraud able to receive notifications quickly through the mobile device or personal computer to detect and prevent credit card fraud as shown in Fig. 1.

Fig. 1. System operation.

B. System Design

From the analysis of the old work system that cannot detect risky or suspected fraudulent behavior. Therefore, design a new work system to be able to analyze the risky behavior or suspect fraud and transmit the data in real-time by designing the work of the system, there are steps as follows:

From the study of documents related to credit card fraud of the company Krungthai Credit Card Public Company Limited From January 2019 - December 2019, it was found that most of the fraud was caused by the following main behaviors.

1) Fraudulent behavior by pressing the wrong code to press money.

2) Fraudulent behavior by conducting transactions in a high balance.

3) Fraudulent behavior by conducting transactions during improper time.

4) Fraudulent behavior by conducting transactions from suspected merchants to be fraudulent.

5) Fraud behavior by pressing the wrong money press code and having transactions that can be exchanged for cash such as gold, mobile phones electronic device.

6) Fraudulent behavior by customers over 60 years old and conducting online shopping transactions.

7) The behavior of fraud through high- balance transactions and 30- day history of fraudulent transactions from merchants.

8) Fraud behavior by conducting multiple transactions in a period of one hour.

9) Corruption behavior by conducting transactions from many countries when they cannot travel.

10) Fraud behavior by using credit card magnetic stripe instead of credit card chip card from credit card machines that can accept credit cards with chip cards.

Functional design for investigating suspicious behavior of credit card fraud with a diagram showing the sequence of steps of the algorithm. In order to clearly see the workflow, use the flowchart symbol and show the direction of data flow when the data meets the specified conditions. Fig. 2, and Fig. 3 describe the working process of the algorithm. When the system searches for information and when the search data meets the conditions set the system will forward the information to the group line that has been created. if the data does not meet the specified conditions the system will continue to check against the conditions and if the data does not meet all the specified conditions then the verification will be ended and the work will be finished.

Data Dictionary Data Dictionary of Credit Card Fraud Surveillance System Alerts on LINE Application Can explain and display various details as in the table I and TABLE II.

Information from Fig. 4, the information display in the notification via the LINE application contains information. ID card information, name, surname, credit card limit, store name, transaction amount. Credit card fraud protection officers are required to use the ID card number to search for customer information by verifying the validity of the name. Last name and credit card limit information and store name are used in decision making in order to make it easier to investigate items that are subject to or suspected fraud.

TABLE I. GENERAL CUSTOMER INFORMATION

No.	Field Name	Type	Size	Description	Key
1	Card_No	Varchar	16	Customer credit card number	PK
2	ID	Varchar	13	Customer ID card number	-
3	Name	Varchar	50	Customer name	-
4	Sub_name	Varchar	50	Customer surname	-
5	Birthday	Varchar	9	Customer date of birth	-
6	Credit_Limit	Varchar	20	Customer credit card limit	-
7	Avail	Varchar	20	Client's remaining limit	-

TABLE II. CREDIT CARD TRANSACTION INFORMATION

No.	Field Name	Type	Size	Description	Key
1	Card_No	Varchar	16	Customer credit card number	PK
2	Date_Time	Varchar	20	Date and time of the transaction	-
3	Merchant _name	Varchar	100	Name of the store that performed the transaction	-
4	City	Varchar	15	Transaction city	-
5	Country	Varchar	2	Transaction country	-
6	MCC	Varchar	4	Types of merchants doing transactions	-
7	Amount	Varchar	20	Transaction amount	-
8	Response	Varchar	3	The results of the transaction	-
9	POS	Varchar	3	Transaction characteristics	-
10	Merchant_ID	Varchar	20	Store code for transaction	-
11	Terminal_ID	Varchar	20	Transaction ID	-

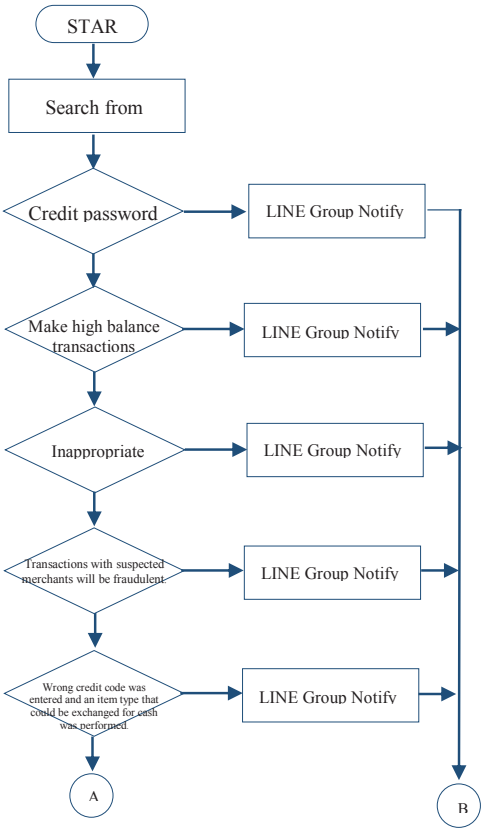


Fig. 2. Working chart of corrupt behavior detection conditions.

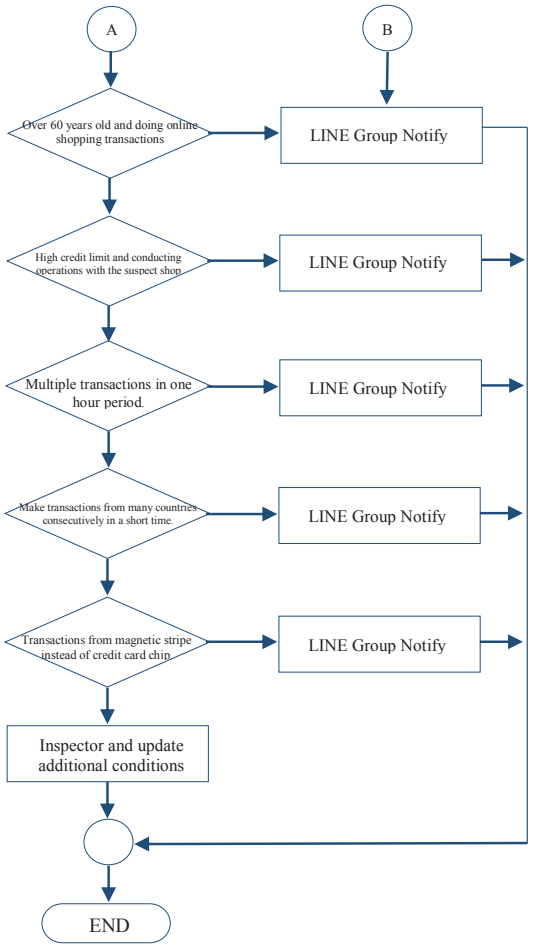


Fig. 3. Working chart of corrupt behavior detection conditions (cont.).

Fig. 4. The screen shows information notifications via the LINE application.

C. *Implementation*

After analyzing and designing the system, it makes it clearer to know the working process of the system. In the process of analyzing and designing the system, it will know the scope of the system development and know that there are processes.

- Acquiring an Access Token to use the API Line Notify, access the website. https://notify-bot.line.me/en/ Login to the system and select Generate token menu, then

enter the information in the Token name, which is the display name when prompted and select the chat room. Vim set to receive notifications. After that, press the Generate token button, the system will generate a Token, which must keep this data set. In order to contact the API Line Notify and the Line Notify system will not store the Token that has been created. User will have to generate a new token every time.

- Create files to connect to the database and control how the database searches.

- Create a file to connect to the database with your Username Password.

- Create a file to find information from the database with certain conditions.

- Design the system to automatically search for information from the database It is set to search for information every 5 minutes and search every day.

- The system record information from searching in the database.

- Send received information to Line Notify service.

- Control the transmission of information to the group line. By sending the information received from searching the database to the API Line Notify.

- Testing and Debugging at this stage it is a step that must be done in conjunction with the system development process. Due to the development of the system, there may be some modifications or changes to the system during development and system readiness is checked before the system is put into production and the system must be tested. In order to find flaws and fix them completely with usage tests as follows

- The credit card holder's information is correct and complete.
- Detects any behaviors that fall within the scope of fraud and correct the conditions.
- The duration of searching for information from the database and sending the information is not more than 5 minutes per time.

TABLE III. PERFORMANCE EVALUATION

Scoring Level		Significance
Qualitative	**Quantitative**	
Very good	80% – 100%	Performance is very good.
Good	70% – 79%	Performance is good.
Moderate	60% – 69%	Performance is moderate.
Little	50% – 59%	Performance is little.
Scanty	0% – 49%	Performance is scantly.

Performance evaluation Development of a surveillance system for credit card fraud through the notification of the LINE application. The researcher assessed the quality by responding to the test after testing the system. To be used for checking the efficiency and speed of the work process Is it meeting the objectives? If any errors are found, they will be corrected and tested again before proceeding to the next step. Then the results were obtained to find percent values to find the correctness by analyzing the data for conclusions by scoring according to the percentage comparison criteria.

IV. RESULT

Development of the analysis of credit card fraud risk behavior and the application of Line Notify technology to notify via Line application To prevent credit card fraud By researchers looking for errors that arise from the system Then test the operation of the system. The results of the research are as follows.

A. Credit card holder details are correct and complete.

From Table IV, the measure of performance of credit card holder data From testing, it was found that All results are valid for 13 times and not all results are correct 2 times, therefore, the results of the measurement of the credit card holder's information are correct and complete. when the percentage of accuracy is used, the result is equal to 100. 86.67 each when compared with the specified criteria, making it the most accurate level.

TABLE IV. TEST RESULTS FOR ACCURACY AND COMPLETENESS OF INFORMATION

Test item	Time	Results
The details of the credit card holder's information are correct and complete.	1	✗
	2	✓
	3	✓
	4	✓
	5	✓
	6	✓
	7	✗
	8	✓
	9	✓
	10	✓
	11	✓
	12	✓
	13	✓
	14	✓
	15	✓
	Percentage	86.67

B. Detects any behaviors that fall within the scope of fraud and correct the conditions.

From Table V., the detecting behaviors that fall within the scope of corruption are correct according to the specified conditions. From testing, it was found that All of the results were valid 12 times and the results were incorrect 3 times. Therefore, the results of measuring the effectiveness of detecting behaviors that fall within the scope of fraud were correct according to the specified conditions. When the percentage accuracy was used, the results were 80% when compared with the specified criteria, making it the most accurate level.

TABLE V. VALIDITY RESULTS ACCORDING TO THE SPECIFIED CONDITIONS

Test item	Time	Results
Detects any behaviors that fall within the scope of fraud and correct the conditions.	1	✓
	2	✓
	3	✗
	4	✓
	5	✓
	6	✓
	7	✗
	8	✓
	9	✓
	10	✗

Test item	Time	Results
	11	✓
	12	✓
	13	✓
	14	✓
	15	✓
	Percentage	80.00

C. The period of searching for information from the database and sending the information is not more than 5 minutes per time.

From Table VI, the time for searching the database and sending the data is not more than 5 minutes per time. All results were valid for 13 times and total incorrect results for 2 times. Therefore, the results of the measurement of the results of the time of searching the data from the database and sending the data not more than 5 minutes per time. Was equal to 86.67 percent when compared with the specified criteria, making it the most accurate level.

TABLE VI. DURATION TEST RESULTS

Test item	Time	Results
Detects any behaviors that fall within the scope of fraud and correct the conditions.	1	✓
	2	✓
	3	✓
	4	✓
	5	✓
	6	×
	7	✓
	8	✓
	9	✓
	10	✓
	11	✓
	12	✓
	13	✓
	14	×
	15	✓
	Percentage	86.67

V. CONCLUSION

This research proposes the development of a credit card fraud surveillance system through an online application alert using a Rule-based technique to analyze credit card fraud risk behavior. Its purpose is to develop a notification system for risky or suspected credit card fraud through LINE Application and to measure the accuracy of the alert to prevent the risk of credit card fraud or to prevent damage that Caused by credit card fraud Which after developing the system, the researcher has measured the performance of the system By conducting a system test to evaluate the results of 3 aspects: 1. The accuracy and completeness of the information. 2. The validity according to the specified conditions 3. Duration The evaluation results have a mean percentage of efficiency in each area which is very good. In conclusion, the assessment results are very good. And in accordance with the objectives Develop an alert system for risky or suspicious behavior of credit card fraud through the LINE application. And measure the accuracy of alerts to prevent the risk of credit card fraud or damage caused by credit card fraud, and in accordance with the research hypothesis, Able to correctly notify problems with a percentage greater than or equal to 80.

REFERENCES

[1] L. Changjian and H. Peng, "Credit Risk Assessment for Rural Credit Cooperatives Based on Improved Neural Network," 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), Changsha, 2017, pp. 227-230, doi: 10.1109/ICSGEA.2017.161.

[2] W. Yu and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," 2009 International Joint Conference on Artificial Intelligence, Hainan Island, 2009, pp. 353-356, doi: 10.1109/JCAI.2009.146.

[3] F. Kouser, Nagaratna, P. V. R., B. Sree and Ravikiran, "Highly Secure Multiple Account Bank Affinity Card-A Successor for ATM Card," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 115-119, doi: 10.1109/ICDI3C.2018.00033.

[4] K. Bobkowska, K. Nagaty and M. Przyborski, "Incorporating iris, fingerprint and face biometric for fraud prevention in e-passports using fuzzy vault," in IET Image Processing, vol. 13, no. 13, pp. 2516-2528, 14 11 2019, doi: 10.1049/iet-ipr.2019.0072.

[5] W. Boonsong, "Smart Intruder Notifying System Using NETPIE through Line Bot Based on Internet of Things Platform," 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 2019, pp. 2208-2211, doi: 10.1109/ICCC47050.2019.9064030.

[6] K. Vongprechakorn, N. Chumuang and A. Farooq, "Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045560.

[7] J. Chaisuwan and N. Chumuang, "Intelligent Credit Service Risk Predicting System Based on Customer's Behavior By Using Machine Learning," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045452.

[8] C. Chang and H. Lin, "On similarities of string and query sequence for DGA botnet detection," 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 104-109, doi: 10.1109/ICOIN.2018.8343094.

[9] S. Kaewkorn, C. Joochim, S. Prasertprasasna, C. Leartrussameejit, H. Kuhataparuks and A. Kunapinun, " Notifying problems of a machine by using Machine Learning," 2019 Research, Invention, and Innovation Congress (RI2C), Bangkok, Thailand, 2019, pp. 1-6, doi: 10.1109/RI2C48728.2019.8999912.

[10] N. Balasupramanian, B. G. Ephrem and I. S. Al-Barwani, "User pattern based online fraud detection and prevention using big data analytics and self organizing maps," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, 2017, pp. 691-694, doi: 10.1109/ICICICT1.2017.8342647.

[11] B. Omair and A. Alturki, "Taxonomy of Fraud Detection Metrics for Business Processes," in IEEE Access, vol. 8, pp. 71364- 71377, 2020, doi: 10.1109/ACCESS.2020.2987337

Face Detection System for Public Transport Service Based on Scale-Invariant Feature Transform

Narumol Chumuang[1], Sansanee Hiranchan[2], Mahasak Ketcham[3], Worawut Yimyam[4], Patiyuth Pramkeaw[5], Tanapon Jensuttiwetchakult[6]

[1]Dep. of Digital Media Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand.
[2,3,6]Dep. of Management Information System, King Mongkut's University of Technology North Bangkok, Thailand.
[4]Dep. of Business Information Management, Phetchaburi Rajabhat University, Thailand.
[5]Media Technology Program, King Mongkut's University of Technology Thonburi, Thailand.

lecho20@hotmail.com1, s6207021910051@kmutnb.ac.th2, mahasak.k@it.kmutnb.ac.th3,
worawut_yimyam@hotmail.com4, patiyuth.pra@kmutt.ac.th5, tanapon.j@it.kmutnb.ac.th6

Face Detection System for Public Transport Service Based on Scale-Invariant Feature Transform
Narumol Chumuang, Sansanee Hiranchan, Mahasak Ketcham, Worawut Yimyam,
Patiyuth Pramkeaw, Tanapon Jensuttiwetchakult

***Abstract*— This paper proposed to reduce the complaints about the use of public transport. We applying the principles of digital image processing with the Eigen face detection and the Scale-Invariant Feature Transform (SIFT) matching technique. The system shown that the face of the person who interested will be mark and detect. After that it do an emotional analysis and show the emotional results immediately. We evaluated the effectiveness of our system based on the accuracy for detecting human faces. In the experimental, total testing 100 times with both of the straight and inclined faces. The results are as follows: A person's face detection with a constant light is 90%, a 45-degree tilted face has a constant illumination of 79%, a 45-degree tilted face, a constant light of 55%, and a thin masked face section, it cannot be work.**

Keywords*— *Face Detection, Transport Service, Scale-Invariant Feature Transform, SIFT

I. INTRODUCTION

At present, there are more public transport as taxi car in Bangkok and major cities around the world [1]. Since in the year 1992 the government by the Ministry of Transport has a policy to increase the number of public taxi without limitation and due to the traffic conditions in Bangkok. That have a changed greatly both in terms of the volume of vehicles and the population, together with the encounter of traffic congestion [2]. They have an impact on travel affect a result, more and more people tend to use taxi. In each time people use public taxi, there will be problems more service complaints such as service provided by taxi, friendliness and manners of drivers while using the service or the driver, careless, thrilling driving, which at present, the agency that controls the service of taxi and related agencies [1],[3]. Lack of serious attention to solving such problems or it may be too slow to fix.

According to statistics of Public Relations and Corporate Communications, the Department of Land Transport has compiled statistics on complaints of taxi from October 2011 to April 2013, which has the number of complaints from passengers increasing. The transport statistics division, the division of planning, has considered the satisfaction of passengers a lot [1]. Therefore has been collected the subject of the complaint that the passenger has, Department of Land Transport, which is the most reported two issues: the bus driver shows impolite verbal behavior, reckless driving, and fearfulness. It can be seen that according to statistics, each year will the number is constantly increasing. There is no quick and immediate resolution of the problem, so the problem cannot be solved [4].

From the above problem, We have an idea to develop a system to detect emotions from different faces. From the face of the passenger and face recognition or sensing through the face of passengers to get a way to detect faces and feelings of passengers. Effective in practical use and to assess the emotional state of the passenger at that time from the first use of the service until reaching the destination and is immediately sent to the Department of Land Transport. In order to get the problem resolved immediately without delay, it can also be used to check customer satisfaction. The analysis behavior patterns of job interviewers. The analysis the emotion of people at different parties. The analysis the facial expressions or feelings of dumb people and examines the behavior of hotel customers.

Therefore, we are interested for studying the system of detecting facial emotions procedures and methods for performing face detection and detailed recognition. The usage facial recognition techniques for passenger emotions such as normal facial expressions, happiness, heartbreak, panic, anger, and drowsiness. In an implementation, we used an OpenCv with a Raspberry pi to create facial recognition technology and Eigen Face technique, which uses digital image processing technology to be used for verification to hope that the system can be used to assist in various services and to be able to use it for the most benefit.

II. LITERATURE REVIEW

A. Digital Image Processing

Digital image processing [5] is about the converting image data into digital format. In order to be able to bring this information through various processes with computer. The operation of the computer, the system for receiving data in or out of data is only in digital format. In digital image processing, once the system has received the image data, it will calculate and output as data that can be used instead of those digital images saving image data in computer memory [6]. This can be done by reserving the machine's memory in the form of an array variable Fig. 1 by the values in each channel of the array represents the properties of the pixel and the position of the array. Determines the position of the image point from the use of memory to store images in the manner

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

mentioned above. The image storage space can be calculated as $M \times N \times g$, where g is an integer representing the number of bits of data at each preview point. If g is equal to 8 bits, we can store the possible color difference up to 256 level, and the M and N values indicate the resolution of the image [5]-[7].

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

Fig. 1. Equation to calculate the brightness or color of each pixel.

The image processing is a process that includes a variety of techniques and methods for processing the numeric data of images, which can be applied to suit the image data being processed [8]. Image data is characterized by varying shapes, textures, colors and structures. It depends on the object and its surrounding environment. Digital image processing can be run as either hardware or software with the principles of image processing theory [9].

B. Face Recognition

The Face recognition system is designed to compare interested person's face with an existing face database [10],[11]. The algorithms used in the template generation and comparison procedures may vary from one person to another. System design of each system but whether there will be how does the algorithm work in the template construction phase and the comparison process? The overall work process of the system is still the same. Typically, a face recognition system consists of three steps [12]. The first on is face detection is the function of locating all the faces in the image. The second is feature extraction is a function of separating features on the face image and then store it in a database or compare it with the facial features feature information in case of retrieval process [13]. In the last is the part that carries information about different parts of the face image obtained from the second step [14]. Let's compare them with different features of face images in the database and then display the results of the face image that is most similar to each other, as shown in Fig. 2.

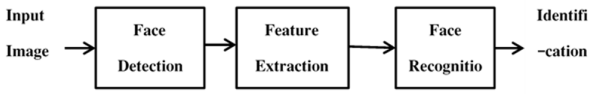


Fig. 2. Shows the main of the face recognition system.

1) Face recognition algorithm [6]

Knowledge- based methods is a method for determining the relationship between distinctive features on the face with fixed basic main positions and elements on the face [15] - [17]. For example, we study of hierarchical knowledge methods, which is divided into 3 hierarchies, as shown in Fig. 3.

- Level 1 consider the position of the overall probability of the face.
- Level 2 histogram analysis with the face's boarder.
- Level 3 analysis the internal features of the face such as symmetry of both eyes, vertical and horizontal position of the nose and mouth.
- Level 4 search base a zoning rule similar to Yang and Huang's method. The difference is that it looks at both the vertical and horizontal histograms, but does not eliminate the problem of the image with complex backgrounds and problems on a variety of faces including the posed that are different.

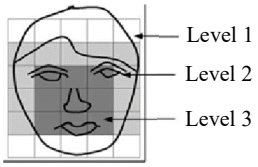


Fig. 3. Hierarchical knowledge methods.

2) Feature-based methods

It uses an algorithm to look at the features and structure of the face, as well as the dynamics of various elements of the image [18]. Use contrasts of light, shadows and contours as models to find faces. The model consisted of two less bright points (dark spots) to represent the eyes and three brightness points (weak points) to show the cheekbones and nose, and to find the correlation between the distance and the location of the differences to select a suitable face model [19], [20]. This system is limited in that when the shadows of the surroundings change, it will affect the search performance.

3) Template matching methods

It is a comparison of the desired image to the model structure of a standard face by collecting relational data independently of the parts [21]. For example, Sakai, Nagao and Fujibayashi, (1969) propose a sub-model plate structure. (subtemplates model) using a sobel filter to find the border to find the probability locations of various subsets on the face that best match with the subtemplate model to the position of the face. Later, a qualitative model for face pattern (QMF) was proposed, using the parameters of illumination and facial contour as a model of the face [22],[23]. Another popular technique is the principal component analysis (PCA) by studying the exact location of faces using basic geometric principles as shown in Fig 4.

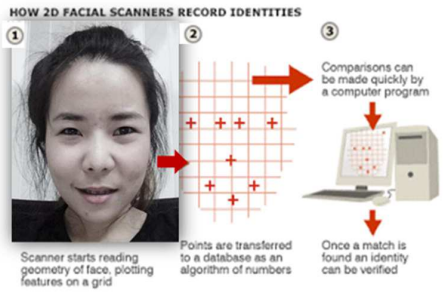


Fig. 4. Face model from the MPEG-4 database.

4) Appearance-based methods

This method is a visual comparison that must search and model structure of the face were learned and trained in the

system to recognize and apply knowledge in the database to consider [24].

5) Eigenface Methods

Basic facial composition analysis with a variety of facial statistical processes a characteristic face is a set of Eigen vectors that can be obtained from a covariance matrix, creating a face model that combines the characteristics of a sample face image to find the specific value of the elements on the face [25]. For example, Turk and Pentland would use a grayscale image and convert it to a vector to determine the eigenvalue and take the eigenvalue of the portrait sample. Let's create a unique face model to locate the position of the face. Distribution-Base Methods represent the distribution of sample data formats with pageline and non-page as the basis for decision making [26]. For example, Sung and Poggio take the Gaussian function around a group. The distribution of the sample mean is shown in Fig. 5.

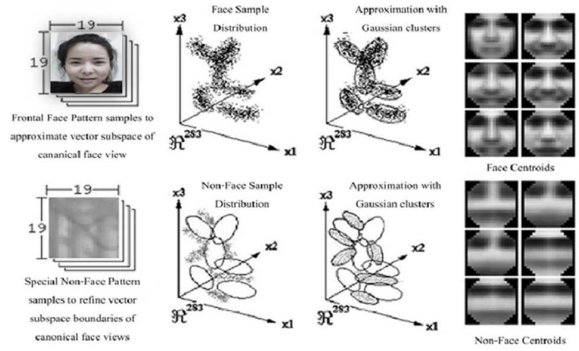


Fig. 5. Distribution of sample data.

III. RESEARCH METHODOLOGY

Implementation of the development of facial recognition system from the camera by image capture and image processing. The working method, divides the development process according to the system development cycle into the following steps.

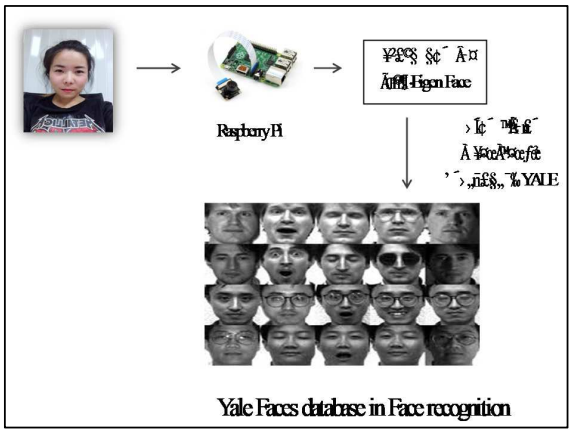


Fig. 6. The concept of the system.

A. Personal face dataset

The dataset of face image from 10 persons, each person will have 50 faces in color with different elements of the environment and different mood, light, contrast, resolution.The dataset used must be .jpg / pad extension with an image size of 300x300 pixels, the image must be straight-faced. Clearly and not blurry. Normal light quality showed in the example as in Fig. 7. The dataset tested, there are 500 images in total, divided into 350 images for training set, and 150 for test set.

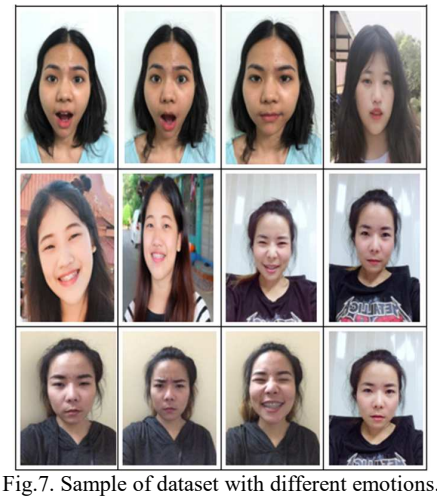


Fig.7. Sample of dataset with different emotions.

B. Design of Human Emotion Identification System

In this paper, the Eigen face was used as an indicator of human mood. In this topic, we have studied how to detect faces using the Haar-like feature and in the communication between hardware and software [27]. In the program, it looks like connecting devices as in Fig. 8. The first step, image processing part is divided as follows.

1. Prepare the equipment by connect a camera with a Raspberry Pi board.

2. Face detection after receiving face image data. The system will take preprocessing as:

2.1 Resize to standard image with 244 x 244 pixel.

2.2 Normalization with extraction of non-facial features and image enhancement to achieve a sharper image using the image enhancement as eq. (1).

$$G(i,j) = med[x(i+r, j+s) \in A(i,j) \in Z^2 \qquad (1)$$

where *G (i, j)* is the middle average filtering, it is a nonlinear filter. The advantages of image enhancement with center average filter are creates a blurring of the image so that the pixel intensity points are in the same tone [28].

2.3 Multi-perspective facial composition analysis (PCA) using Eigen face technique to represent vector data. The key work steps are perform converting the data matrix structure to row vectors. Calculates for the Eigen vector corresponding to the eigen value and then used to calculate the distinctive features of the image to be used for face recognition.

3. Face recognition and compare with face images in the database for matching.

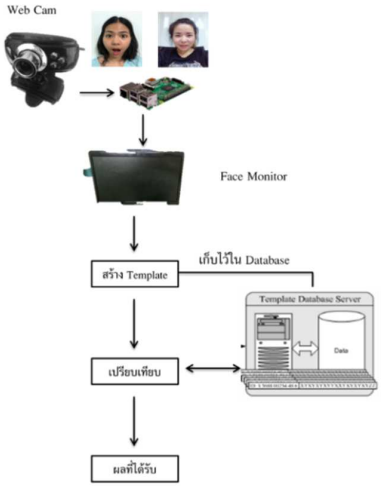


Fig. 8. The work flow of the system.

3.1 Face analysis and face recognition techniques using triangular geometry of the eye and mouth which is used to find the position of the eyes and mouth in Fig .8 by selecting the triangle with the least inclination weight as eq. (2).

$$ow(i,j,k) = \prod^2 e^{-3(1-cos^2(\theta_r(i,j,k)))}$$

$$\text{Cos}(\theta_r(i,j,k)) = \frac{\overrightarrow{u_r} \cdot \overrightarrow{v_r}}{\left\|\overrightarrow{v_r}\right\|} \quad (2)$$

where ow is the weight of the slope of the triangle

$\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ is the vector between the midpoint of *i*, *j* and *k*.

θ_1 is the angle between the vectors $\overrightarrow{u_1}$ and $\overrightarrow{v_1}$

θ_2 is the angle between the vectors $\overrightarrow{u_2}$ และ $\overrightarrow{v_2}$

$\overrightarrow{u_1}$ is a vector that originates at the midpoint of $\overline{ij}$ which is perpendicular to the vector $\overline{ij}$

$\overrightarrow{u_2}$ is a vector originating at k that is perpendicular to the horizontal.

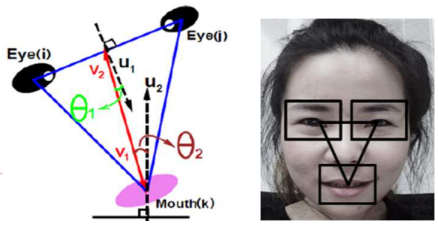


Fig. 8. Triangular geometric simulation of the eye and mouth.

3.2 SIFT consists of 4 important steps [29]:

3.2.1 The key point in dimensions, size and distance (Scale-space extreme detection) The first step of finding the key point of an image that does not depend on the image size. The search will be performed by changing the size of the image and blurring the image with the Gaussian function, as follows eq. (3).

$$L(x,y,\sigma) = G(x,y,\sigma)I(x,y)$$

$$G(x,y,\sigma) = \frac{1}{(2\pi\sigma^2)} e - (x^2,y^2)/2\sigma^2 \quad (3)$$

where (x,y) is the coordinate on the image.

$L(x,y,\sigma)$ is that the image is blurred with the Gaussian function

$G(x,y,\sigma)$ is a type filter with Gaussian function.

I is the original image

σ is the size of the blur variable, the more value the more the blur.

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G\,(x,y,\sigma))xI(x,y)$$
$$= L(x,y,k\sigma) - L(x,y,\sigma) \quad (4)$$

$D(x,y,\sigma)$ is parameters of image size that have been blurred by the Gaussian function will be compared to determine Difference-of-Gaussian: DOG).

3.2.2 Keypoint localization by the maximum and minimum values of the image obtained from the DOG process are determined by 8 neighbors [30] on the same layer as the point considered, and 9 on the upper level and nine on the lower level adjacent to the layer.

3.3.3 Orientation assignment to find the size and direction of the gradient (Gradient) around the key point in order to direct the key point using the eq. (5).

$$m(x,y) = \sqrt{((L(x,y)+1,y) - L(x-1,y))^2\,(L(x,y+1) - L(x,y-1))^2}$$
$$\theta(x,y) = tan^{-1}((L(x,y)+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \quad (5)$$

where $m(x,y)$ is the size of the gradient.

$\theta(x,y)$ is the direction of key point

3.3.4 Keypoint desciptor by finding the size of the gradient and the direction around the key point, then creating a 16x16 windows around the key point and then generating eight orientation histograms in each sub area. So we get a characteristic vector unique to each key point that is equal to 128 data [31].

3.3.4.1 The face images are compared with the face images in the database for data matching completed and bring the results to the monitor. LCD shows whether the face image shows the face according to the mood in the database or not.

Evaluation

The evaluation for the validity of our work is a statistical of emphasis on accuracy rate and comparative analysis of recognition performance. In the analysis of the accuracy, a image-to-image method is used to verify the meaning of the target with the real meaning derived from translations by professionals relevant to the content of each respective document. Then all the correct converted target images are collected to calculate the percentage of accuracy in the interpretation of each document with the formula [32] as the following eq. (6).

$$Accracy\ rate = \frac{\text{Number of Classified Correctly}}{\text{All of Classified}} \times 100\% \quad (6)$$

IV. EXPERIMENTAL AND RESULTS

This paper conducted to test the ability to recognize a person's emotions. The results were obtained from the testing of the sample dataset to collect data for the color, area and eye size as.

A. Results of individual's emotion feature

Eigen Face with SIFT, LBP and Fisiher techniques, which results from each of these algorithms. We will use image data obtained from install a webcam. It is a collection of real images from 10 persons, each 10 images, and then the images obtained will be processed for the accuracy of determining the characteristics of human emotions.

Fig.9. The results of detecting faces and emotions of the system.

TABLE I. A COMPARISON OF EIGEN FACE AND SIFT, LBP AND FISIHER.

Method	Person 1	Person 2	Person 3	Accuracy (%)
Eigen face & SIFT	94.00%	93.00%	95.00%	94%
LBP	90.00%	88.00%	89.00%	89%
Fisiher	74.00%	78.00%	72%	74.66%

The experiment the dominant characteristics of the individual emotions in different ways of all 10 individuals, 10 images were found that Eigen Face and SIFT technique had the most accurate results.

B. System test results.

The results of multiple angles, such as a straight face, a 45-degree tilted face, and a 90-degree tilted face, the test was divided into two parts and total of 100 times in each part as follows.

TABLE II. THE RESULTS OF EMOTION DETECTION WITH BRIGHTNESS

Degree and Distance	Accuracy (%)	
	Normal Light	Low Light
Straight face at a distance of 30 cm.	87	72
Straight face at a distance of 50 cm.	80	0
Side face 45-degree at a distance of 30 cm.	75	67
Side face 45-degree at a distance of 50 cm.	70	0
Side face 90-degree at a distance of 30 cm.	43	35
Side face 90-degree at a distance of 50 cm.	0	0

In our work, the human face recognition system has been developed to indicate the emotions expressed through the faces for use with public transport. It can detect emotions from human faces using various image processing technologies. Evaluation of the performance of the developed system can be put into practice and accurate face recognition.

V. CONCLUSION

From our work, it was found that there were still problems among passengers who were satisfied with the public transport service. Therefore, to increase efficiency, improve and reduce defects from service. The image processing techniques were applied in the work by designing and developing a face detection system for using Eigen face and SIFT matching technique, development language is Python 3 and library used is OpenCV to detect a person's face and alert the Department of Transportation. According to the objectives and scope set this will show the results of detecting a person's face through the web cam and give an alert to the LCD display screen and can also clearly distinguish the emotions of people with facial expressions. The system can work accurately and precisely to detect the faces of people who express emotions. Detecting a face in constant light accounted for 90 %t of the face detection tilted 45-degrees with constant light accounted for 79 % and constant-light 90-degree tilted face detection. For 55 %, but still found that the system still has errors in detecting human faces There are certain faces that the system will not be able to identify with a person's emotions. This error may be due to the fact that the person's face is too tilted, which may make the face appear unclear or there is insufficient light in the area, and this system can be used to immediately reduce the problem of solving the satisfaction of the public bus service, and the system can be applied to the driver to see the characteristics.

However, the Eigen Face method has its limitations Obscuring the face of a person Here it is the masking of a person's face, which the system can detect accurately only if that person's face must not be obscured by an object, for example, that person must not wear a hat or sunglasses in order to detect the face and express emotions of that person but if the person is wearing a hat or wearing sunglasses This

will make the system unable to detect the person's face and express their emotions. Facial features of a person, the system will detect a person's face with a clear angle, such as a straight face, 45-degree tilted face, and a 90-degree that is, if the person has an excessively tilted detection face. This may result in the system being unable to detect the person's face, or if the person's face was captured, the system's accuracy in expressing emotions decreased.

REFERENCES

[1] C. Rujichansiri, K. Kungcharoen, P. Palangsantikul, P. Porouhan and W. Premchaiswadi, "Pair Selection of Appropriate Taxi Drivers Using Social Network Analysis Models," 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, 2018, pp. 1-5, doi: 10.1109/ICTKE.2018.8612363.

[2] S. Dey and M. Rahman, "Application of Image Processing and Data Mining Techniques for Traffic Density Estimation and Prediction," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 2019, pp. 1-6, doi: 10.1109/ICACCP.2019.8882878.

[3] R. Drakoulis et al., "A Gamified Flexible Transportation Service for On-Demand Public Transport," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 3, pp. 921-933, March 2018, doi: 10.1109/TITS.2018.2791643.

[4] Q. Man, Q. Wu, P. Dong and X. Yang, "Analysis of urban land expansion and simulation of urban land use in Loukou City," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 5109-5112, doi: 10.1109/IGARSS.2017.8128152.

[5] H. Lin, J. Si and G. P. Abousleman, "Orthogonal Rotation-Invariant Moments for Digital Image Processing," in IEEE Transactions on Image Processing, vol. 17, no. 3, pp. 272-282, March 2008, doi: 10.1109/TIP.2007.916157.

[6] A. Castillo, J. Ortegon, J. Vazquez and J. Rivera, "Virtual Laboratory for Digital Image Processing," in IEEE Latin America Transactions, vol. 12, no. 6, pp. 1176-1181, Sept. 2014, doi: 10.1109/TLA.2014.6894017.

[7] F. Yang, Y. M. Lu, L. Sbaiz and M. Vetterli, "Bits From Photons: Oversampled Image Acquisition Using Binary Poisson Statistics," in IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 1421-1436, April 2012, doi: 10.1109/TIP.2011.2179306.

[8] Nasser Kehtarnavaz; Shane Parris; Abhishek Sehgal, "Smartphone-Based Real-Time Digital Signal Processing," in Smartphone-Based Real-Time Digital Signal Processing , Morgan & Claypool, 2015.

[9] G. Lu, X. Zhang, W. Ouyang, D. Xu, L. Chen and Z. Gao, "Deep Non-Local Kalman Network for Video Compression Artifact Reduction," in IEEE Transactions on Image Processing, vol. 29, pp. 1725-1737, 2020, doi: 10.1109/TIP.2019.2943214.

[10] J. Detsing and M. Ketcham, "Detection and facial recognition for investigation," 2017 International Conference on Digital Arts, Media and Technology (ICDAMT), Chiang Mai, 2017, pp. 407-411, doi: 10.1109/ICDAMT.2017.7905002.

[11] Akira Hirose, "Quaternionic Fuzzy Neural Network for View-Invariant Color Face Image Recognition," in Complex-Valued Neural Networks: Advances and Applications , IEEE, 2013, pp.235-278, doi: 10.1002/9781118590072.ch10.

[12] N. V. Boulgouris; Konstantinos N. Plataniotis; Evangelia Micheli-Tzanakou, "PersonSpecific Characteristic Feature Selection for Face Recognition," in Biometrics: Theory, Methods, and Applications , IEEE, 2010, pp.113-141, doi: 10.1002/9780470522356.ch5.

[13] Ethem Alpaydin, "Pattern Recognition," in Machine Learning: The New AI , MITP, 2016, pp.55-84.

[14] Anthony T. S. Ho; Shujun Li, "On Forensic Use of Biometrics," in Handbook of Digital Forensics of Multimedia Data and Devices , IEEE, 2015, pp.270-304, doi: 10.1002/9781118705773.ch7.

[15] W. Yimyam, T. Pinthong, N. Chumuang and M. Ketcham, "Face Detection Criminals through CCTV Cameras," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 351-357, doi: 10.1109/SITIS.2018.00061.

[16] E. L. Anggraini, N. Suciati and W. Suadi, "Parallel computing of WaveCluster algorithm for face recognition application," 2013 International Conference on QiR, Yogyakarta, 2013, pp. 56-59, doi: 10.1109/QiR.2013.6632536.

[17] Y. Jihua, X. Guomiao and H. Dan, "Research and implementation of of multi-modal face recognition algorithm," 2013 25th Chinese Control and Decision Conference (CCDC), Guiyang, 2013, pp. 2086-2090, doi: 10.1109/CCDC.2013.6561280.

[18] Pradipta Maji; Sankar K. Pal, "Introduction to Pattern Recognition and Data Mining," in Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging , IEEE, 2012, pp.1-20, doi: 10.1002/9781118119723.ch1.

[19] Z. Wang and Z. Miao, "Feature-based super-resolution for face recognition," 2008 IEEE International Conference on Multimedia and Expo, Hannover, 2008, pp. 1569-1572, doi: 10.1109/ICME.2008.4607748.

[20] P. Kocjan and K. Saeed, "A Feature Based Algorithm for Face Image Description," 2011 International Conference on Biometrics and Kansei Engineering, Takamatsu, Kagawa, 2011, pp. 175-178, doi: 10.1109/ICBAKE.2011.17.

[21] X. Qi and L. Miao, "A Template Matching Method for Multi-Scale and Rotated Images Using Ring Projection Vector Conversion," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, 2018, pp. 45-49, doi: 10.1109/ICIVC.2018.8492726.

[22] J. Lu, V. E. Liong, X. Zhou and J. Zhou, "Learning Compact Binary Face Descriptor for Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 10, pp. 2041-2056, 1 Oct. 2015, doi: 10.1109/TPAMI.2015.2408359.

[23] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 1002-1014, 1 April 2018, doi: 10.1109/TPAMI.2017.2700390.

[24] A. Memiş and F. Karabiber, "Face recognition on mobile environment images using appearance based methods," 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, 2016, pp. 169-172, doi: 10.1109/SIU.2016.7495704.

[25] T. Mantoro, M. A. Ayu and Suhendi, "Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods," 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, 2018, pp. 1-5, doi: 10.1109/ICMCS.2018.8525935.

[26] C. Wu, W. Wei, J. Lin and W. Lee, "Speaking Effect Removal on Emotion Recognition From Facial Expressions Based on Eigenface Conversion," in IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 1732-1744, Dec. 2013, doi: 10.1109/TMM.2013.2272917.

[27] J. Ai, R. Tian, Q. Luo, J. Jin and B. Tang, "Multi-Scale Rotation-Invariant Haar-Like Feature Integrated CNN-Based Ship Detection Algorithm of Multiple-Target Environment in SAR Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 12, pp. 10070-10087, Dec. 2019, doi: 10.1109/TGRS.2019.2931308.

[28] C. Lee and D. Kim, "Visual Homing Navigation With Haar-Like Features in the Snapshot," in IEEE Access, vol. 6, pp. 33666-33681, 2018, doi: 10.1109/ACCESS.2018.2842679.

[29] T. Yingthawornsuk, N. Chumuang and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 418-423, doi: 10.1109/SITIS.2017.75.

[30] S. Suwannakhun, N. Chumuang and M. Ketcham, "Identification and Retrieval System by Using Face Detection," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 294-298, doi: 10.1109/ISCIT.2018.8587856.

[31] N. Chumuang, P. Chansuek, M. Ketcham, A. Silsanpisut, T. Ganokratanaa and P. Selarat, "Analysis of X-ray for locating the weapon in the vehicle by using scale-invariant features transform," 2017 Fourth Asian Conference on Defence Technology - Japan (ACDT), Tokyo, 2017, pp. 1-6, doi: 10.1109/ACDTJ.2017.8259599.

[32] P. Pramkeaw, M. Ketcham, W. Limpornchitwilai and N. Chumuang, "Analysis of Detecting and Interpreting Warning Signs for Distance of Cars using Analyzing the License Plate," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-8, doi: 10.1109/iSAI-NLP48611.2019.9045252.

Image Classification of Forage Plants in Fabaceae Family Using Scale Invariant Feature Transform Method

Thidarat Pinthong[1] ,Worawut Yimyam[2], Narumol Chumuang[3], Mahasak Ketcham[4],Patiyuth Pramkeaw[5], Nattavee Utakrit[6]

[1,2]*Dep. of Business Information Management,Phetchaburi Rajabhat University Phetchaburi, Thailand.*
[3]*Dep. of Digital Media Technology, Muban Chombueng Rajabhat University Ratchaburi, Thailand.*
[4,6]*Dep. of Management Information System, King Mongkut's University of Technology North Bangkok, Thailand.*
[5]*Dep.Media Technology Program, King Mongkut's University of Technology Thonburi,Thaland.*

worawut_yimyam@hotmail.com[1], thidarat.pin@mail.pbru.ac.th[2], lecho20@hotmail.com[3], mahasak.k@it.kmutnb.ac.th[4],patiyuth.pra@kmutt.ac.th[5], nattaveeu@kmutnb.ac.th [6]

Abstract—**This paper proposes a novel method for the image classification of forage plants in fabaceae family by using Scale Invariant Feature Transform (SIFT) method. The color image extension jpeg color mode RGB adjust the image to 1000x1000 pixels to get a single image of the template file. All of the sample images, four prototype images were standard scaled and rotated. The image was obtained through the image extraction process using SIFT implements and matching dataset of Forage Plants leaves with matching points to evaluate the accuracy of flea leaf identification, it was found that Senna siamea, Clitoria ternatea and Pithecellobium dulce leaves 100% accuracy but Sesbania grandiflora Desv was obtained with 0% accuracy. The total accuracy of all 4 plants 75%, indicated that the photosynthesis of SIFT leaves was suitable for Senna siamea, Clitoria ternatea and Pithecellobium dulce Because it is 100% accurate, but not with Sesbania grandiflora Desv leaves. The accuracy is 0% because the leaves are dark green. The leaves are not clear. And the leaves are slender, evenly spaced leaves, which makes it a very rare feature. While Senna siamea, Clitoria ternatea and Pithecellobium dulce leaves are clear. Leaf edge is unique. Include appropriate techniques for recognition and classification.**

Keywords-Image Classification; Image Classification; Forage Leaf; SIFT;

I. INTRODUCTION

In present, a pattern of animal raising has developed to become at an industry level. The main cost of animal raising covers animal feeds at 70 percent [1],[2]. In raising ruminants including beef cattle, dairy cattle, buffaloes, goats, sheep, deer and rabbits, forage plants are considered to be main feeds for such animals [3].

Each forage plant is adaptable, growing and giving effective production under different environment [2], [4], for example, areas, soil fertility and climate conditions, including temperature, light range, humidity, volume and distribution of rain to grow fruitful forage plants, it is necessary to consider types of Gramineae or Fabaceae to select suitable type for certain area and to consider its benefits such as forage plants for grazing, fresh forage grass for animal feeding or preserved or dried grass and silage. Each forage plant is suitable for different usage [5].

Fig. 1. Show the sample of forage plants.

Forage plants contain different nutrients and limitations on usage [6]. Some forage plants have limitations on usage including toxicity, nutrient inhibitors and growing [4] – [7]. In addition, some forage plants are similar so it is difficult to classify forage plants accurately. Therefore, classification of forage plants is very important.

Image classification is a statistical processing to classify all composing image point of a study site into subgroups [8]-[10], using statistical features to determine the differences among image point groups. The image points in the same group contain the statistical features in the same orientation [11].

SIFT is a method in computer vision to calculate keypoints in an image and to calculate features of key point [12]. SIFT uses a keypoint, not depending on scale, orientation, position, viewpoint, brightness, shadow to compare with other key points in an image, which is easy and accurate [13]. This method was

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

developed by Professor David Lowe of British Columbia in 1999. In general, a keypoint means a pixel in an image with Two -Dimensional change of Pixel Intensity around the key point to analyze an image using SIFT to localize a keypoint applies repeated octave [14]. The number of key points and positions change the number of octaves for suitable key point requires a test of such key point [15]. It randomly identifies the number of octaves of analysts, leading to delay or unsuitable key points [15]-[17]. Therefore, the researcher is interested in developing classification techniques for forages by localizing key points, using SIFT, which is very similar to forage classification performed by specialists.

In this paper, we divided aspects and sequenced the presentation as follows. The second topic addresses related related work. Third we describe about our process research methodology. The 4th discusses research results and conclusion. The last topic addresses the research discussion.

II. Related Works

A. Forage

Forage refers to any plants that animal eat and contributes benefit to them without toxicity [18]. It consists of two families including Gramineae such as grass, millet, corn, etc. and Fabaceae such as Centosema pubescens [19],

A leave of Gramineae Family consists of a leaf sheath which both sides overlap to tightly support the trunk, and leaf blade or lamina [20]. Leaves of Gramineae Family have longitudinal parallel vein. A lamina is small and narrow, yet it very long, compared to the leave width. It looks similar to a spear. A leaf base is wider than other parts. A leaf surface is smooth or rough [19]. A connecting part between the lamina and the internal leaf sheath has thin tissue called ligule. It has different patterns. On each side of Gramineae leave has one auricle [21]. The lamina functions as a common leave such as photosynthesis and dehydration. In addition, it protects and helps sprouting. The ligule of each Gramineae has individual feature which can be used for Gramineae classification. The ligule is a thin white or brown tissue. Some ligules have hair around the edge or do not have hair with smooth edge [19],[20].

A leaf of Fabaceae Family is an alternate compound leaf [22]. Each leaf consists of a petiole. On the petiole has several leaflets. A size and a shape are different, depending on types of plants. A petiole of leaflet is called petiolule and a petiole between petiolule of each leaflet is called rachis [23]. There is a stipule at a leaf base. The special feature of leaflet is a leaflet base. Each petiole is large, which is called pulvinus. Sometimes, a leaf may change to climbing tendril. The leaf of Fabaceae Family usually has 3 leaflets.

B. Feature extraction

It is a process of localizing key points of an image to use as representatives of such image. To translate a color image, 3 types of data consisting of the followings are applied. 1) Spectral data is an average of tone change within different frequencies. 2) Textural data is spatial distribution of tone change within frequency range. 3) Contextual data is a result of image processing. The difference between tone frequency data and textural data is frequency data, focusing on identifying grey value of key points. The textural data focuses on distribution of grey tone across the image [24]. The method of identifying features are various as follows. (1) Edge Detection is used for identifying the edge of object and patterns of the image. The advantage is that it shows differences of image effectively. It is widely used with the sample with different shapes and contains fewer details. The disadvantage is that it is difficult to perform effective threshold and it also depends on details of an object [25],[26]. (2) Unwrap image is a method of copy an image by identifying a center of rotation point. Then, it is spread as a sheet to retain suitable features for circular object [27] through the use of Histogram Graph to use this method, it is not necessary to change an image into grey-colored image because each RGB value is collected and used as features. (3) Rotation Invariance is a change of orientations in positioning an image and collect features in each orientation to make a comparison. (4) Area identification within an object based on different types of images such as Ripple Feature, Vein Feature is to determine suitable threshold for edge detection or pixel. This method is suitable for an object with different physical features such as a leaf.

C. Scale Invariant Feature Transform (SIFT)

It is used to localizing key points of an image, not depending on its size or orientations of objects in the image. The outstanding features of the similar objects are similar. It is therefore often used to match objects from two images containing the same objects. It is widely studied to improve performance in recognition or classification of objects in an image. The key processes include feature detection and key point matching, which is effective and fast recognizing [28]. SIFT is applied to analyze key points of an image. The key points of the image of the similar objects contain similar features. SIFT consists of 4 main processes [32] as follows.

- Scale-space extrema detection

Localizing keypoints of an image, not depending on size or orientation, by blurring the image through gaussian function of each octave. Each octave contains different levels of blurring, starting from normal blurring level to increasing level of σ (scale parameter). This results in more blurred image. It is continually repeated. The image size of the next octave is half of the image size of the first octave.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y)$$
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

L is an image blurred, G is gaussian filter with σ, I is an original image x,y is an coordinate on the image, σ is a variable of blurring level; the higher the

level is, the more blurred the image will become, * is an image convolution *I*, using gaussian blur *G*.

- Keypoint localization

Scale-space extrema detection results in keypoint localization [29]. It matches images of each octave and determine Difference of Gaussian (DoG). All images in each octave are performed.

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (2)$$

Keypoint localization [30] considers the maximum value and the minimum value of the image retrieved from DOG, considering 8 points at the same layer of the considering point, 9 points at the top layer and another 9 points at the low layer next to the considering point. As shown in Fig 3, if the considering point contains the maximum value and the minimum value, such point is localized as a keypoint. If the keypoint contains the difference of low contrast or if the keypoint is the edge, such keypoint is deleted.

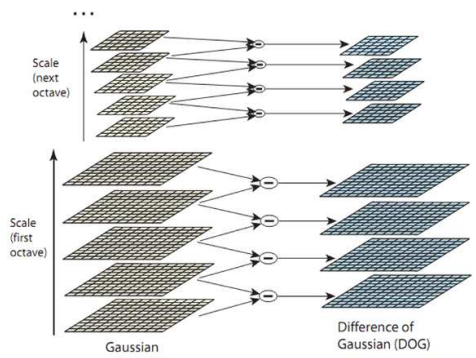


Fig. 2. Identification of difference of gaussian of each octave.

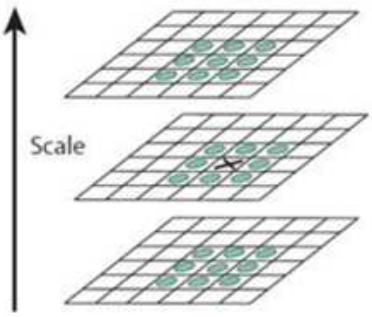


Fig. 3. Shows X which is a reference point, comparing the surrounding 26 points.

- Orientation assignment

Procedures of size and orientation assignment of Gradient around the keypoint in order to determine size *m(x,y)* and orientation *θ(x,y)* of keypoint applies the following equation [31].

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$
$$\theta(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \quad (3)$$

- Keypoint descriptor

A 16x16 window (window is the area to collect data of orientation with the size of 1.5*σ) around keypoint is created [32]. 16 sets of 4x4 windows are created. Each set calculates sizes and orientations of gradient to create a 8-bin histogram (x axis is divided into 8 ranges, each range is 45 degrees). The size of a histogram depends on the size*weight (distance from the keypoint). Therefore, when the calculation is finished, the outcome is 4x4x8 = 128. It is normalized and is used as a feature vector of each keypoint as shown in Fig 4.

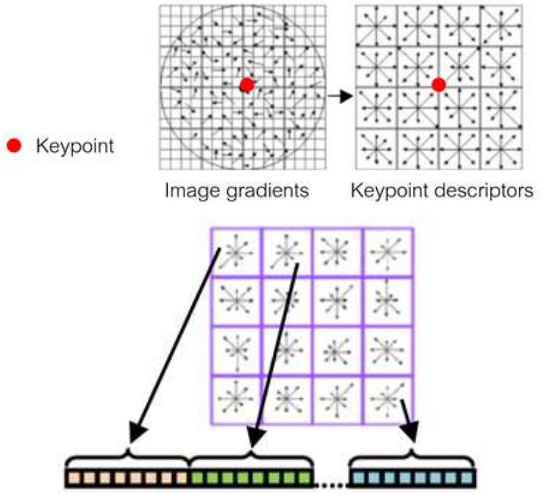


Fig.4. Show keypoints of an image.

S. Lavania and P. S. Matey [33] proposed feature detection using SIFT for leaves of local herbs. Performance measurement of feature detection was performed using images of 3 types of local herbal leaves including moringa leaves, mint leaves and centella asiatica leaves. According test and performance measurement by finding F-measure values, the test sample that had the best feature detection was the 3rd test sample (centella asiatica leaves) with the F-measure value of 60.0 %. The average of all test sample consisted of the F-measure value, the Precision value (P) and the Recall value (R) at 67.0%, 69.0% and 63.0% respectively.

A. A. Bharate and M. S. Shirdhonkar [34] proposed the rapid recognition to classify grape's leaves using SVM and vector features with high dimension. The researchers studied feature extraction from documents and research studies proposing methods and processes of rapid and accurate feature extraction. The followings are proposed by the researcher; Shape Description, Invariant Moment, Border Description, Color Description (Centroid Radii) and Texture Feature. After that, the extracted features were processed together, resulting in highly accurate outcome for recognition. When compared to outcomes of other studies, the outcome was highly accurate at 99% when performing Linear SVM and at

98.8% on Radiant Basis Function and Support Vector Machine.

Z. Long, M. Xiaoyu, L. Zhigang and L. Yong, [35] proposed feature extraction of tobacco leaves in the image processing and signal conference 2008. They employed several techniques to solve problems of feature extraction and analyze tobacco leaves including Color Features, Shape Features and Texture Features. GLCM (Grey-Level scale Co-occurrence Matrix was employed to extract texture features. The statistical data of GLCM including Texture Energy, Texture Entropy and Contrast value were included to the calculation. It was concluded that the data of feature extraction could be used to develop automatic machines that classify tobacco leaves.

T. Yingthawornsuk, N. Chumuang and M. Ketcham [36] proposed an automatic Thai baht coin calculation system using image processing technique. The results showed that the system could classify and calculate the total value of coins, which is very useful for visual impaired persons. The system employed an image processing technique for coin detection and SIFT algorithm was employed to classify groups and sound matching. The experiment was performed with 4 common Thai baht coins. They designed different cases. Each case contained different coins, ranging from the total number of 5 baht to 40 baht. The accuracy rate was 99%.

III. RESEARCH METHODOLOGY

The research methodology was divided into 2 steps including preprocessing, feature extraction and matching. The experiment was performed with images of 4 forages as show in Fig 5.

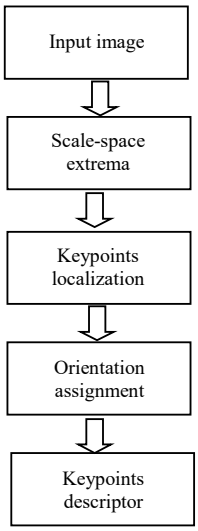


Fig. 5. The overview process work.

A. Preprocessing

Four types of forages were selected including Agasta, butterfly pea, manila tamarind and cassia. The leaves of the selected forages were captured by a digital camera. The leaves were placed on white paper. The images were in JPG in RGB mode. The size was 1000x1000 pixel with the resolution of 72 pixel/inch. The background was deleted to create an original image of each leave.

The testing image was created by placing leaves of 4 forages together. Among the 4 testing images, there was 1 original image, which was reduced its size and rotated.

B. Feature Extraction

The created images were extracted their features, using Scale invariant feature transform or SIFT

C. Matching

Images of forage leaves with the same keypoints were matched.

D. Evaluation

Accuracy of forage leave recognition was evaluated as eq (4). The percentages of accuracy for each type and of the total accuracy were calculated. The following eq (5) was applied.

$$\%Accuracy\ for\ each\ type = \frac{O}{C} x100 \qquad (4)$$

where O is number of certain forage leave giving accurate outcome, C is number of leave forage leave.

$$\%total\ accuracy = \frac{TO}{TL} x100 \qquad (5)$$

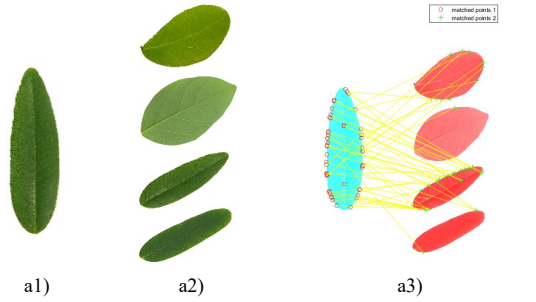
where TO is total number of certain forage leave giving accurate outcome, C is total number of leave forage leave.

IV. EXPERIMENTAL AND RESULTS

To conduct image classification of 4 Fabaceae forage leaves including Agasta, butterfly pea, manila tamarind and cassia using SIFT, the researcher conducted the experiment using matlab program. The results showed that cassia leaves (Figure 5), butterfly pea leaves (Figure 6) and manila tamarind (Figure 7) yielded the accuracy of 100% while Agasta leaves yielded the accuracy of 0 % (Figure 9). The total accuracy was at 75% as shown in Table I.

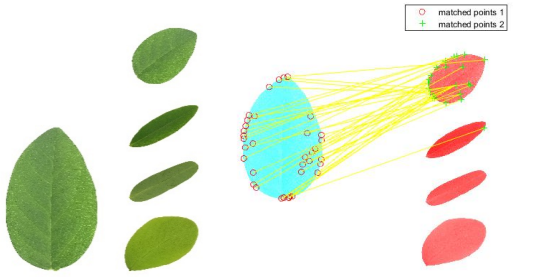
TABLE I. EVALUATION OF RECOGNITION ACCURACY OF FORAGE LEAVES THROUGH THE MATCHING OF SIFT

Types of forage leave	% of accuracy
Cassia	100
Butterfly pea	100
Manila tamarind	100
Agasta	0
% total accuracy	75



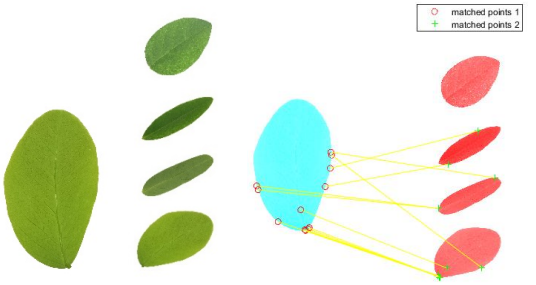
a1) a2) a3)

Fig. 6. Show the result as a1) Cassia leave model , a2) image dataset , a3) the results of classification and matching using SIFT.



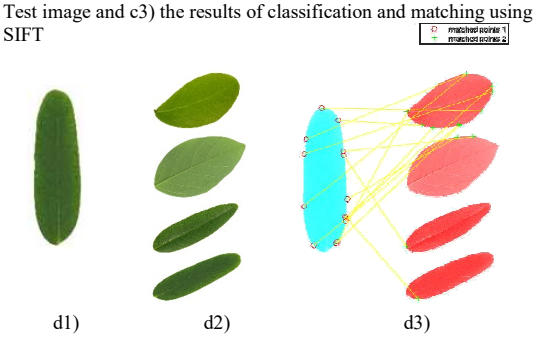
b1) b2) b3)

Fig. 7. Show the result as b1) Butterfly pea leave model, b2) Test image and b3) the results of classification and matching using SIFT.



c1) c2) c3)

Fig. 8. Show the result as c1) Manila tamarind leave model, c2) Test image and c3) the results of classification and matching using SIFT



d1) d2) d3)

Fig. 9. Show the result as d1) Agasta leave model, d2) Test image, d3) the results of keypoint localization using SIFT and matching results

The results showed that image classification of Fabaceae forage leaves could classify cassia leaves, butterfly pea leaves and manila tamarind leaves accurately. However, it could not classify Agasta leaves because the color of Agasta leaves is dark green. The leave is long with unclear and fewer patterns. The edge is not outstanding. This is totally different from cassia leaves, butterfly pea leaves and manila tamarind leaves, which are bright with clear patterns and outstanding edge.

V. CONCLUSION

It could be conclusion that image classification of 4 Fabaceae forage leaves including Agasta, butterfly pea, manila tamarind and cassia using SIFT is suitable for cassia, butterfly pea and manila tamarind as it yielded 100% of accuracy. However, it is not suitable for Agasta leaves.

Recognition and classification of Fabaceae forage leaves which are the important protein sources for ruminants such as beef cattle, dairy cattle, goats and sheep is essential to development of animal feed information system. If agriculturists could classify and recognize nutrients, how to use them for animal feeds, they can manage animal feeds effectively. Collecting forage leaves to create a data set requires complete and mature leaves without leave diseases, causing spots on the leaves. Furthermore, an investigation of age and types of Fabaceae forage as well as suitable techniques for recognition and classification shall be further studied.

REFERENCES

[1] C. Yongqiang, L. Shaofang, L. Hongmei, T. Pin and C. Yilin, "Application of Intelligent Technology in Animal Husbandry and Aquaculture Industry," 2019 14th International Conference on Computer Science & Education (ICCSE), Toronto, ON, Canada, 2019, pp. 335-339, doi: 10.1109/ICCSE.2019.8845527..

[2] T. Hirata, Y. Shiroma, T. Miyagi, M. Nakahodo, Y. Nagata and S. Tamaki, "Development of farmers support system on dairy and meat industry of goat utilizing Information and Communication Technology," The SICE Annual Conference 2013, Nagoya, Japan, 2013, pp. 2434-2440.

[3] C. Thoroughgood, "Animal Husbandry: Aquaculture Versus Agriculture," Proceedings OCEANS '83, San Francisco, CA, USA, 1983, pp. 993-997, doi: 10.1109/OCEANS.1983.1152016.

[4] T. Hirata, Y. Shiroma, T. Miyagi, M. Nakahodo, Y. Nagata and S. Tamaki, "Development of farmers support system on dairy and meat industry of goat utilizing Information and Communication Technology," The SICE Annual Conference 2013, Nagoya, Japan, 2013, pp. 2434-2440.

[5] P. Moaveni, "Effect of drought stress on dry forage yield, plant height, prussic acid and LAI in four varieties of sorghum bicolor," 2010 International Conference on Chemistry and Chemical Engineering, Kyoto, 2010, pp. 395-397, doi: 10.1109/ICCCENG.2010.5560406.

[6] A. Rajani, R. Darussalam, R. I. Pramana and A. Santosa, "Simulation of PV - Biogas Integration on Hybrid Power Plant using HOMER: Study Case of Superior Livestock Breeding Center and Forage of Animal Feed (BBPTU - HPT) Baturraden," 2018 International Conference on Sustainable Energy Engineering and Application (ICSEEA), Tangerang, Indonesia, 2018, pp. 69-74, doi: 10.1109/ICSEEA.2018.8627132.

[7] S. S. Chouhan, A. Kaul, U. P. Singh and S. Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant

Pathology," in IEEE Access, vol. 6, pp. 8852-8863, 2018, doi: 10.1109/ACCESS.2018.2800685.

[8] S. Thaiparnit, N. Khuadthong, N. Chumuang and M. Ketcham, "Tracking Vehicles System Based on License Plate Recognition," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 220-225. doi: 10.1109/ISCIT.2018.8588008

[9] T. Pinthong, W. Yimyam, N. Chumuang and M. Ketcham, "License Plate Tracking Based on Template Matching Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 299-303, doi: 10.1109/ISCIT.2018.8587992.

[10] B. Narin, S. Buntan, N. Chumuang and M. Ketcham, "Crack on Eggshell Detection System Based on Image Processing Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 1-6. doi: 10.1109/ISCIT.2018.8587980

[11] W. Yimyam, T. Pinthong, N. Chumuang and M. Ketcham, "Face Detection Criminals through CCTV Cameras," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 351-357, doi: 10.1109/SITIS.2018.00061.

[12] A. Farooq, M. Irfan, A. Saeed, T. Ansar, S. Arshad and N. Chumuang, "Design and Development of a Low-Cost Indigenous Solar Powered 4-DOF Robotic Manipulator on an Unmanned Ground Vehicle," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-5, doi: 10.1109/iSAI-NLP48611.2019.9045302.

[13] B. Zhong and Y. Li, "Image Feature Point Matching Based on Improved SIFT Algorithm," 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), Xiamen, China, 2019, pp. 489-493, doi: 10.1109/ICIVC47709.2019.8981329.

[14] M. Safdari, P. Moallem and M. Satari, "SIFT Detector Boosted by Adaptive Contrast Threshold to Improve Matching Robustness of Remote Sensing Panchromatic Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 2, pp. 675-684, Feb. 2019, doi: 10.1109/JSTARS.2019.2892360.

[15] Lowe D.G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60(2): 91–110.

[16] N. Chumuang, P. Chansuek, M. Ketcham, A. Silsanpisut, T. Ganokratanaa and P. Selarat, "Analysis of X-ray for locating the weapon in the vehicle by using scale-invariant features transform," 2017 Fourth Asian Conference on Defence Technology - Japan (ACDT), Tokyo, 2017, pp. 1-6, doi: 10.1109/ACDTJ.2017.8259599.

[17] W. Yimyam and M. Ketcham, "The automated parking fee calculation using license plate recognition system," 2017 International Conference on Digital Arts, Media and Technology (ICDAMT), Chiang Mai, 2017, pp. 325-329, doi: 10.1109/ICDAMT.2017.7904985.

[18] J. Francis, Anto Sahaya Dhas D and Anoop B K, "Identification of leaf diseases in pepper plants using soft computing techniques," 2016 Conference on Emerging Devices and Smart Systems (ICEDSS), Namakkal, 2016, pp. 168-173, doi: 10.1109/ICEDSS.2016.7587787.

[19] Stebbins, G. (1956). Taxonomy and the Evolution of Genera, with Special Reference to the Family Gramineae. Evolution, 10(3), 235-245. doi:10.2307/2406009

[20] Renvoize, S. (1981). The Sub-Family Arundinoideae and Its Position in Relation to a General Classification of the Gramineae. Kew Bulletin, 36(1), 85-102. doi:10.2307/4119008

[21] Ahmad M. Alqudah, Helmy M. Youssef, Andreas Graner, Thorsten Schnurbusch Natural variation and genetic make-up of leaf blade area in spring barley, Theoretical and Applied Genetics 131, no.44 (Jan 2018): 873–886. https://doi.org/10.1007/s00122-018-3053-2

[22] Cosio, E.G., Feger, M., Miller, C.J. et al. High-affinity binding of fungal β-glucan elicitors to cell membranes of species of the plant family Fabaceae. Planta 200, 92–99 (1996). https://doi.org/10.1007/BF00196654

[23] Adinarayana, D., Ramachandraiah, P. & Rao, K.N. Flavonoid profiles of certain species ofRhynchosia of the family Leguminosae (Fabaceae). Experientia 41, 251–252 (1985). https://doi.org/10.1007/BF02002621

[24] Cecilia Di Ruberto and Lorenzo Putzu. “A Fast Leaf Recognition Algorithm based on SVM Classifier and High Dimensional Feature Vector.” 2014 International Conference on Computer Vision Theory and Applications (VISAPP), 2557.

[25] Xinhong Zhang and Fan Zhang. “Images Features Extraction of Tobacco Leaves.” 2008 Congress on Image and Signal Processing, 2551: 773-776

[26] Ketcham M., Yimyam W., Chumuang N. (2016) Segmentation of Overlapping Isan Dhamma Character on Palm Leaf Manuscript’s with Neural Network. In: Meesad P., Boonkrong S., Unger H. (eds) Recent Advances in Information and Communication Technology 2016. Advances in Intelligent Systems and Computing, vol 463. Springer, Cham. https://doi.org/10.1007/978-3-319-40415-8_7

[27] W. Yimyam and M. Ketcham, “The automated parking fee calculation using license plate recognition system.” In 2017 International Conference on Digital Arts, Media and Technology (ICDAMT) (pp. 325-329). IEEE.

[28] M. Ketcham, W. Yimyam, and N. Chumuang, “ Segmentation of overlapping Isan Dhamma character on palm leaf manuscript’s with neural network”. In Recent Advances in Information and Communication Technology 2016 (pp. 55-65). Springer, Cham.

[29] Y. Li, J. Zhou, A. Cheng, X. Liu and Y. Y. Tang, "SIFT Keypoint Removal and Injection via Convex Relaxation," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1722-1735, Aug. 2016, doi: 10.1109/TIFS.2016.2553645.

[30] Y. Li, J. Zhou, A. Cheng, X. Liu and Y. Y. Tang, "SIFT Keypoint Removal and Injection via Convex Relaxation," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1722-1735, Aug. 2016, doi: 10.1109/TIFS.2016.2553645.

[31] Y. Xiang, F. Wang and H. You, "OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 6, pp. 3078-3090, June 2018, doi: 10.1109/TGRS.2018.2790483.

[32] Nitin Agarwal; Huan Liu, Modeling and Data Mining in Blogosphere , Morgan & Claypool, 2009, doi: 10.2200/S00213ED1V01Y200907DMK001.

[33] S. Lavania and P. S. Matey, "Leaf recognition using contour based edge detection and SIFT algorithm," 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2014, pp. 1-4, doi: 10.1109/ICCIC.2014.7238345.

[34] A. A. Bharate and M. S. Shirdhonkar, "Classification of Grape Leaves using KNN and SVM Classifiers," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 745-749, doi: 10.1109/ICCMC48092.2020.ICCMC-000139.

[35] Z. Long, M. Xiaoyu, L. Zhigang and L. Yong, "Application of Hyperspectral Imaging Technology in Classification of Tobacco Leaves and Impurities," 2019 2nd International Conference on Safety Produce Informatization (IICSPI), Chongqing, China, 2019, pp. 157-160, doi: 10.1109/IICSPI48186.2019.9095975.

[36] T. Yingthawornsuk, N. Chumuang and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 418-423. doi: 10.1109/SITIS.2017.75

A Development Heat Stroke Detection System Integrated with Infrared Camera

Worawut Yimyam
Dep. of Business Information Management,
Phetchaburi Rajabhat University,
Phetchaburi,Thailand
worawut_yimyam@hotmail.com

Mahasak Ketcham
Dep. of Management Information System,
King Mongkut's University of Technology
North Bangkok,
Bangkok,Thailand
mahasak.k@it.kmutnb.ac.th

Narumol Chumuang
Dep. of Digital Media Technology,
Muban Chombueng Rajabhat University
Ratchaburi,Thailand
lecho20@hotmail.com

Nattavee Utakrit
Dep. of Management Information System,
King Mongkut's University of Technology
North Bangkok,
Bangkok,Thailand
nattaveeu@kmutnb.ac.th

Montean Rattanasiriwongwut
Dep. of Management Information System,
King Mongkut's University of Technology
North Bangkok,
Bangkok,Thailand
montean.r@it.kmutnb.ac.th

Sansanee Hiranchan
Dep. of Management Information System,
King Mongkut's University of Technology
North Bangkok,
Bangkok,Thailand
s6207021910051@kmutnb.ac.th

***Abstract*— Currently, the problem of global warming is increasing heat illness called Heat Stroke disease, body core temperature has risen over 41 Celsius, affected to central nervous system failure and death cased. This research proposed A Development Heat Stroke Detection System Integrated with Infrared Camera to detect people with body temperatures above 39 degrees Celsius or people who are at risk of heat stroke. The experiment with sample group's photo result show that system able to diagnosis Human Body detection accuracy rate 90 % Temperature measurement 60 % Heat Stroke detected notification 100 % and responsive time 60 % total accuracy validation method summarized 77.5 percent. This study is collect a satisfaction measurement about Visualize, Usage and Contribution from program expertise with good satisfaction result. The work can be invention to screen people outdoor activities who have heat stroke's risk for first aid assistant.**

Keywords*— Heat stroke, Infrared camera, Image processing, Body temperature

I. INTRODUCTION

Climate change is a major global problem and hot weather problem related to Heat illness symptom able to cause death which dividing as dangerous level such as Heat Exhaustion or Heat Stroke [3],[8]. Previous work, Detecting humans in image is a challenging task owing to their variable appearance and the wide rage of poses that can adopt especially human body heat detection need to improve performance for diagnosis crowed in public with short-time responsive.[6] This research aim to develop Heat Stroke Detection System Integrated with Infrared Camera advocated effective detection people at risk of heat stroke and notification alert to officer according to a remote monitoring system called smart camera with a wireless transmission [1] and a facial verification by analyzing thermal data from different organs of the human face [2] concept model.

II. LITERATURE REVIEW

A. Heat stroke

Heat Stroke is emergency symptom [3] [8] caused by the body cannot release body temperature. Most of patients will have 3 pattern 1) high fever 40 ºC 2) no sweating and 3) the central nervous system malfunctioning such as fainting, restlessness, aggressive behavior, hallucinations, depression, staggering, unconsciousness, etc.

Normally, the body has a mechanism to adjust the body temperature to around 37 degrees Celsius or 98.6 degrees Fahrenheit. Over time, if the body has accumulated a lot of heat, such as metabolism, muscle movement the body will eliminate heat from the body by radiation, which is to distribute heat from the body to the surrounding air cooler body but if the outside air is hotter than 35 degrees celsius or hotter than body temperature.

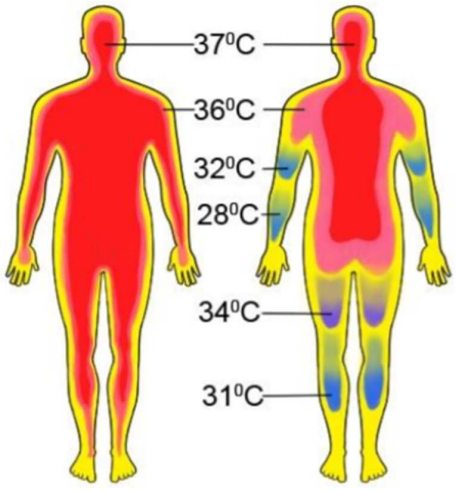


Figure 1 Body temperature and body heat transfer

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

The cause of heat illness [8] is caused by exposure to hot weather such as the occurrence of heat waves greater than 39.2 degrees Celsius (102.5 degrees Fahrenheit) for more than 3 days consecutively. Heat-borne diseases include young children, people over 65 years old, obese people, people with chronic diseases such as heart disease, diabetes, toxic goiter, Parkinson's disease, People who take certain medications that block the mechanism for the elimination of heat from the body (such as antipsychotic, antihistamines, anti-colic drugs), people who drink alcohol or Cocaine users or amphetamine (amphetamines) or from exercise or laboring heavily in hot and humid weather or in a closed, hot room causing the body to produce more heat than can be eliminated. Is a common cause among young people with strong bodies such as athletes, military workers etc.

B. Human Detection

Detection of people using image processing technology (Image Processing) [4], [5], [6]. The working principle is separation of personal characteristics the object of interest background with the object not interested which has the algorithm need to be detected. The Histogram of Oriented for Human Detection has the following sequence of operations.

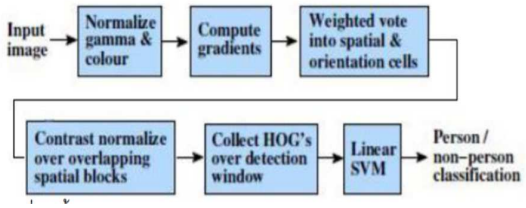


Figure 2 Histogram of Oriented for Human Detection steps

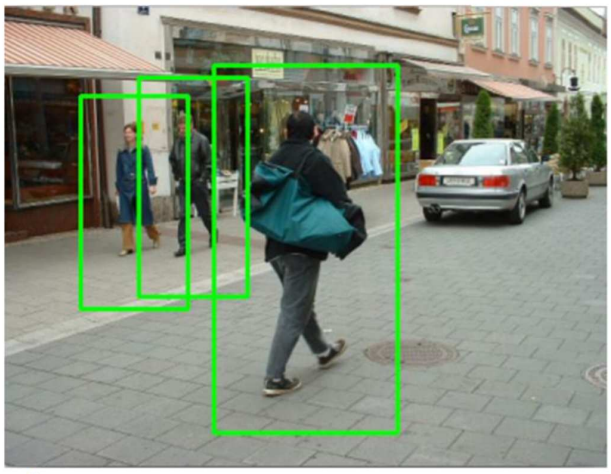


Figure 3 Histogram of Oriented for Human Detection sample

C. Infrared Camera

Infrared Camera or Thermography able to generate images from infrared radiation comparative with an ordinary camera, different from an ordinary camera. Imaging of visible light with wavelengths in the range of 450–750 nm are used, but thermal imaging camera work with electromagnetic waves at a wavelength of approximately 14000 nm, which are not visible to the naked eye. Also known as the heat wave in theory, a thermal imaging camera has a temperature range from temperatures above zero degrees completely. In practice, however, the operating range of the measuring instrument depends on the technology used to make the device, as well as the wavelength range chosen by the manufacturer. Some thermal imaging cameras can now measure temperatures as low as -20 °. C [2], [7]

Figure 4 Infrared camera vision

III. METHODOLOGY

A. Study and analysis solutions

Study and analysis solution is a step of understanding how to identified heat stroke people with temperature measurement equipment such as a thermometer that only use time by time and limited to check multiple people at once. At present, the Army Medical Department has recommended the use of a thermometer in the armpit, which is easy to measure but the obstacles, while measuring body temperature using mercury, cannot always be measured requires a period of 2-3 minutes with hundreds of soldiers. While the heat build-up in the body is like a time bomb that can cause heat stroke all the time. The purpose of research have an idea of the application of image processing techniques and use a camera that can capture images in infrared spectrum to detect people at risk of heat stroke so that they can early find people at risk of heat stroke quickly and efficiently. This research will allow people who do activities in the sun or places with extremely high temperatures, such as soldiers, troops, or other operators. Safe from pain and death from heat stroke (heat stroke).

B. System design

The step of system design have dividing in 2 part, type of camera and system detail below :

Type of camera

- Infrared Camera
- Visible Camera

A Infrared Camera deploy vision input to system analysis for thermography image and Visible camera captured ordinary image. The images from both cameras are sent to the server where the image data is stored and transferred to the Python Script, which has algorithms for image quality, slice detection and color value to alert the person to the information in the Database and data stored where users can access the information stored by Web Browser (Bootstrap Platform) via the Internet on various devices such as Smart Phone, Laptop or Tablet etc.

The Flowchart of Python Script in a system to detect people at risk of heat stroke using infrared cameras and image processing. Starting with quality improvement It is then a process to detect people

(Human Detection) and will take the color from the images that detect them. Compare the values in the temperature mapping table to get the temperature of that person if they match the table. Will show the person's temperature value to the display screen and if the temperature exceeds the set (Risk of heat stroke) will alert the user that a person is at risk of heat stroke. But if the corresponding color value in the color-temperature comparison table cannot be found, the "Error Mapping Temperature" value will be displayed.

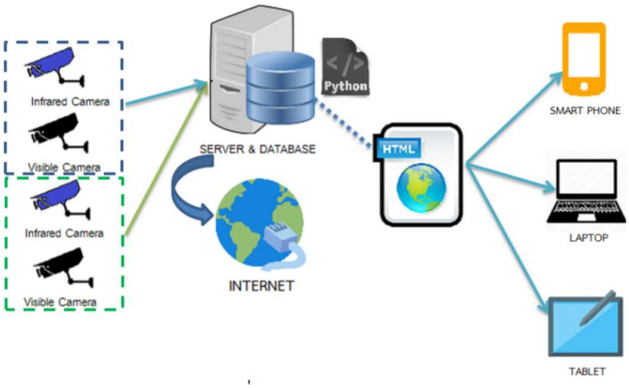


Figure 5 Overall system design

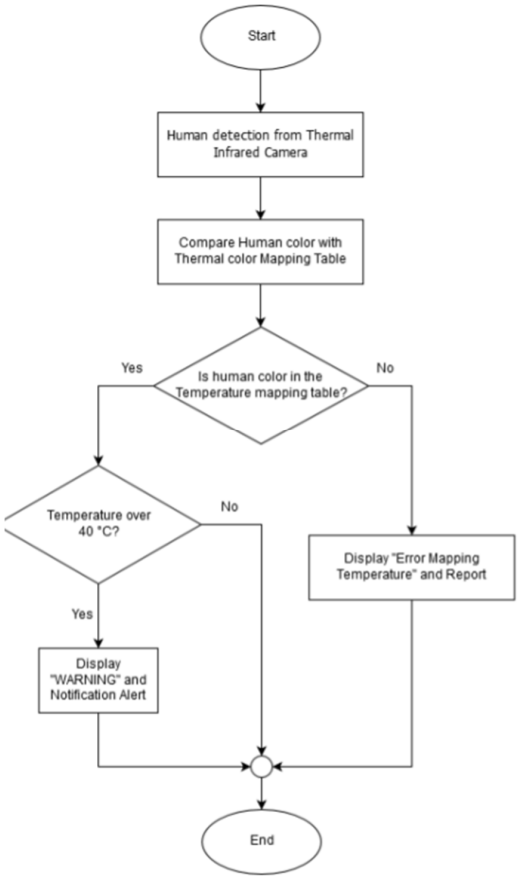


Figure 6 Python Script decision model for heat stroke detection

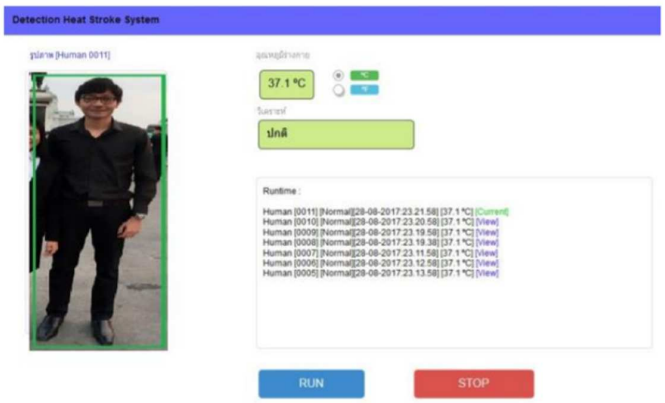


Figure 7 Monitor Display

C. *System development*

The system to detect people at risk of heat stroke using infrared cameras and image processing, the image analysis algorithm is located in Python Script, which starts from receiving the image from the camera which receives the image. Comes from both cameras as the received image is less sharp and noise Therefore, it is necessary to adjust the quality of the received image to be clear, improving the quality of that image according to Figure 8.

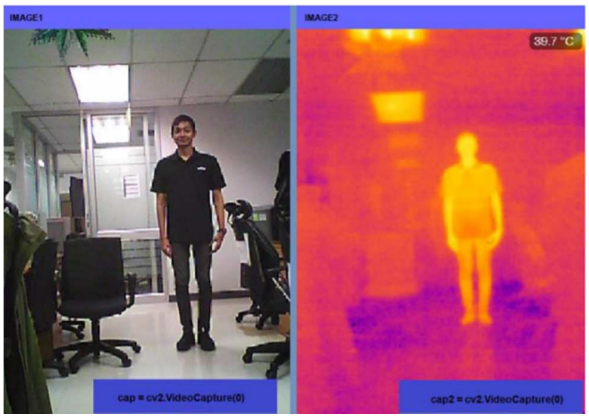


Figure 8 Infrared Camera output

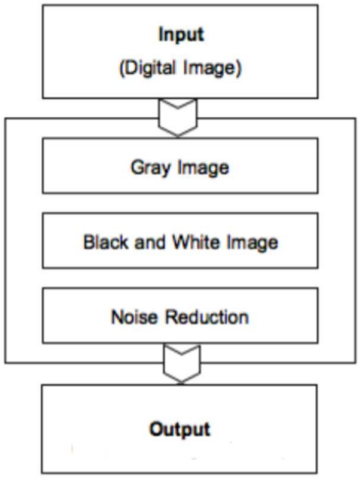


Figure 9 Quality image processing

After that, the image is submitted to the human detection process by using the Histogram of Oriented for Human Detection, the results will be shown in Figure 10.

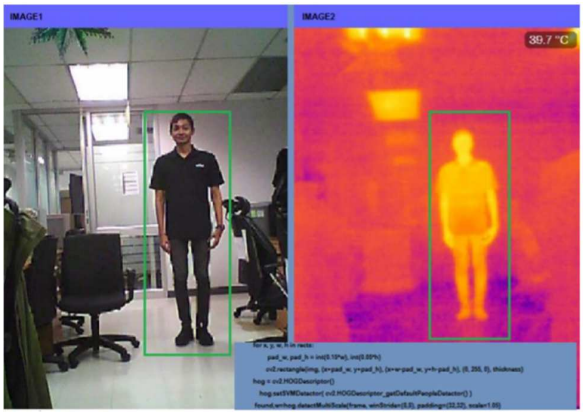


Figure 10 A results after entering the human detection process

According to heat stroke symptom a core temperatures (chest, abdomen, or rectum) [3] must be over than 41 ºC. The experiment selected a image's center or abdominal area for comparison temperature value. If it is less than the set temperature of risk, the program evaluated "normal temperature condition". In case of temperature between 39 ºC to 41 ºC, the program will evaluated "high temperature condition" then show output as heat stroke risk and store data in database.

Table 1 Heat_Stroke Data table

No.	Filed name	Data type& Data size	Description	Key
1	Id	Int(11)	Identity number	PK
2	Im_visible	varchar(300)	Digital image	
3	Im_Infrared	varchar(300)	Infrared image	
4	Date_time	Datetime	Date&time	
5	Temperature	Float(11)	Temperature	
6	Status	Int(11)	1= Normal 2= Heat stroke risk	

The display is developed in the form of Web Application developed using Bootstrap, JSP, HTML, Javascript, Ajax and Jquery, which can be accessed via the Internet on various devices such as Smart Phone, Laptop or Tablet, etc., in which the screen has The results are in 4 parts as follows: In Section (1) it will be a picture of a person whose data has been analyzed by the program. This section retrieves information from the "im_visible" field. Later, part (2) shows the temperature value, which is extracted from the "temperature" field. It can be viewed in degrees Celsius or degrees Fahrenheit. Section (3) is the analytical part in which the analysis results It can be obtained from the "status" field where the value 1 is equal to "normal" and the value 2 is equal to. Finally, in Section (4) Runtime, this section displays all information and updates automatically (Real-Time) when there is information on a person at risk of heat stroke (Heat Stroke). Will be red and a notification sound will sound to alert you. And users can check the data on detection of persons at risk of heat stroke (heat stroke) retrospectively, the final result will be as in Figure 12.

Figure 11 Result Display

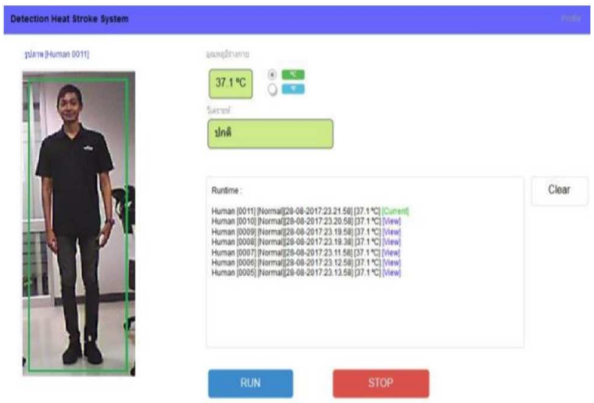


Figure 12 Final Result Display

IV. RESULT

A. Performance validation

A Development Heat Stroke Detection System Integrated with Infrared Camera is a human heat detection system for prediction heat stroke sickness. The experiment shows result from sample 100 images from the Infrared camera and system detected abnormal condition which is temperature between 39 ºC to 41 ºC and display result accuracy 77.5% as detail shown in table 2.

Table 2 Validation accuracy

Validation method	Accuracy (%)
Human body detection	90
Temperature measurement	60
Notification when heat stroke is detected	100
Respond time	60
summarized	77.5

B. Satisfaction measurement

The Development Heat Stroke Detection System Integrated with Infrared Camera satisfaction assessment divided into 5 levels which is based on Likert Scale method. The questionnaires were validated for the conformity of the questions to measurement objectives (IOC) with the program development expertise.

Table 3 The Development Heat Stroke Detection System Integrated with Infrared Camera satisfaction assessment

Measurement list	Satisfaction level		
	X	S.D.	Result
Visualize	4.27	0.43	Good
Usage	4.25	0.42	Good
contribution	4.40	0.56	Good
Summary	4.32	0.42	Good

V. CONCLUSION

The system for detecting people at risk of heat stroke using infrared cameras and image processing. They can be used to quickly and effectively detect and alert people who are at risk of heat stroke. The results from 100 photos and images taken from thermal imaging cameras to be tested for use with the system were found The system for detecting people at risk of heat stroke by using infrared cameras and image processing was accurate to 77.5% and the system could be enabled for a variety of applications due to the development of a display in the form of Web Application. Make the evaluation results of satisfaction in using the system at a good level Yes, with the mean overall satisfaction at 4.45 and standard deviation (SD) of 0.42

VI. DISCUSSION

The Development Heat Stroke Detection System Integrated with Infrared Camera can be used to quickly and effectively detect and alert people who are at risk of heat stroke. In order to bring the patient to first aid promptly Before it can be dangerous Which makes it safe for doing outdoor activities or at high temperatures Additionally, the system detects people at risk of heat stroke using infrared cameras and image processing that detects a person's body temperature. Then alert Therefore can be applied to other diseases Regarding body temperature, people at risk of heat stroke are detected using infrared cameras and image processing. It helps detect all people caught by the camera to be at risk of heat stroke, however, there are certain groups of people at greater risk than others. For example, young children, older people, obese people, people with chronic diseases, or people who have had heat stroke before have the right to heat stroke more than others. This research may study face recognition techniques in order to obtain priorities. Make these groups a top risk person to be extra careful with. To increase the safety of people who have to do activities in the heat even more.

REFERENCES

[1] T. Medjeldi, and S. Guillemette, “Wireless system for detecting human temperature,” Science and Information Conference (SAI),2013.

[2] Chule Yang, Danwei Wang, and Prarinya Siritanawan, “OrganBased Facial Verification Using Thermal Camera,” Multimedia(ISM), 2016.

[3] The Army Medical Department, “Surveillance, prevention and first aid for illness Due to heat”. Health Promotion and Preventive Medicine Division, Royal Thai Army Medical Department, 2555.

[4] Matt Davis, and Ferat Sahin, “HOG feature human detectionsystem,” Systems Man and Cybernetics (SMC), 2016.

[5] Luciano Spinello, and Kai O. Arras, “People Detection in RGBD Data,” Intelligent Robots and Systems (IROS), 2011.

[6] Navneet Dalal, and Bill Triggd, “Histograms of OrientedGradients for Human Detection,” CVPR, 2010.

[7] SzeYen Lee, and Mutharasu Devarajan, “Thermal analysis of multi-chip LED package with different position and ambient temperatures,” Photonics (ICP), 2011.

[8] Matirut Moongtin, Ram Rungsin, Worrutchanee Imjaijit et al, “A qualitative study for the prevention of heat stroke in military personnel”, The Military Medical Association of Thailand under the Royal Patronage, 2555

Enhancing and Evaluating an Impact of OCR and Ontology on Financial Document Checking Process

Worawut Yimyam
Dep. of Business Information Management, Phetchaburi Rajabhat University,
Phetchaburi, Thailand
worawut_yimyam@hotmail.com

Mahasak Ketcham
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok,Thailand
mahasak.k@it.kmutnb.ac.th

Tanapon Jensuttiwetchakult
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok,Thailand
tanapon_j@it.kmutnb.ac.th

Sansanee Hiranchan
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
s6207021910051@kmutnb.ac.th

Patiyuth Pramkeaw
Media Technology Programe,King Mongkut's University of Technology Thonburi,
Bangkok,Thailand
patiyuth.pra@kmutt.ac.th

Narumol Chumuang
Dep. of Digital Media Technology, Muban Chombueng Rajabhat University,
Ratchaburi, Thailand
lecho20@hotmail.com

***Abstract*— This research objective is to increase the efficiency of financial document auditing process for reimbursement by using Optical Character Recognition (OCR) and Ontology. On the researched system, user can use the system for checking completeness of document completeness, in accordance with the disbursement accounting standards, with digital photo after paid immediately. In the past it took more than a week to submit the evidence and make an account disbursement document. This research can reduce the time since the occurrence of the program. (Transaction) up to the creation of documents with more than 50% disbursement.**

Keywords*— *Reimbursement, OCR on receipt, Financial document checking process, Ontology

I. INTRODUCTION

Nowadays, information technology in the accounting process and financial process are increase instead of using paper or books that helps manage accounting, finance, and processing for reporting.

The cash payment in each organization need the document for accounting records submission in each transaction. But checking financial documents process before accounting records submission rely on accounting staff, which may cause delays, errors, and lose the document. According to the Revenue Department, accounting records should have accurate and complete supporting documents for tax purposes [1]. Revenue Department will call for the evidence of payment if incomplete accounting document were sent, that may impact to the financial system in company.

If that offense is to avoid or fraudulent from 10 million baht per year tax up, or tax refunds by fraud of 2 million baht per year or more These are the basic offenses under the money laundering law. Under the Revenue Code Section 37 term, which specifies offenses that fall within the scope of money laundering law in 3 sections, namely Section 37, Section 37 Bis and Section 90/4 [2]

This research objective is to increase the efficiency of financial document auditing process for reimbursement by apply optical character recognition (OCR) and ontology. users can take a photo of the receipt via the mobile phone camera after making payment immediately. The system will check the completeness and accuracy of the information on documents according to the standard accounting principles for disbursement Which in the past had to take More than a week to submit evidence and make an account disbursement document This research can reduce the time since the occurrence of the program. (Transaction) up to the creation of documents with more than 50% disbursement

Fig. 1. Disbursement process with text recognition (OCR) on financial documents

II. LITERATURE REVIEW

A. *Criteria for the preparation of accounting documents that can be taxable*

In the case of actual expenses paid but the recipient does not have evidence Accepting sufficient funds to account the business [1], can prepare supporting documents for accounting. with details as follows

- Document the receipt of the payee.
- Invoice or receipt showing the name, eg electricity or water bill, but the entity has evidence to prove that is company payment.
- Prepare payment vouchers for use in cases where there is no evidence as above.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

B. Principles of petty cash disbursement

Petty Cash Book [1] used for the trading business. product or services that tend to have issues with cash-related oversight especially cash, bank notes and coins. Each day usually has an expense of trade items with a small amount of money that usually cash payment than cash checks. The low-cost expense is hardly to receive the document in every transaction. an organization should record the transaction with a small amount of money in a small passbook to use to record the small amount of money received as a check.

C. Optical Character Recognition;OCR

OCR is the process of converting printed media, such as magazines, contracts, into text data. OCR (Optical Character Recognition) technology was first applied to document management in 1929 by Mr Tausheck [3]. In 2012 A. Singh, K. Bacchuwar [4] analyzed and compiled the application of OCR technology including OCR development techniques. OCR may be used in the financial even more. In 2015 A. D. Gross, D. G. Neely and J. Sidgman [5] describe the trend in using OCR technology that will play a significant role in the transition to this digital age. By discussing the challenges of applying OCR on information systems efficiency and data security. The format of documents without explicit forms by applying OCR technology. There are many open source applications for application development. In 2016 P. Chirag Patel [6] Tesseract is an open source development developed by HP company. Comparison of efficiency with Transym shows that Tesseract gives average accuracy of 71% black and white images and 61% color images, but Transym has average results at 47% The time it takes to process Tesseract takes only 1 second.

1) Image Pre-Processing

Pre-processing, is one of the three image processing processes [7], to improve image quality before entering the character recognition process.

In 2011, KS Xujun Peng Huaigu Cao [8] evaluate image quality from mobile phones before processing into the OCR process, because the image quality is count on User Experience, and impact the OCR program. Optimization OCR can be used for comparison and errors prone technique that have three methods including Full-Reference, No Reference, and Reduced Reference. In 2016 S. Sharma [9] research on image processing techniques with picture from low quality cameras for optimizing OCR. The researcher tested the image improvement using De Noising, De Blurring, Contrast Enhancement and Regularized Filter techniques. The best results are compared between Lucy Richardson and Weiner in the Imadjust technique in the step contrast adjust in characters images.

2) Recognition

This process is the core of OCR that define the character in image which have recognition techniques include Template Matching, Statistical Approach, Structural Analysis, Neural Network.

a) Post-Processing

After the recognition process has been completed, the characters image sent resulted in a letter code. However, the OCR output have accuracy less than 100% in every language. With the program accuracy improvement, the spell checking, and spell editor were added as post processing.M. Alwani [10] adapt a spell check from Google which can increase the accuracy to good level.

3) Ontology

The ontology in the computer industry is the definition of interest, which is always true fact. For example, the definition of a person analyzes that the components involved by people with two arms, two legs, two eyes, male and female, have white, yellow, or black skin, aged, family relationships, etc. These data can convert into the knowledge of computer. Y. Tijerino [11] applied the ontology theory to OCR and built the relationship structure on the receipt by creating ORM (Object Oriented Model) and domain for each word. In example, the system recognizes the word such as “Price” then returned variable type is Double (numbers).

III. METHODOLOGY

This research objective is to improve the financial document inspection process for reimbursement by using Optical Character Recognition (OCR), ontology and mathematical calculations.

The challenge of the research is low accuracy if using only text recognition [9]. The solution to increase the accuracy is adding processes PRE-PROCESSING, to increase the sharpness of the image, before the recognition process phase, and Post-Processing which matching keyword in the database.

A. Document acquisition as training set

The scope of the research use 2 type of the document including cash bill and carbon slip. The 100 good quality documents, which not fade and scratch, and 100 low quality documents were gathered. the OCR model has limitation on handwrite character recognition and recognize only number character, which are taxpayer and price.

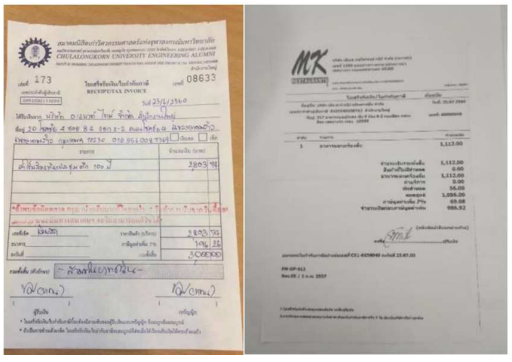


Fig. 2. The example of acquisted document

3.2 Financial document verification process with OCR and ONTOLOGY

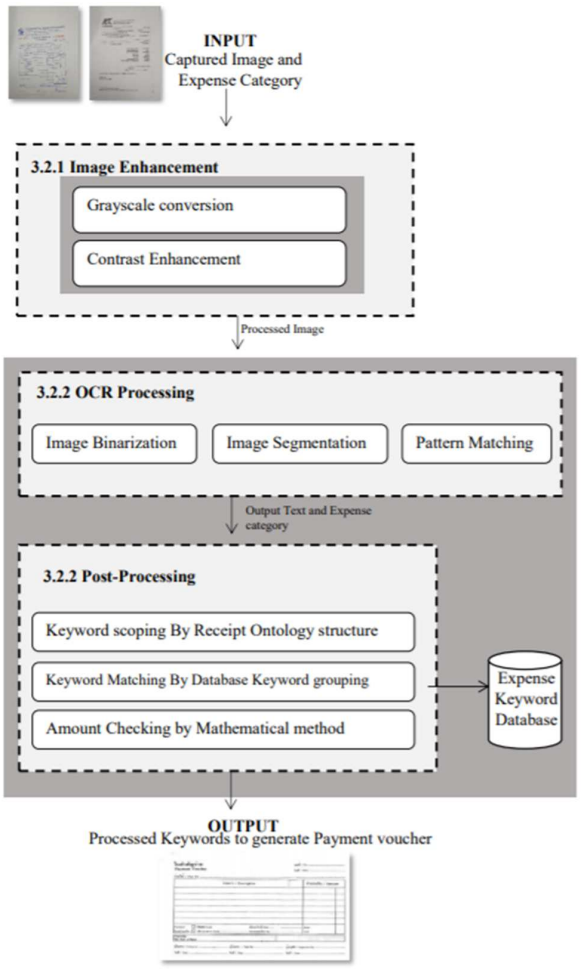


Fig. 3. Framework of financial document verification process with OCR

1) Image enchancment

a) Gray Scale

The image received in the first step is in the RGB color space. Each coordinate image is representing the value of R, the value of G and the value of B. In this step, RGB image were changed in a grayscale image for reducing dimension of image data.

b) Contrast Enhancement

Contrast is the color difference value, In term of histogram, the graph has a large color intensity. The original numerical image data grouped together in a narrow intensity range. The image will get a sharper if spread out a narrow intensity. However, the research uses Linear Contrast Stretch technique, which extends the range of the original data intensity until the full range 0-255.

2) Optical character recognition;OCR

a) Image Binarization

The procedure for replacing black in every pixel on the image with the concentration $Ii, j < T$ (Threshold), and replacing with white on every pixel that has more concentration than the specified value T.

b) Image Segmentation

The paragraph image was detected as tax line for word finding then analysis process to divide the image data, which contains a character, into small component that known as word chopping.

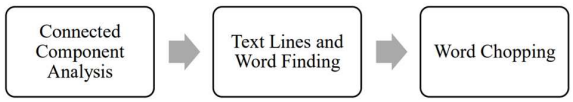


Fig. 4. Word and letter wrapping process

Fig. 5. Document pictures that have passed the image segmentation process

The character wrapping process uses the histogram or Verticle Projection inspection technique to see the spaces of each character. In which the density of the vertical black pixel group has the lowest value, density of vertical black pixels can be calculated in equation (10) and creating a bounding box for the next pairing process, as figure 6

$$p(x) = \sum_{y=0}^{y=ymax} P(x, y) \quad (1)$$

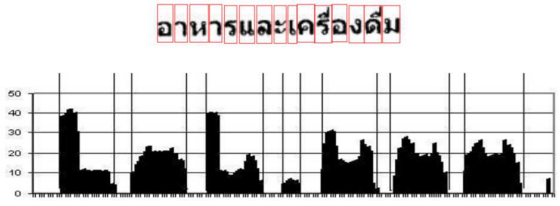


Fig. 6. Character clipping with histogram checking and bounding box

c) Pattern Matching

Character templates are created to compare the character images obtained from the image segmentation. To measure the similarity of the image and character template, key positioning principles were adapted, for distinguish between each character, to identify the word.

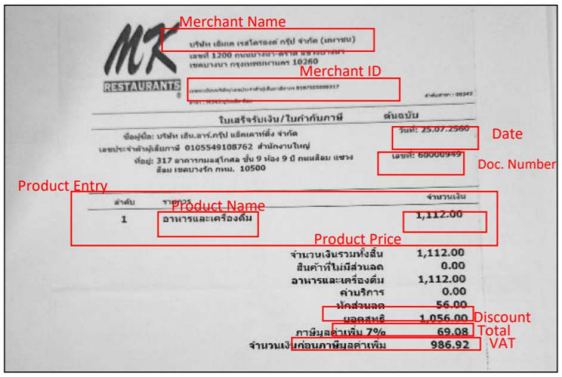


Fig. 7. Image of receipt with character recognition

3) Post-processing of character recognition

After the result of the character recognition process, there may have incorrect words. In this step, this will increase the accuracy of the vocabulary as follows;

a) Ontology

According to the structure of financial document, there are same element components on receipt or tax invoice which revenue department have defined. Such as name of establishment, address of establishment, identification number of companies, date of document, document ID. These details about payment are used to create an ontology structure as figure 8.

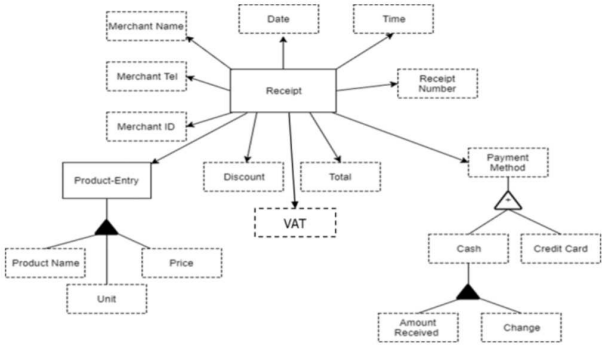


Fig. 8. Ontology-based automatic receipt accounting system

In which the rules and formats of the results can created according to the example in Figure 6.

```
Price
Internal representation: Double
External representation: \$[1-9]\d{0,2}
Context keyword: รวมจำนวนเงิน, รวมจำนวนเงินทั้งสิ้น, Total
Unit: Baht, บาท
Output method: toThaiBahtFormat
...
end
```

Fig. 9. example of the result of price by using the ontology

b) Comparative the words with the vocabulary database.

Assuming users choose the vocabulary category as expenses related to food. The data can be compared to the vocabulary by Levenshtein Distance method [12] by considering the approximation and the length of the vocabulary by using the formula as follows:

$$cost = (s[i] == t[j])? \ 0 : 1 \quad (1)$$

$$d[i, j] = Min\ (d[i-1, j]+1\ ,d[i, j-1]+1,\ d[i-1, j-1]+cost) \quad (2)$$

Define s as first word and t as second word. after acquiring similar terms, the histogram were measured the differential by using Character length. If the difference is less than 0.5 then the vocabulary will be changed by using the equation

$$D = \frac{Lev\ Distance\ (a,b)}{Min{-}L(a,b)} \quad (3)$$

Where D is the difference result, Lev Distance (a, b) is the difference of Levenshtein Distance of the words a and b, Min-L (a, b) is the shortest length of the words a and b.

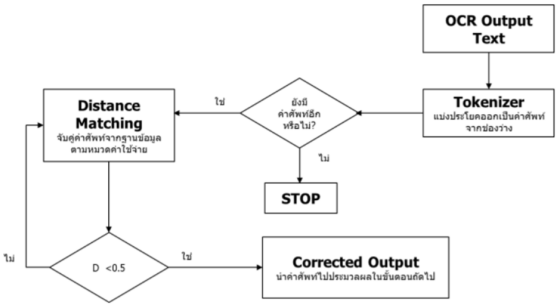


Fig. 10. Word and vocaulary database matching work diagram

c) Mathematical and financial calculations principles

Since the recording of payment documents is not necessary to collect details of all paid items thar using only the total amount paid. The researcher put the vocabulary in category topic relating to amount of money. Mathematical calculations get the correct amount from figure 5. There are categories relating to the amount of money is product price, discount, total, value added tax (VAT), amount paid (amount received) and change. According to the equations (4), (5), (6) or (7), which designed the calculation logic when detected information in the document to select the equation as shown in figure 11 to calculate the net amount for automatic accounting.

Total= Product Price (4)

Total=(Product Price-Discount) (5)

Total=(Product Price-Discount)+vat (6)

Total=Amount Received-Change (7)

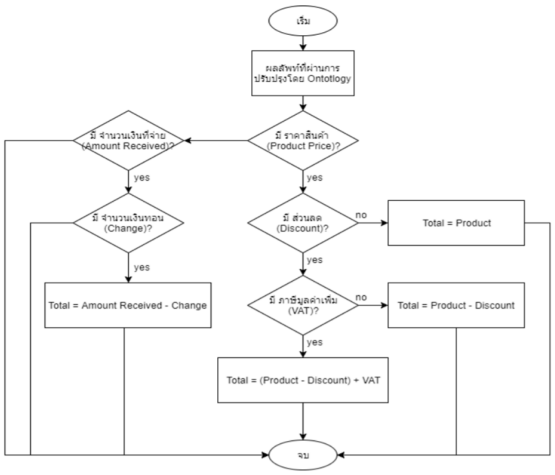


Fig. 11. Flowchart of equation selection with multi-cariable

4) System performance evaluation

The system performance is measured by obtain results from the character recognition processor on the picture that has been quality measured using the following equation.

$$Accuracy\ of\ vocabulary\ improvement = \frac{Correct\ word}{Total\ word} \quad (8)$$

The researcher will evaluate system performance according to 2 types of documents, which are cash bills and carbon receipts. The position of character recognition improvement 11 area on receipt including date, document number, distributor name, dealer tax ID, product name, product price, discount, VAT, and total amount. To be able to know Efficiency, suitability of the techniques used

IV. RESULT

A. The result of pattern matching based on OCR

The result of character recognition still has an error in product name and the total expense that can't be used to create disbursement document. There need techniques to improve efficiency of OCR which takes into consideration the similarity and length of the vocabulary correct as in Table 1.

TABLE I. VOCABULARY EXAMPLES FROM CHARACTER RECOGNITIONUSING LEVENSHTEIN DISTANCE TECHNIQUE.

#	Wording type	Ontology result	Variance	Th… c…
1	-	23-9-2559	0	23-9-2559
2	-	0105555149092	0	0105555149092
3	อาหารและเครื่องดืม	อาหารและเครองดืม	0.5	อาหารและเครื่องดืม
4	-	9330.0	0	9330.0
5	-	9	0	9

B. OCR and pattern matching result

The evaluation results were obtained from testing with 200 samples with good quality and bad quality document image. The test results show that the system can check cash bills with up to 90% accuracy and carbon slip up to 95% accuracy. Low quality documents will produce less accurate results because the quality of the documents affects the accuracy which is effective as in Table 2. The process to increase the reimbursement process OCR and ONTOLOGY. The financial documents were compared with the traditional process that workers must submit the actual document for disbursement. The result of OCR and Ontology can reduce the time by up to 50%, as shown in Table 5.

TABLE II. VOCABULARY IMPROVEMENT RESULT FOR BUILDING PAYMENT CERTIFICATE FOR DISBURSEMENT

Document type	Accuracy of vocabulary improvement	
	Good quality	Low quality
Cash bill	90%	70%
Carbon Slip	95%	85%

TABLE III. COMPARATIVE OF BUILDING PAYMENT SLIP FOR DISBURSEMENT WITH EXIST PROCESS AND OCR SYSTEM

Exist process		New process	
Disbursement procedures	Time (minute)	Disbursement procedures	Time (minute)
Gather evidence	30	Get a document image	2
Create document	40	OCR and ONTOLOGY	1
Prove and cross check	1 day		
approve	1 day	approve	1 day

C. The results of the calculation of the equation to create a payment voucher

The image data was auctioned to select the equation for cost calculation according to the conditions specified in the document. in the sample data of creating vouchers, when processed according to the conceptual framework in figure 11. the calculated data in equation (4) will cause the total amount to become 9330.00 which is correct as in the original image, and able to be built disbursement document.

However, this part of system is unable to recognize the information or incomplete information input. Users can check and edit before confirming the creation of the voucher before sending to the users.

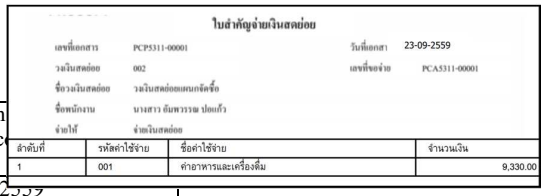


Fig. 12. Examples of payment vouchers created after acquiring the vocabulary and amount pairs

V. CONCLUSION

Research results By considering the research objectives To develop and increase the process of checking official documents Money with a mobile phone for reimbursement using character recognition (OCR). In this section, a study of possibility of research that Similar to both Thailand and overseas, dating back 5-6 years, resulting in data Reference in the application of OCR Including

improvement techniques Vocabulary quality from character recognition To allow the disbursement process The night is even more effective. Which the results of the efficiency evaluation accurate in creating disbursements, an average of 85% and 50% less process time. That means time and costs About making documents for disbursements actually reduced

The optimization of financial documents inspection process with a mobile phone to be used for reimbursement as well. Financial accounting is still dependent on humans or employees, which may cause delays and errors. From the evaluation results The average efficiency on documents is 85% that shown time and cost associated with making documents for disbursements can actually be reduced.

VI. DISCUSSION

Problems, that mostly encountered in this research, is transparent cash bill receipts, which makes it difficult to control the quality of the photos. Suggestions from research, development and increase efficiency process of checking financial documents with a mobile phone for use in Reimbursement with character recognition (OCR) may increase the scope of the group. More examples in the future to increase the variety of preparation reimbursement category and has user management in the company systematically.

REFERENCE

[1] Revenue Department, “Preparation of accounting documents,” Thailand, 2016.

[2] THANSETTAKIJ, “operation guidelines at 2017”, 2017

[3] G. Tauschek, ―Reading machine‖ U.S. Patent 2026329, Dec. 1935

[4] A. Singh, K. Bacchuwar, A. Bhasin, “A Survey of OCR Applications,” 2012

[5] A. D. Gross, D. G. Neely, J. Sidgman, “When Paper Meets the Paperless World,” The CPA Journal; New York, pp. 64-67, 2015.

[6] A. P. P. a. D. P. Chirag Patel, “Optical character recognition by open source OCR tool Tesseract: A case study,” International Journal of Computer Applications; New York, 2016.

[7] Processing, Digital Image Processing and Digital Signal, “Digital Image Processing and Digital Signal Processing,” Thailand, 2007.

[8] K. S. Xujun Peng Huaigu Cao, “Automated image quality assessment for camera-captured OCR,” IEEE Xplore document, 2011.

[9] P. S. a. S. Sharma, “Image processing based degraded camera captured document enhancement for improved OCR accuracy,” IEEE Xplore document, 2016.

[10] Y. B. a. M. Alwani, “OCR post-processing error correction algorithm using Google online spelling suggestion,” arxiv, 2012.

[11] Z. S. a. Y. Tijerino, “Ontology-Based automatic receipt accounting system,” Digital Library, 2012.

[12] I. Setiadi ,”Damerau-Levenshtein Algorithm and Bayes Theorem for Spell Checker Optimization”, 13511073 Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia,2013

COVID19 X-Ray Image Classification using Voting Ensemble CNNs Transfer Learning

Phuwadol Viroonluecha
Department of Information and Communication Technologies
Technical University of Cartagena
30202 Cartagena, Spain
phuwadol.viroonluecha@upct.es

Thanwarat Borisut
Department of Statistics
Ramkhamhaeng University
10240 Bangkok, Thailand
thanwarat.borisut1@gmail.com

Jose Santa
Department of Electronics and Computer Technology
Technical University of Cartagena
30202 Cartagena, Spain
jose.santa@upct.es

***Abstract*— COVID-19 is a novel pandemic and infected COVID-19 people are overgrowing, involving an outbreak that is changing lifestyles around the world. A global issue when trying to contain the propagation of the illness is how to efficiently detect infected people and isolate them. Medical image classification is one of the medical screening tools being used nowadays. Apart from manual inspection, several automatic methods can be applied to exploit artificial intelligence, such as Convolutional Neural Networks (CNNs). Transfer learning can also be applied to make predictive models more effective worldwide, due to the expected small amount of COVID-19 chest X-ray images available. In this paper, we propose the Voting Ensemble CNNs Transfer Learning to recognize the COVID-19 footprint and classify it in a chest X-ray image. The dataset used for training and evaluation was collected from several sources: COVID-19 image data collection and NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories. Our voting ensemble comprises CNNs architectures: ResNet18, ResNet34, and AlexNet. The results illustrate that our model performs with an accuracy of 0.9, recall of 0.825, precision of 0.971, and F1 score of 0.892.**

Keywords*— coronavirus disease 2019, CXR image, majority voting, deep learning, convolutional neural networks.

I. Introduction

The COVID-19 virus outbreak has created chaos around the world, impacting the economy, societies, political measures, and people's livelihoods. Infected people are rapidly growing globally, as shown in Fig. 1. The number of cases is 17 million people counted at the end of July 2020. Currently, the mortality rate is about 3.90% [1]. However, all of us acknowledge that this viral epidemic is more severe, rapid, and broader than it has been assessed and has murdered too many people around the world [2]. As a result, all frontline public health professionals, including public and private sector workers worldwide, have worked hard to heal this enormous impact. People must harmonize to keep social distancing, especially staying at home to stop the transmission and infection during this challenging time [3]. In parallel, innovations and technologies are identified as useful tools for making prevention and treatment more effective against COVID-19. These include the automatic inspection of chest X-ray images by using image recognition.

Image classification can become one of the first essential mechanisms that hospitals use to screen patients and identify people who are being infected with COVID-19 [4]. In addition to measuring the temperature, controlling, and tracking the data of people at risk of infection, separating COVID-19 from pneumonia using medical or X-rays images is an additional method of detecting illness. In the laboratory, the genetic material of the virus is detected by Reverse Transcription PCR (RT PCR), but it may be insufficient because the number of reagents is limited to the rapidly increasing number of patients. Image classification could support and confirm the results from the laboratory, save time, and perform the lab's results more accurately [5].

Deep Learning is a subset of machine learning which mimics the function of the human brain to do specific tasks. Deep learning could be applied for medical image classification, such as lung cancer and pneumonia [6]. Convolutional Neural Networks (CNNs) is a well-known model for image classification and segmentation [7]. CNNs are a class of deep learning that automatically learns the filters that humans have to do in traditional algorithms. This paper proposes an ensemble approach using CNNs and transfers learning to classify COVID-19 infection in X-ray images derived from medical image libraries. We modified three CNNs architectures, which are ResNet18, ResNet34, and AlexNet. These three models are applied in our voting ensemble.

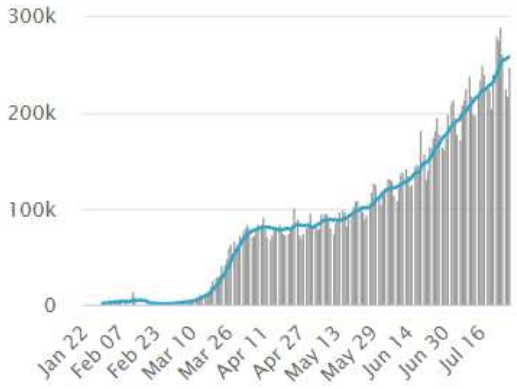


Fig. 1 Global daily new cases of COVID-19 with 7-day moving average (blue line) from January to July 2020 [1].

This paper's structure is the following: after describing the reason for COVID-19 image classification, Section II reviews related works. Section III includes the proposed methodology with the CNN-based voting ensemble transfer learning. In Section IV, we evaluate the proposal discussing the results assessing the suitability of the automatic COVID-19 image classification.

This work was supported by the European Commission, through the PHARAON project (ref. 857188), and the Spanish Government, through the projects ONOFRE-2 (ref. TEC2017-84423-C3-2-P) and Go2Edge (ref. RED2018-102585-T).

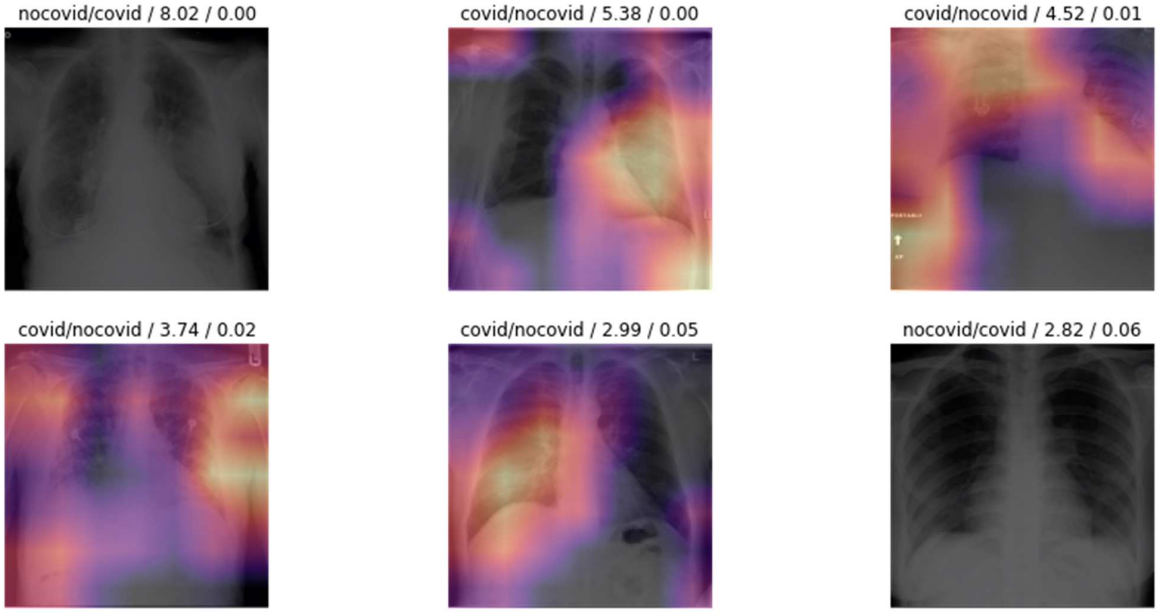


Fig. 2 Top losses images with heatmap.

II. Related Works

The emergence of the new coronavirus encourages the attention of medical image classification research to itself. Various researchers proposed methods to define the disease in X-Ray images such as A. Abbas et al. proposed Decay Transfer and Writing (DeTraC) [8], E. E. Hemdan et al. presented COVIDX-Net, consisting of seven CNNs models [9]. R. M. Pereira et al. propose multi-class and hierarchical classification to detect COVID-19 from normal and pneumonia chest X-ray images [10]. The pre-training model is applied to COVID-CAPS presented by P. Afshar et al. The results of all proposed models acquired high accuracy [11].

From the literature review, we found that all of them used CNNs as a core algorithm. Moreover, ensemble and pre-trained models are interesting and could improve medical image classification accuracy. As a result, we applied CNNs with two approaches: pre-trained and ensemble in our method, described in the next section.

III. Methodology

There are five steps in our research. First, data collection is describing in the dataset section. Data preparation is how we processed images with augmentation, including improvement of augmentation using heatmap from the top loss images. There are two main steps in modeling: separate model training and consolidate three models to the voting ensemble. In the end, the evaluation is presenting with chosen metrics.

A. Dataset

The dataset is taken from the chest x-ray images collected from two different sources: COVID-19 image data collection [12] and NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories [13]. The dataset obtains two classes of images: COVID-19 class, including 200 images, and non-COVID-19 class, with another 200 images. While positive COVID-19 images are in the first class, the second class includes healthy chest images and chest images with regular pneumonia in the same ratio.

B. Data Preparation

The dataset is divided into three parts: 60% for the training set, 20% as the validation set, and the remaining 20% as the testing set. After that, we trained the model with a few epochs and analyzed the incorrectly identified images (images with the highest loss value) using a heatmap, as shown in Fig. 2. This analysis was applied to determine image augmentation parameters. Image augmentation is a regularisation technique to support training models for computer vision. It performs small random transformations but not changing the core of the image. In this experiment, the parameters used in this process are do flip horizontal, do not flip vertical, maximum rotation = 25, and maximum zoom = 1.17. Fig. 3 shows a sample of normal and COVID-19 images extracted from the dataset.

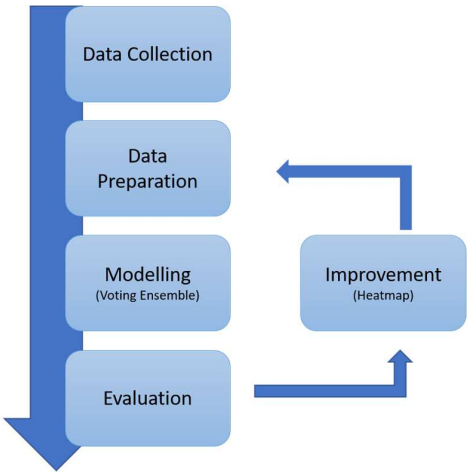


Fig. Methodology of this research

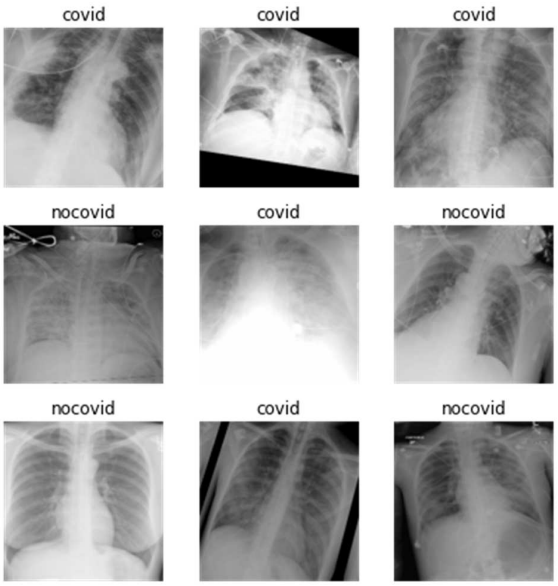


Fig. 3 Example images after image augmentation.

C. Data Modeling

We want to obtain the optimal algorithms in our study. Therefore, we compare 14 pre-trained CNNs models available in the FastAI library:

- ResNet18, ResNet34, ResNet50, ResNet101, ResNet152
- SqueezeNet1_0, SqueezeNet1_1
- DenseNet121, DenseNet169, DenseNet201, DenseNet161
- VGG16_bn, VGG19_bn
- AlexNet

The performance of the 14 models with train loss, valid loss, accuracy, and execution time is shown in the results included in Table I. The selection criteria followed considers the models with the highest accuracy of 3 models, ResNet18, ResNet34, and AlexNet. If there are models with the same accuracy, it will be determined by the valid loss.

Then, we selected the top three models by accuracy and validation loss scores. As presented in Table I, the top three models are AlexNet, ResNet18, and ResNet34. These three algorithms will be the voters in the majority voting system for COVID-19 classification.

- Alexnet is designed by A. Krizhevsky [14]. It is an eight-layer CNNs in which the first five are convolutional layers.
- ResNet is designed by K. He, X. Zhang, S. Ren, and J. Sun [15]. It is a convolutional neural network up to 152 layers deep. We picked ResNet18 and ResNet 34, which have 18 and 34 layers deep, respectively.

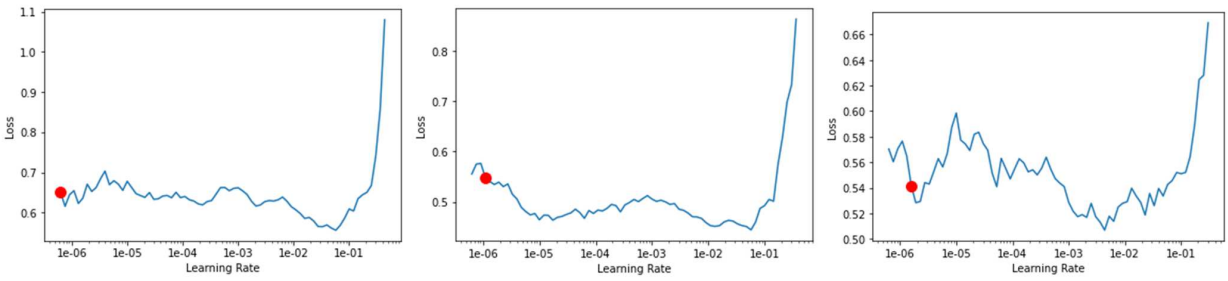
Before training models, we defined the optimal learning rate for each model, as shown in Fig. 4. After the model selection, every model has been trained for another 30 epochs with a predefined learning rate.

TABLE I. PERFORMANCE OF 14 REFERENCE MODELS

Model	Train loss	Valid loss	Accuracy	Time
resnet18	0.69976	0.75995	0.7625	0:22
resnet34	0.60526	0.8179	0.75	0:22
resnet50	0.91096	2.03975	0.6375	0:24
resnet101	1.0916	5.28986	0.6125	0:25
resnet152	1.24716	3.70274	0.5625	0:26
squeezenet1_0	0.79798	0.9064	0.75	0:22
squeezenet1_1	0.91698	3.36416	0.6125	0:22
densenet121	0.89914	2.03398	0.6375	0:25
densenet169	0.63103	1.15615	0.725	0:25
densenet201	0.91675	5.21541	0.5625	0:26
densenet161	0.71805	2.3503	0.5875	0:37
vgg16_bn	0.90201	2.13799	0.55	0:28
vgg19_bn	0.75693	0.88414	0.6	0:24
alexnet	0.90383	0.72282	0.8125	0:22

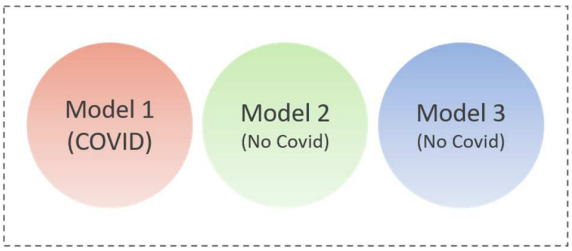
D. Voting Ensemble

Finally, the majority voting method was created with ResNet18, ResNet34, and AlexNet. The voting system is a simple system designed to improve the prediction performance of our models. Each model provides only one predicted result, outputting a 0 value for the positive COVID-19 cases and a 1 value for the negative COVID-19 cases. This voting schema avoids false positives when the predictions are not in agreement. For instance, if the three models do not have consensus output, the highest frequency value from models will represent the final output [16]. Fig. 5 shows the example of the final result with negative COVID-19 (top) and the final result with positive COVID-19 (bottom).



(a) Resnet18 (b) Resnet34 (c) AlexNet

Fig. 4 Learning rates of Resnet18, Resnet34 and Alexnet



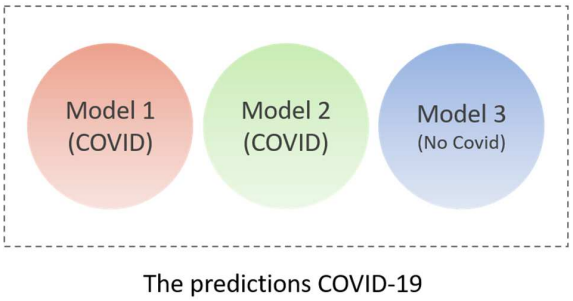


Fig. 5 Example outputs from voting ensemble when the predictions are not in agreement

E. Evaluation Metrics

The conventional method to evaluate the classification approach is to calculate metrics from the confusion matrix. The confusion matrix is a table with two columns and two rows that demonstrates the number of true positives – correct prediction of positive COVID-19, true negatives – correct prediction of negative COVID-19, false positives – incorrect prediction of positive COVID-19, and false negatives – incorrect prediction of negative COVID-19. Table II shows the confusion matrix.

TABLE II. CONFUSION MATRIX

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

We chose the accuracy, precision, recall, and F1 score to evaluate our model performance. The accuracy shows how many correct predictions our model outputs. The formula of accuracy is shown as:

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

Precision is a measurement that gives a ratio of correct positive predictions as compared to the positive predictions (true and false). The formula is shown as:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall is known as the sensitivity ratio of correct positive predictions as compared to the addition of true positives and false negatives. The formula is shown as:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1 score is a combination of recall and precision. The formula is shown as:

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

All metrics range from 0 to 1, being 1 the best possible score.

TABLE III. PERFORMANCE OF PROPOSED MODEL WITH ITS VOTERS

Metrics	**Voting ensemble**	**ResNet18**	**ResNet34**	**AlexNet**
Accuracy	**0.9**	0.825	0.8375	0.8625
Recall	0.825	0.725	**0.85**	0.75
Precision	**0.970588**	0.90625	0.829268	0.967742
F1-Score	**0.891892**	0.805556	0.839506	0.84507

TABLE IV. COMPARISON OF CONFUSION MATRICES FROM THE FOUR MODELS

	PREDICTED							
	Voting Ensemble		**ResNet18**		**ResNet34**		**AlexNet**	
	COVID	**NOCOVID**	**COVID**	**NOCOVID**	**COVID**	**NOCOVID**	**COVID**	**NOCOVID**
ACTUAL COVID	33	7	29	11	34	6	30	10
ACTUAL NOCOVID	1	39	3	37	7	33	1	39

IV. RESULTS

We applied the specific learning rates indicated in Section III. The learning rates for training ResNet18, ResNet34, and AlexNet models are e^{-7}, e^{-6}, and e^{-6}, respectively.

Table III shows that the voting ensemble model based on three CNNs voters performed well, with true negatives results equal to the ones obtained with AlexNet. Both the voting ensemble and the AlexNet model classified for non-COVID-19 class correctly at 97.5%, with only one incorrect prediction. As seen further in Table IV, ResNet34 overperformed our proposed model when attending to recall, presenting only one prediction different. The ranking for positive COVID-19 class prediction (TP cases) is ResNet34, our voting ensemble, AlexNet, and ResNet18. Furthermore, the ranking for negative COVID-19 class prediction (TN cases) is, first, our proposed model and AlexNet, and then ResNet18 and ResNet34.

As a result, the voting ensemble performed better than the other models in three metrics. The voting model obtained 90% accuracy, 82.5% recall, 97.06% precision and 89.19% F1-score. AlexNet, ResNet34, and ResNet18 performed at 2nd, 3rd, and 4th positions by accuracy, respectively.

V. CONCLUSION

The unexpected pandemic caused by COVID-19 disease is a global problem that all countries around the world are fighting. Numerous tools are developed, and some are still under development, to fight against the virus to prevent new cases and support treatments. Machine learning is identified as a useful supporting tool in the area of automatic medical image classification. It is a supporting tool for medical professionals to detect COVID-19 using chest X-ray images.

In this paper, we propose to use CNNs to identify COVID-19 cases from X-ray images. Our voting ensemble CNNs transfer learning method is able to efficiently classify COVID-19 from medical images. The voting system consists of ResNet18, ResNet34, AlexNet as its voters. Our proposed model performed better than its individual voting members with an accuracy of 90%, recall of 82.5%, the precision of 97.06%, and F1-score 89.19%. The F1-score of our proposed model performed over a multi-class approach of [11] and performed equivalently to a hierarchical classification of [11] and healthy image classification of [9]. This means that it is an efficient supporting tool to support common PCR tests when X-ray imaging is possible, and a number of patients must be checked. Accuracy detected is, in fact, in the same range of PCR tests currently being used. Moreover, this system is able to detect patients not suffering from COVID-19, but at risk of developing the illness critically, due to other respiratory problems.

As a part of our future work, we plan to extend our dataset to better adapt our prediction models and improve accuracy and reliability. Also, we intend to pre-train the model with other chest X-ray images to improve performance because pre-trained models in FastAI is based on ImageNet but not X-Ray images. Finally, we plan to develop and experiment the system with a more complex and weighted (discounted) voting system to get better prediction results.

ACKNOWLEDGMENT

We would like to express my very great appreciation to Associate Professor Dr. Esteban Egea López for his valuable and constructive suggestions for this research work. His willingness to give his time so generously has been very much appreciated.

REFERENCES

[1] Worldometer, "Coronavirus Update (Live) from COVID-19 Virus Pandemic." Accessed: Jul. 30, 2020. [Online]. Available: https://www.worldometers.info/coronavirus

[2] WHO, "Coronavirus Disease (COVID-19) Dashboard," Accessed: Jul. 30, 2020. [Online]. Available: https://covid19.who.int/

[3] WHO, "Argentina: There is no economy without health," Accessed: Jul. 30, 2020. [Online]. Available: https://www.who.int/news-room/feature-stories/detail/argentina-there-is-no-economy-without-health

[4] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no.5, pp. 1207–1216, May 2016, doi: 10.1109/TMI.2016.2535865.

[5] LabCorp, "ACCELERATED EMERGENCY USE AUTHORIZATION (EUA) SUMMARY COVID-19 RT-PCR TEST (LABORATORY CORPORATION OF AMERICA)," Accessed: Jul. 24, 2020. [Online]. Available: https://www.fda.gov/media/136151/download

[6] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol.113, no.1, pp. 202–209, January 2014, doi:10.1016/j.cmpb.2013.10.011.

[7] L. Wang and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," 2020. [Online]. Available: arXiv:2003.13815.

[8] A. Abbas, M. M. Abdelsamea, M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020. [Online]. Available: arXiv:2003.09871.

[9] E. E. Hemdan, M. A. Shouman, M. E.l Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," 2020. [Online]. Available: arXiv:2003.11055.

[10] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla, and Y. M.G. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine*, vol.194, pp. 1-18, 2020.

[11] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, "COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 cases from X-ray Images," 2020. [Online]. Available: arXiv:2004.02696.

[12] *COVID-19 Image Data Collection: Prospective Predictions Are the Future*, J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, 2020, [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[13] *NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories*, National Institutes of Health, Sep. 27, 2017, [Online]. Available: https://nihcc.app.box.com/v/ChestXray-NIHCC

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," In *European conference on computer vision*, 2016, pp. 630-645.

[16] A. M Rossetto, W. Zhou, "Ensemble Convolution Neural Network with a Simple Voting Method for Lung Tumor Detection," In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017 pp. 729–734, doi:10.1145/3107411.310

Estimation of Oil Content in Oil Palm Fresh Fruit Bunch by Its Surface Color

Sutat Sae-Tang
Image Processing and Understanding Research Team (IPU)
Artificial Intelligence Research Group (AINRG)
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
sutat.sae-tang@nectec.or.th

***Abstract*— Oil palm is one of the potential tree crops in Thailand. However, the production of oil palm has been experienced many aspects. Price factor is also one of the problems. Price of oil palm depends on the amount of oil content in the oil palm fruit which are estimated by an expert. The main consideration is the ripeness of the oil palm fresh fruit bunches. An expert determines using its surface color. A different experience of experts leads to a different estimation. The problem may be solved using the chemical analysis methods which more accurate. However, it takes time and uncomfortable. In this research, artificial intelligence (AI) will be applied to estimate the oil content in a fresh fruit bunch (FFB). Two popular types of oil palms in Thailand are used in this work. The Nigrescene fruit, color varies from dark purple to red orange depending on its gene and ripeness. The Virescene fruit, color changes from green to orange. The surface color of an oil palm fruit and structure of the bunch were considered as the feature set. An oil palm FFB image from a smartphone camera was fed to the model for predicting the oil content in FFB. Several models such as multi linear regression, artificial neural network and convolution neural network will be observed. The measure of the quality's model uses the root mean square error (RMSE). The convolution neural network produces the average of RMSE at 7.27 for Nigrescene and at 4.83 for Virescene.**

Keywords*— *Oil palm, Oil content in fresh fruit bunch (FFB), Ripeness, Surface color, Convolution neural network

I. Introduction

Oil palm with the scientific name of "*Elaeis guineensis Jacq.*" is an important industrial crop in Thailand. Oil palm bunches can be used as raw material in the process of palm oil production as a form of safe edible vegetable oil, margarine and shortening. They can be used not only in food products, but also in non-food products, which are the basic ingredient in many manufacturing products such as soap, candle. The kernel can be used as a main ingredient in cosmetic manufacturing, as well. An oil palm fresh fruit bunch (FFB) may contain up to 2,000 fruits with an individual weight of 3 – 30 grams and 2 – 5 centimeters in size. There are about two species more popularly grown in Thailand: Nigrescene and Virescene fruit. The Nigrescene fruit, color varies from dark purple to red orange depending on its gene and ripeness. The Virescene fruit, color changes from green to orange. The oil content for different stages of FFB ripeness also varies, and it is commonly stated as oil-to-bunch ratio ([1], [2] and [3]). As the oil content of an oil palm FFB is a main indicator of the ripeness, it is crucial that the FFBs are harvested at the optimum ripeness.

At present we use an expert to estimate the ripeness of an oil palm FFB. A different experience of experts leads to a different estimation. The problem may be solved using the chemical analysis methods which more accurate. However, it takes time and uncomfortable.

However, many attempts using image processing and machine learning have been made to solve the challenge. In [4] and [5], the linear regression models are proposed for estimating the oil content in oil palm fruits. The RGB values of an oil palm fruit image are used as a set of features. The L*a*b* color space values are also tested and compared the results with the RGB. In [6], The RGB color model and artificial fuzzy logic is introduced. The purpose of the grading system is to distinguish among the three different classes of oil palm fruits which are under-ripe, ripe and overripe. In [7], The ripeness classification of oil palm fruits using artificial neural network is implemented. The MLP neural network is segregated the fruits into four ripeness categories: unripe, under-ripe, ripe and overripe. The results of models trained with RGB images and reduced features based on the Principal Component Analysis (PCA) are compared.

Moreover, other approaches also have been introduced such as [8] using four-band optical sensor. As the reviews, most of them tried to classify the ripeness stage of oil palm fruits, such unripe, ripe and overripe, rather than to estimate the oil content of an oil palm FFB. A few tried to predict the oil content from the mesocarp (a piece of an oil palm fruit), but not the oil-to-bunch (oil pre bunch).

The objective of this work is to develop a method for automatic estimated the oil content in an oil palm FFB based on its surface color in a natural light environment. The color measurement was based on RGB, HSV and L*a*b* color spaces of oil palm FFB images. The results were presented and discussed.

The rest of the paper is presented as follows: In Section II, the process for collecting the oil palm FFB pictures and steps involved in developing the methods are explained. Then the

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

experimental results are presented and discussed in Section III. The conclusion and future work are given in Section IV.

II. Materials and Methods

In this section, the collection of an oil palm FFB picture is described concisely. The prediction and evaluation methods are also explained.

A. Image Acquisition and Labelling

Two popular grown oil palms in Thailand (Nigrescene and Virescene) were used in this work. An average of 25-kilogram weight of an oil palm FFB is chosen. Different stages of FFB ripeness was considered. The total of 394 samples was used. The oil palm FFB images were captured from a digital cellphone camera in a controlled light (in a semi-light controlled box) and natural light environment. The front and back side of each bunch were captured. An example is shown in Fig. 1.

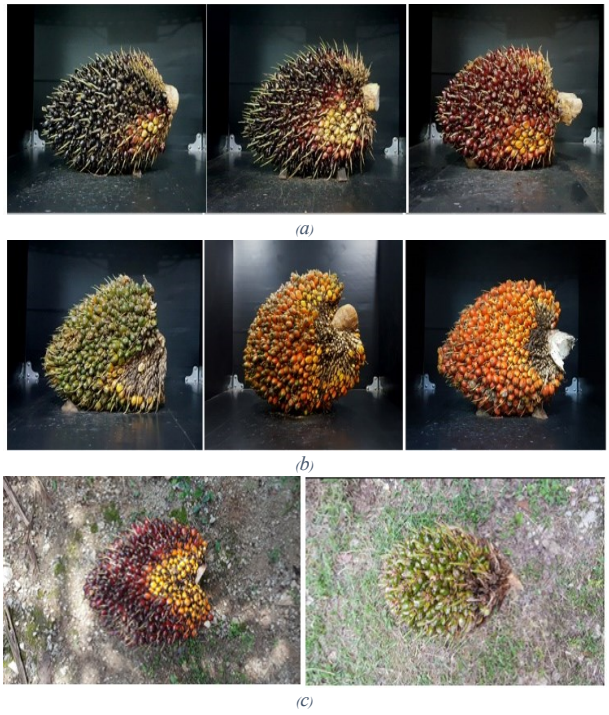


Fig. 1 Examples of oil palm FFB image: (a) Nigrescene (raw, semi-ripe and ripe), (b) Virescene (raw, semi-ripe and ripe) in a controlled light, and (c) Nigrescene and Virescene in a natural light.

In the process of labelling, the bunch analysis and oil palm extraction were applied. Each FFB is consist of many parts, for examples, core, palm berry and seed. The mesocarp is the important part for oil extraction. The more mesocarp will yield more oil as well. Laboratory analysis (Soxhlet extraction method) were carried out simultaneously using the dry mesocarp samples to determine the change in oil content. Bunch analysis were employed to convert the percentage of oil content from the dry mesocarp to the percentage of oil-to-bunch.

B. Prediction and Evaluation Methods

At present, artificial intelligence has been widely used for data analysis, classification and prediction in many domain including agriculture. In this section, several methods were selected for comparing the performance among them.

Multi linear regression (MLR): The MLR, also known simply as multiple regression, is a basic statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of MLR is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. The MLR was performed as a baseline method in order to compare with other advance methods.

Artificial neural network (ANN): The ANN has been generally applied to map input patterns with their desired outputs. The practical applications are far and wide, ranging from classification to prediction and visualization. In this work, the multilayer perceptron (MLP) neural network with backpropagation learning, a commonly used ANN architecture, was investigated.

Convolutional neural network (CNN): The CNN is a popular deep learning algorithm designed for image analysis. It always consists of two basic operations, namely convolution and pooling. The convolution operation using multiple filters is able to extract features (feature map) from the data, through which their corresponding spatial information can be preserved. The pooling operation, also called subsampling, is used to reduce the dimensionality of feature maps from the convolution operation. Max pooling and average pooling are the most common pooling operations used in the CNN. Nowadays, many pre-trained CNN models have been presented ([9], [10] and [11]), but most of them aim to solve the classification problems. In this work, the oil-to-bunch ratio is predicted from an oil palm FFB picture. It is a regression problem, so a new CNN will be built from scratch rather than use the transfer learning technique.

Evaluation Metrics: The evaluation metrics are used to measure the quality of the statistical or machine learning model. Evaluating machine learning models or algorithms is essential for any project. There are many different types of evaluation metrics available to test a model. Mainly, these can be grouped into two types: evaluation metrics for classification and for regression problems. These include classification accuracy, logarithmic loss, confusion matrix, and others. As mention above, this research aims to solve a regression problem. The mean square error (MSE) seems to be suitable for this task. However, the root mean square error (RMSE) was chosen rather than the MSE because of easy to understand. The average of RMSE is calculated according to Equation (1).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2} \quad (1)$$

where $\hat{Y}_i$ is the prediction

Y_i is the target

n is the total of samples

III. Experimental Settings and Results

In this section, the oil palm FFB dataset used in the experiments are described. The comparison of the color space models is presented The experimental results consisting of the multi linear regression, multilayer perceptron neural network and convolutional neural network are shown and discussed

A. Oil Palm FFB dataset

In the experiments, the 394 oil palm FFBs were collected in different stages of ripeness. These were divided into 196 and 198 bunches from Nigrescene and Virescene respectively as shown in detail in TABLE I.

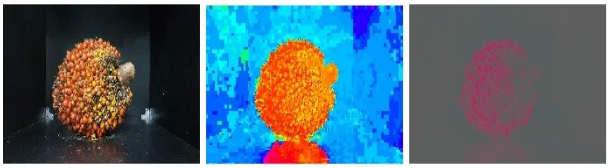
TABLE I Number of oil palm fresh fruit bunches.

Type	Stage of ripeness		
	Raw	Semi-ripe	Ripe
Nigrescene	65	65	66
Virescene	65	68	65

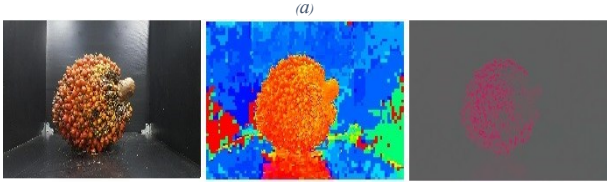
Each bunch was captured with a digital mobile phone camera in a controlled light and natural light environment and in multiple directions (front, back and others). The oil palm FFB dataset consists of 5,568 images from 394 subjects. The resolution of images in the dataset is 1,328x747 pixels. The dataset was separated into two groups. The first group, namely indoor dataset, consist of the images in a controlled light environment only. The second, called outdoor dataset, consist of the images in a natural light environment only. The indoor dataset was divided into training (70%) and test sets (30%) randomly, which the whole outdoor dataset was set as another test set.

B. Color Feature Extraction

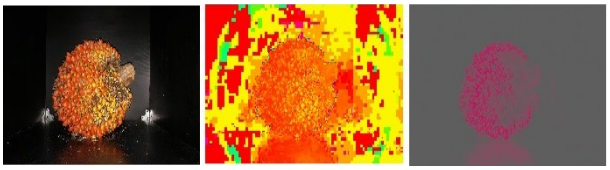
The color features of the oil palm FFB image are analysed based on the HSV and L*a*b* color space models rather than the RGB color space models. The Fig. 2 shows images in RGB, HSV and L*a*b* color space. The RGB image was converted into HSV and L*a*b* image. The only H (hue) channel was display for the HSV image while the only a* (green to red) channel was display for the L*a*b* image. Based on the experiments, the L*a*b* image can be performed very well in many light environments.



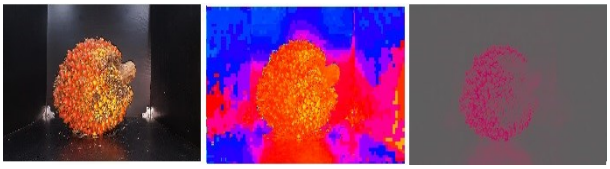
(a)



(b)



(c)



(d)

*Fig. 2 Oil palm FFB images in a controlled light (RGB, HSV and L*a*b*): (a) without LED light, (b) with LED light, (c) using flash, without LED light, and (d) using flash, with LED light.*

C. Experimental Results

Frist of all, the MLR was trained as a baseline model. The L*a*b* images were used in the experiment. The a* and b* channel were chosen to represent the color of an oil palm FFB. From the center of the oil palm FFB, 323 positions in the image are chosen as the training feature. The performances of the MLR model are shown in TABLE II. As the results, The MLR model performed not very well on the indoor2 dataset as well as outdoor dataset that the unseen data.

TABLE II Average of RMSE for MLR model in a controlled light and in a natural light.

Dataset	RMSE of MLR	
	Nigrescene	Virescene
Indoor1	2.58	1.49
Indoor2	10.15	9.70
Outdoor	10.97	10.29

Next, the MLP model was evaluated on the same dataset. Fig. 3 shows the structure of MLP model in the experiment. It is a fully connected in each layer, consisting of one input layer (646 nodes), six hidden layers (323, 323, 160, 80, 40 and 20 nodes, respectively, and one output layer (1 node).

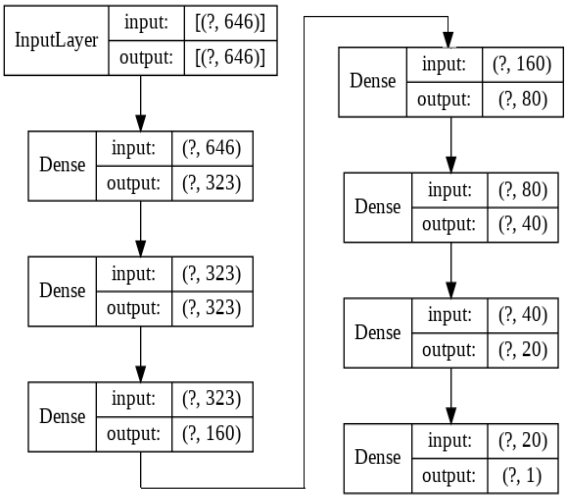


Fig. 3 Structure of MLP neural network.

The results are shown in TABLE III. For the results, it shows that the MLP model performed better than the MLR model. However, the problem still occurs on the outdoor dataset as seen on high values of RMSE.

TABLE III Average of RMSE for MLP model in a controlled light and in a natural light.

Dataset	RMSE of MLP	
	Nigrescene	Virescene
Indoor1	1.19	0.66
Indoor2	8.97	9.34
Outdoor	10.39	10.13

Finally, the CNN model was introduced. The CNN model normally applied for the image classification problem. In this case, the CNN was created for solving the regression problem was shown. The input image for the CNN model is a 256x256 L*a*b* and HSV images. The model consists of five convolution layers with 16, 32, 64 and 128 filters, two dense layers (16, and 4 nodes), one output layer (1 node). The structure of CNN model is shows in Fig. 4.

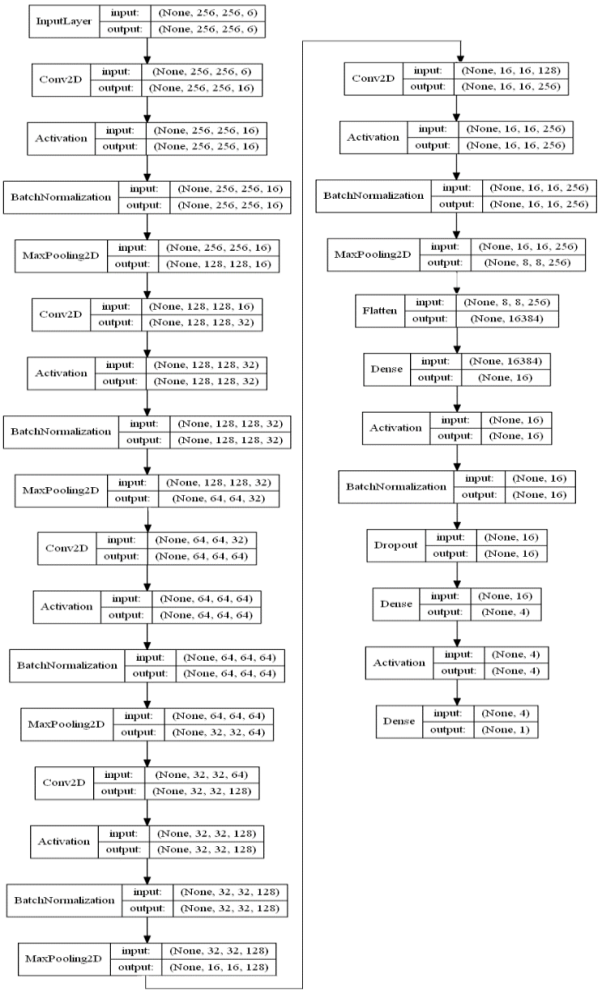


Fig. 4 Structure of CNN.

In TABLE IV, the results show that the CNN model performed the best among the purposed models. Although the RMSE of indoor1 dataset for the CNN higher than the RMSE of indoor1 dataset for the MLP, the results of the unseen data (indoor2 and outdoor) are better. It could be implied that the CNN model may be flexible and suitable for predicting an oil palm FFB in real world.

TABLE IV Average of RMSE for CNN model in a controlled light and in a natural light.

Dataset	RMSE of CNN	
	Nigrescene	Virescene
Indoor1	7.51	3.39
Indoor2	6.41	3.30
Outdoor	7.90	7.81

IV. Conclusions

The main objective of this paper is to predict the oil-to-bunch ratio of an oil palm FFB from its image. First, the dataset and labelling are created. Second, three different models, the multi linear regression (MLR) as a baseline, multilayer perceptron (MLP) neural network and convolutional neural network (CNN) were presented and evaluated. Finally, the CNN model outperformed other models.

In future work, the investigation of CNN, such as a varied number of convolution and dense layer, to improve the performance will be performed. The oil-to-bunch of oil palm FFB prediction system will be implemented as an intelligent agricultural chatbot.

Acknowledgment

I would like to express my deepest appreciation to all those who provided me the possibility to complete this work. This research was supported by the Agricultural Research Development Agency (ARDA). A special thanks goes to my co-researchers at Suratthani Oil Palm Research Center who help me for data collection and gave suggestion about the oil palm industrial and the oil palm FFB manual ripeness classification.

References

[1] S. Balasundram, P. Robert and D. Mulla, "Relationship Between Oil Content and Fruit Surface Color in Oil Palm (Elaeis guineensis Jacq.)," *Journal of Plant Sciences,* vol. 1, no. 3, pp. 217-227, 2006.

[2] V. Ormzubsin, P. Jumradshine and S. Promcheau, "Effect of oil palm bunch ripeness on bunch component and palm oil quality (in Thai)," *Khon Kaen Agricultural,* pp. 694-700, 2015.

[3] M. S. M. Alfatni, A. R. M. Shariff, H. Z. M. Shafri, O. M. B. Saaed and O. M. Eshanta, "Oil Palm Fruit Bunch Grading System Using Red, Green and Blue Digital Number," *Journal of Applied Sciences,* vol. 8, no. 8, pp. 1444-1452, 2008.

[4] K. Sunilkumar and D. S. S. Babu, "Surface color based prediction of oil content in oil palm (Elaeis guineensis Jacq.) fresh fruit bunch," *African Journal of Agricultural Research,* pp. 564-569, 2013.

[5] M. H. Razali, "A Novel Technology in Malaysian Agriculture," *Advances in Computing,* pp. 1-8, 2012.

[6] Z. May and M. H. Amaran, "Automated Oil Palm Fruit Grading System using Artificial Intelligence," in *International Joural of Video & Image Processing and Network Security*, 2011.

[7] N. Fadilah, J. Mohamad-Saleh, Z. A. Halim, H. Ibrahim and S. S. S. Ali, "Intelligent Color Vision System for Ripeness Classification of Oil Palm Fresh Fruit Bunch," *Sensors,* pp. 14179-14195, 2012.

[8] O. M. B. Saeed, S. Sankaran, A. R. M. Shariff, H. Z. M. Shafri, R. Ehsani, M. S. Alfatni and M. H. M. Hazir, "Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system," *Computers and Electronics in Agriculture,* vol. 82, pp. 55-60, 2012.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision," 2017.

[10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, 2015.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," in *Conference on Computer Vision and Pattern Recognition*, 2015.

A Surveillance System for Children Stuck Inside the Car with Embedded System Technology

Arthi Yooyen and Siriruang Phatchuay
Computer Engineering Technology
RajamangalaUniversity of Technology
Rattanakosin Prachuap Khiri Khan,Thailand
Arthit.yoo@rmutr.ac.th,siriruang.pha@rmutr.ac.th

***Abstract*– The objective of this research is to surveillance and develop a system of warning system for the safety of living organisms in the general automobile passenger compartment. By applying the PIR Motion Sensor for detects infrared radiation waves to detects living organisms suitable for sticking inside the car cabin. As well as the application of the GSM Module SIM800L and Buzzer as an alternative to the security alarm. By working principle, using an embedded system to control and operate the Arduino Uno R3 microcontroller with the accuracy of the system to detection for 80 %.**

Keywords–surveillance, detects, infrared, embedded system

I. INTRODUCTION

Nowadays, news of children trapped inside cars is frequent, such as schoolchildren stuck in shuttle buses or parents forget their children stuck in the car. [1]Which causes harm to children It can be fatal. In Thailand, according to disease surveillance data from the Bureau of Non-communicable Diseases, Department of Disease Control, it was found that in the 5 years (2014-2018) it was found that there were 106 events in which children were left in their cars, so building surveillance equipment to reduce damage and prevent it. Would be very necessary ,But when looking at the current object detects technology It is constantly evolving and is modernity. But there is an important factor, the price is expensive, it is difficult for the general public to buy and install in every car. Causing the researcher to realize the solution to the problem Is the development of surveillance equipment for installing in cars at low cost. By developing a device for safety warning and living organisms in the common car cabin, as well as applying the PIR Motion Sensor to be suitable for the interior of the car cabin.[2] GSM Module SIM800L and Buzzer [3]are also used as an alternative to security alarms and asked for help from the car owner Mentor teacher Parents or people in that area It can also reduce the mortality rate of children as well. Which can be installed and used for real.

II. RELATED WORK AND THEORIES

A. PIR sensors operation

PIR sensors are also called Pyroelectric infrared sensors, passive infrared sensors, or IR motion sensors, which detect differences in temperature, heat radiation, human body or animal. [4, 5] By PIR operation, the infrared sensor detects the distance of the object with infrared rays. When the beam detects an object, by returns to the receiver at an angle after the reflection. The method of triangulation is as shown in Fig 1.

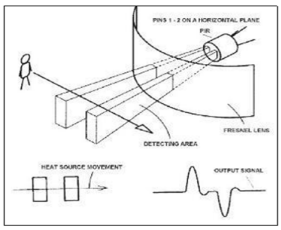


Fig 1. PIR sensors for detection [4]

B. GSM Module system design

The SMS checking process is starting from initialization. Then, check SMS, if it is "TURN ON" then the process will proceed to the "System = 1" check phase. If the "System = 1" process transmitted to the PIR 1 sensor, if the PIR 1 sensor catches human movement the SMS will be sent to the mobile phone by this message "Motion Detected", if it does not find any movement then the process will continue checking the PIR 2 sensor, it also happens to other PIR sensors. If there is getting human movement, the PIR sensor will be forwarded to the Arduino system.[6]

C. Buzzer circuit diagram

Buzzer speaker integrates each part of the Arduino Uno R3 Buzzers microprocessor circuit to complete each function. Buzzer speakers can be found in most common alarm devices.[7]

D. Arduino Uno R3

Arduino is an open source electronic board that has input and output and there is a major component for program controllers that can be written and deleted in a special way that is an ATMega328 based microcontroller chip. Microcontroller itself is a chip or IC (Integrated circuit) that can be programmed using computer. The recorded program aims to enable electronic circuits to read inputs, process and then produce outputs as desired. The result can be a signal, voltage, light, sound, vibration, movement.[8-10]

E. Related Work

Alathari, B., et al., [11] from this research, the concept was obtained and applied in circuit design by using the actual circuit in the file. Virtual environment It consists of five PIR motion sensors used to detect any movement, working within a 360 radius. When the PIR sensors detects motion, the servo motor will rotate to the direction of that movement. Also have Five LEDs, there LEDs indicate which sensor is working when it goes high.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

Mahzan, N., et al. [12] presented the house fire alarm design with an Arduino based system by the way of GSM module with the purpose of home security whose main point is to avoid fire. Accidents happen to residents and property in the home as well. Use Arduino the Uno board together with the ATmega328 chip when the system detected the temperature 400. Notification message will be sent to users. Via short message service (SMS) via GSM module

Joseph, A., V. Parmar, and V. Bagyaveereswaran.[13] presented the robot simulation to detect humans trapped behind walls or under the rubble, which can help the rescue team during it. Natural disasters such as earthquakes Using the Arduino Uno microcontroller.

Al Fani, H., et al.,[14] proposed to watch and look after the parents' babies. Using a sound monitoring system in the children's room It uses an Arduino Atmega 328 microcontroller to detect interference and send a signal via a buzzer to alarm.

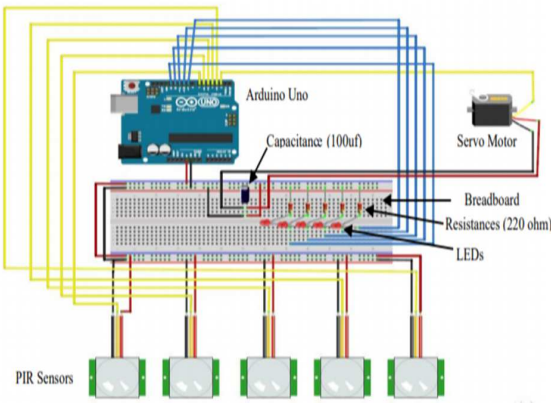


Fig 2. concept Circuit Diagram of the Proposing System Using Fritzing Software[11]

III. Propose Methodology

The work in this algorithm of operations to consists of analysis and design of embedded system parts that can be actually installed in the vehicle for surveillance of children trapped in the car.

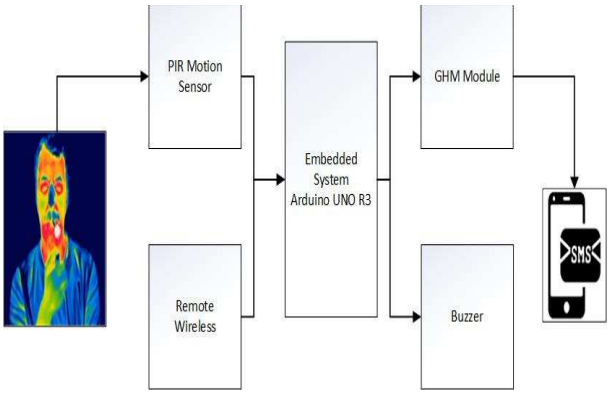


Fig 3. Block Diagram

The working principle of the living organism safety alarm system that starts when the system is turned on by pressing the remote lock from the Wireless Remote and PIR Motion Sensor. Infrared to detection living organisms, if not found, PIR will continuously detect infrared waves until the system is shut down, but if finds organisms in 5 minutes, the processor will send commands to GSM Module SIM800L. The send a message to phone number specified in the system five minutes later after the message has been sent, the system has not been shut down, the processor will continue to send a command to the Buzzer to sound an alarm for assistance from the area.

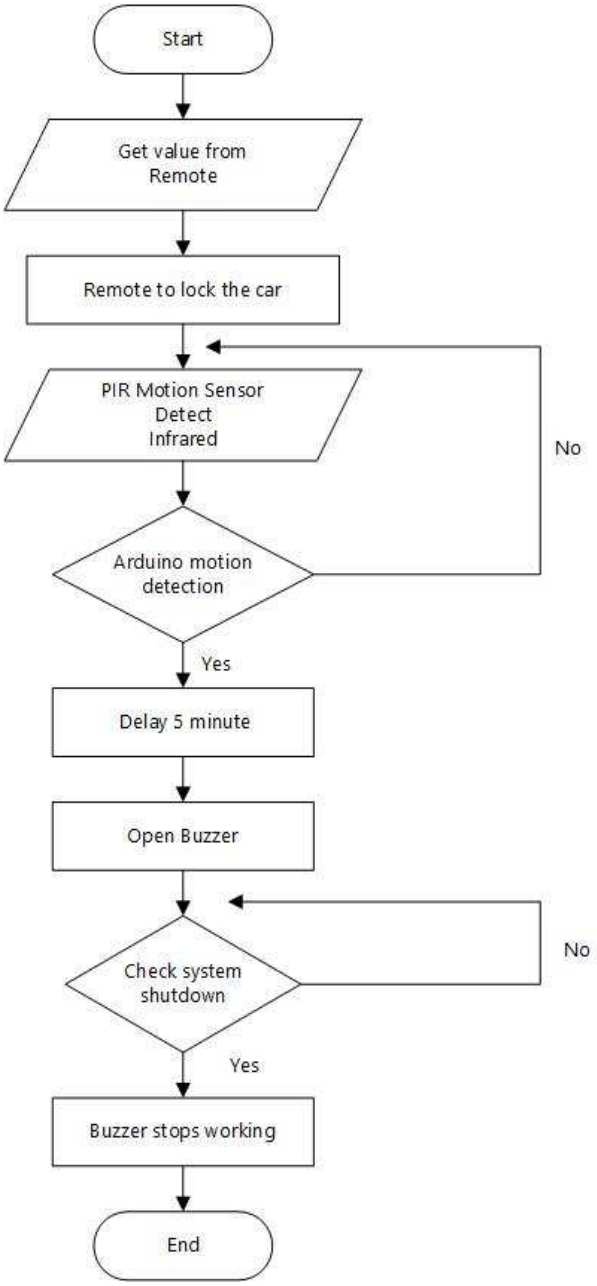


Fig 4. Flowchart of working algorithm

Fig 4. Show the flowchart of the working algorithm. Motion detection and send alert messages to the user's phone, and if there is no response within 5 minutes, the system will send the next command to the buzzer to sound the alert for further assistance from that area.

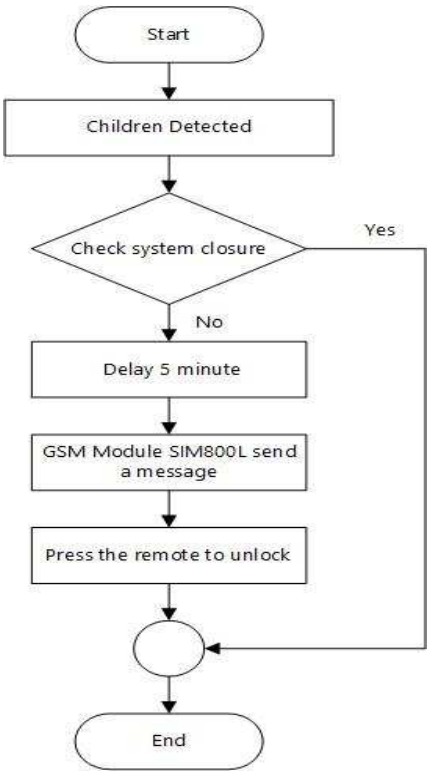


Fig 5. Flowchart of GSM Module SIM800L send a message

Fig 5. Show the flowchart of the working algorithm. The work system start to when PIR Motion Sensor detects organisms. Will check the system shutdown by pressing the remote unlock the car If yes, it will be finished. If not, the delay start of 5 minutes, if after 5 minutes and the GSM Module has not been shut down, SIM800L will send a message to the phone number specified in the system and if the remote is pressed, the car will be unlocked.

```
#include <Wire.h>
#include <TimerOne.h>
#include <SoftwareSerial.h>
#include <LiquidCrystal_I2C.h>
const  byte  RX_PIN      = 10;
const  byte  TX_PIN      = 11;
String  stLock = "";
String  valRemote;
constintmaxMyNumber = 1;
String MyNumber[maxMyNumber] = {"0643140822"};
SoftwareSerialmyGSM(RX_PIN, TX_PIN);
voidsms()
{
while(true)
  {
if(stLock == "") setup();
if(stLock == "") break;
lcd.clear();
lcd.setCursor(0, 0);
lcd.print("!!!SMS!!!");
delay(1000);
lcd.clear();
lcd.setCursor(7, 1);
lcd.print("!!!SMS!!!");
delay(1000);
lcd.clear();
lcd.setCursor(0, 1);
lcd.print("!!!SMS!!!");
delay(1000);
lcd.clear();
lcd.setCursor(7, 0);
lcd.print("!!!SMS!!!");
delay(1000);
checkRemote();
if(countMotionTime == 70)break;
  }
}
```

SECTION 1 : *Hardware design*

The system consists of :

- PIR Motion sensor
- Arduino UNO R3
- GSM Module SIM800L
- Wire-less Remote Control Module
- DC to DC Step Down Module
- Buzzer

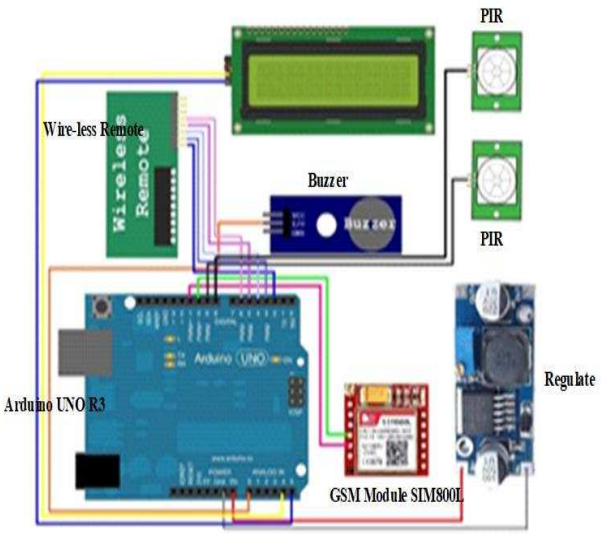


Fig 6. Hardware Circuit Design

SECTION 2 : *Hardware circuit overview*

The Circuit design can be designed by feeding 12V power from the supply to the board. Arduino then supplies the power from the Arduino board from the VIN pin and GND pin of the Arduino to the voltage adjustment circuit. DC to DC Step Down LM2596 Module to adjust the voltage not to cause 9 Volt to supply in various devices to provide steer and sufficient for the voltage required for the equipment. This will have a Wireless Remote Control Module circuit like a switch on - off, then it will be PIR Motion Sensor detecting infrared from living things. When the PIR Motion Sensor detects it, it sends a value. Arduino will continue to process and send commands to GSM Module SIM800L to send a notification message to the specified destination number. After that the processor Arduino sends a command to the buzzer to sound an alarm, ask for help from the local area, and the Liquid Crystal Display (LCD) to show the value through the screen.

SECTION 3 : *Microcontroller*

The microcontroller board (Arduino) connects devices to control various functions. Of the living organism safety alarm system are PIR Motion Sensor, GSM Module SIM800L, Buzzer and Liquid Crystal Display (LCD).

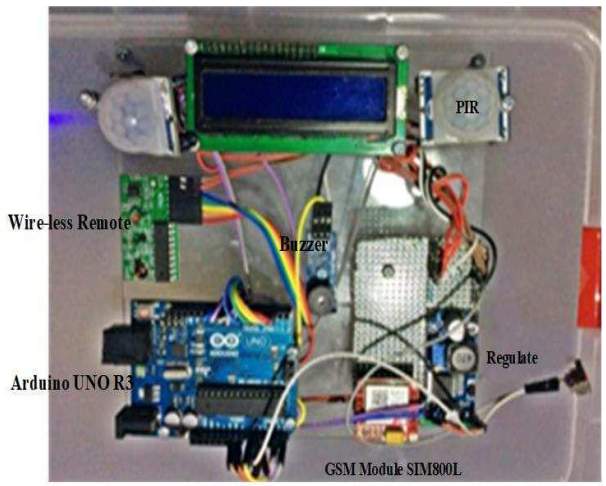


Fig 7. Alarm System Control Circuit Board

The system uses 12V power from the supply to the Arduino board, then the power from the Arduino board from the VIN pin and GND pin of the Arduino to feed the circuit to adjust the DC to DC Step Down LM2596 Module to adjust the voltage not to 9. Volt goes to party in different equipment. The steer and sufficient to the voltage required by the device, which will have a circuit Wireless Remote Control Module as a switch on - off, then a PIR Motion Sensor, a device that detects infrared radiation from living things. PIR Motion Sensor detects and sends the value to the Arduino to process and sends a command to the GSM Module SIM800L to send a notification message to the specified destination number. After that, the Arduino processor sends a command to the buzzer to sound an alarm and the Liquid Crystal Display (LCD) to show the value through the screen.

SECTION 4 : *System design*

PIR Motion Sensor will be installed in 2 positions: Front and Rear.The detection radius of the PIR Motion Sensor is 70 degrees.

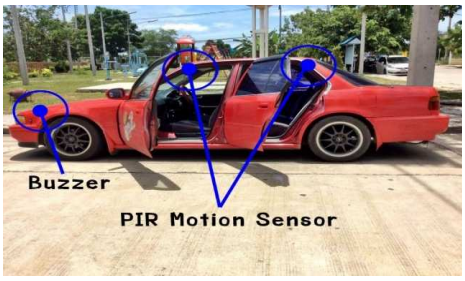


Fig 8. Device Installation Position

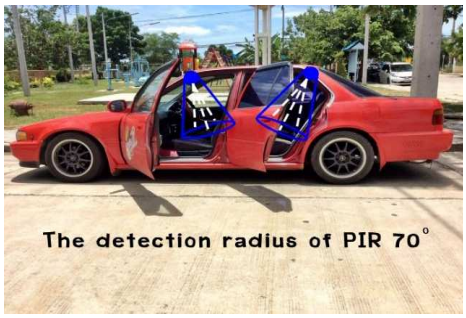


Fig 9. Detection Radius of the Module PIR Motion Sensor

SECTION 5 : *The Install and Test of the system*

Fig. 10 Position PIR Motion Sensor the Front

Fig 11. Position PIR Motion Sensor the Rear

Fig 12. Position Microcontroller Box

IV. Experimental Results

From the test results of the system workflow, using 12V power supply from the source to the Arduino board, then power from the Arduino board from the VIN pin and GND pin of the Arduino to feed the circuit to adjust the voltage DC to DC Step Down. LM2596 Module to adjust the voltage not to cause 9 Volt to supply in different devices. The steer and sufficient pressure required for the equipment, the LCD screen will show "STANBY" indicating the ready state. Until the remote lock is pressed from the Wireless Remote to turn on the system. Shown as in Table 1.

TABLE I. Experimental Results

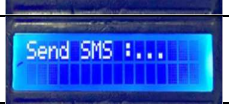
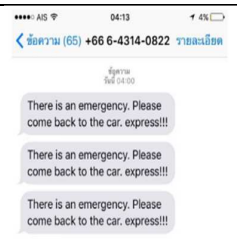
System performance		
Front		***Rear*** 
Step : 1		
Step : 2		
Step : 3		
Step : 4		
Step : 5		
Step : 6		
Step : 7		
Step : 8		
Step : 9		
Step : 10		

TABLE II. System Test Results

Test Surveillance system	**Example sequence**	**Message Results**	**Call Results**	**Buzzer effect notification**
Surveillance system and alerting children in the cabin A sample of 10 people took 10 minutes each.	1	✓	✓	✓
	2	✓	✓	✓
	3	✓	✓	✓
	4	✓	✓	✓
	5	×	×	×
	6	✓	✓	✓
	7	×	×	×
	8	✓	✓	✓
	9	×	×	×
	10	✓	✓	✓
	11	✓	✓	✓
	12	✓	✓	✓
	13	✓	✓	✓
	14	✓	✓	✓
	15	✓	✓	✓
	Percentage	80%		

From Table II, the test No. 5 , No.7 and No.9 the subject may not sit on the spot, so PIR Motion Senses cannot detect infrared radiation.When PIR Motion Senses are inactive, GSM Module SIM800L and Buzzer will not work.

V. CONCLUSION

The testing the operation of the system, if PIR Motion Senses detects organisms, an alarm system can be activated. The system was accuracy 98% by having the subject sit in the cabin for 10 minutes to allow PIR Motion Senses to detection living organisms or not. Specified in the system after another 5 minutes it will turn on the alarm sound if the system has not been shut down.

ACKNOWLEDGMENT

This research was supported by a grant from the Department of Computer Engineering Technology, Faculty of Industry of Technology, Rajamangala University of Technology Rattanakosin.

REFERENCES

[1] Shirkande, S., S. Survase, and T. Vitekari, *Design of Sensors based Hyperthermia Car Alert Surveillance System using GSM.* International Journal of Advance Research, Ideas and Innovations in Technology, 2018. **4**(2).

[2] Hazizan, A., N.M. Lazam, and N. Hassan. *Development of Child Safety Car Alert System Using Arduino and GSM Module*. in *IOP Conference Series: Materials Science and Engineering*. 2020. IOP Publishing.

[3] Lubis, A.M., et al., *Smart Dust Bin With Automatic Lid and Buzzer Alarm.* Journal of Electrical And Electronics Engineering, 2020. **2**(2): p. 50-53.

[4] Ismail, R., Z. Omar, and S. Suaibun. *Obstacle-avoiding robot with IR and PIR motion sensors*. in *IOP Conference Series: Materials Science and Engineering*. 2016.

[5] Erden, F., et al., *Sensors in assisted living: A survey of signal and image processing methods.* IEEE Signal Processing Magazine, 2016. **33**(2): p. 36-44.

[6] Ihsan, A. and E. Mutia. *Wi-Fi and GSM Based Motion Detection in Smart Home Security System*. in *IOP Conference Series: Materials Science and Engineering*. 2019. IOP Publishing.

[7] Ding, B., et al. *Design of Home Intelligent Alarm System Abstract*. in *IOP Conference Series: Materials Science and Engineering*. 2019. IOP Publishing.

[8] Sujadi, H. and A. Bastian. *Design prototype detection tools of Porous Tree using microcontroller Arduino Uno R3 and piezoelectric sensor*. in *Journal of Physics: Conference Series*. 2018. IOP Publishing.

[9] Liang, R. *The Design of GPS Information Display System Based on Arduino UNO R3*. in *2017 7th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2017)*. 2017. Atlantis Press.

[10] Chen, L., J. Zhang, and Y. Wang. *Wireless Car Control System Based on ARDUINO UNO R3*. in *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. 2018. IEEE.

[11] Alathari, B., et al., *A framework implementation of surveillance tracking system based on PIR motion sensors.* Indonesian Journal of Electrical Engineering and Computer Science, 2019. **13**(1): p. 235-242.

[12] Mahzan, N., et al. *Design of an Arduino-based home fire alarm system with GSM module*. in *Journal of Physics: Conference Series*. 2018. IOP Publishing.

[13] Joseph, A., V. Parmar, and V. Bagyaveereswaran. *Design of Human Detection Robot for Natural calamity Rescue Operation*. in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*. 2019. IEEE.

[14] Al Fani, H., et al., Perancangan Alat Monitoring Pendeteksi Suara di Ruangan Bayi RS Vita Insani Berbasis Arduino Menggunakan Buzzer. JURNAL MEDIA INFORMATIKA BUDIDARMA, 2020. 4(1): p. 144-149.

Survey of Query correction for Thai business-oriented information retrieval

Phongsathorn Kittiworapanya
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
60070055@kmitl.ac.th

Nuttapong Saelek
Faculty of Engineering
Khon Kaen University
Khon Kaen, Thailand
Nuttapong_s@kkumail.com

Anuruth Lertpiya, Tawunrat Chalothorn
Kasikorn Labs
Kasikorn Business—Technology Group
Nonthaburi, Thailand
{anuruth.l}{tawunrat.c}@kbtg.tech

***Abstract*—The importance of effective Thai information retrieval (IR) increases as more businesses in Thailand undergo digital transformation. However, previous research on Thai IR systems has mainly focused on web search engines. This study will focus on using query correction to reduce user errors to improve Thai IR. Experiments are conducted on our business-oriented Thai IR task (bTIR). Our investigation presented three notable findings. First, cognitive errors are less of an issue in a business setting. Thus, homophones correction methods provide very little to no benefit for bTIR. Second, approximation based spelling correction methods can significantly reduce search performance. Thus, partial matching on a full dictionary, such as symmetric delete indexing (SymSpell), should be preferred over non-optimal search methods. Third, we introduce a re-ranking algorithm for query corrector, which features multiple sub-correctors (e.g., ThaiQCor 2.0), which results in better performance across multiple configurations.**

***Index Terms*—Natural Language Processing, Text normalization, search engines, information retrieval, text retrieval, Thai language**

I. Introduction

Digital transformation is a vital business strategy in both reducing costs and increasing productivity. Data is often stored in a structured format, enabling computers to reason with the data. However, data within a business (e.g., documents, questions-and-answers, images) is created and used by human operators. Thus, the contents are often either written in or labeled with natural language (e.g., Thai). Thus, providing fast and easy access to these unstructured data is often a challenge for information retrieval (IR) systems.

Indexing techniques in IR systems explored on Thai documents have relied heavily on indexing terms (keywords) to match records. Thus, one of the significant obstacles of IR systems is errors in the search queries [1]. Prior works have incorporated strategies to help mitigate this issue. IR systems adopt a Trie structure for document indexing to perform partial keyword matching with a relatively low-performance penalty [2]. Query correctors have been proposed to correct spelling errors in queries before the search [1], [3]. However, prior works have yet to measure how correction methods affect the end-to-end performance of IR systems.

In this paper, we investigate how current IR systems perform on business-oriented Thai IR (bTIR). Our investigation showed three important findings.

First, cognitive error (user cannot spell the word) in search queries is a non-issue in a business IR. There are two primary types of errors mentioned in the literature on web search engines [3]: typographical errors (typos) and cognitive errors. They are opposed to our investigation in business IR.

Second, for optimal query correction, methods equivalent to full dictionary search should be preferred over approximation-based indexing approaches (e.g., ThaiQCor 2.0 [3]). Since directly computing the edit distance for each search query against every word in the dictionary can be very costly, specialized data structures (SymSpell [4]) have been proposed to reduce computation costs.

Third, we propose a new method to merge multiple lists of candidates produced by multiple query correctors.

This paper is structured as follows. Section II overviews the prior works relating to Thai IR systems. Section III describes our business-oriented Thai IR (bTIR) task, which is used to evaluate each query corrector. Section IV goes over the experiment setup for each query corrector as well as the IR system. Section V discusses the results of our experiments. Lastly, we will reiterate our contributions in Section VI.

II. Related Works

Advancements in Thai IR can be split into two categories: indexing approaches and query correction. A vari-

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

ety of methods have been proposed for document indexing, but research in this area is mainly focused on reducing the computational cost of performing keyword searches [5]–[7]. Whereas, advances in indexing can also result in better search accuracy; for example, the Trie structure can theoretically enable misspelled queries to match appropriate documents via partial matching [2]. However, IR systems that employ index-based partial matching will be limited to text-based errors. Moreover, ThaiQCor has shown that using multiple domains (i.e., text-based for spelling errors and sound-based for cognitive errors) leads to better correction performance [1], [3].

Significant advancements have been made in Commercial IR systems (e.g., web search engines) which employ user data (e.g., search history and online ranking) to better resolve search queries [8] However, we will only focus on the text retrieval task. In text retrieval, IR systems are tasked to return a ranked list of documents when given a text search query [9], [10]. Modern IR systems have shifted towards neural-based methods for document ranking [8], [10]. However, these systems require a massive amount of data to train [11]. Therefore, traditional methods are still often used in open-source search engines[1], which are typically found in enterprise search systems.

III. Business-oriented Thai IR Task

The objective of this research is to improve our current internal IR system. Therefore, we develop the business-oriented Thai IR (bTIR) task to evaluate how each query correction system affects the final search results. In the following sub-sections, we discuss the nature of our data and the evaluation metrics used in this task.

A. Data

Documents and queries data are sampled from our existing IR system. The queries have an average length of 2.5 words, with a minimum of 1 and a maximum of 12. Meanwhile, the documents have an average length of 189.4 words (a minimum of 12 words and a maximum of 2,883 words). Then, the data is annotated with binary relevance labels (587 queries and 203 documents). The annotated queries are then split into three sets: training, validation, and test. The training set and validation set are used for model development while the test set is used for evaluation. This allows us to gauge how the IR systems perform on unseen (new) queries. The size of each dataset split is shown in Table I.

Table I
Query data split

Split	Queries
Train	375
Validation	94
Test	118
All	587

B. Evaluation Metrics

Many metrics have been proposed for evaluating IR systems [8]. Since the data is annotated with binary relevance labels, we chose precision, recall, F-score, mean reciprocal rank (MRR), and top-ten. Precision is the fraction of retrieved relevant documents over retrieved documents. Recall is the fraction of retrieved relevant documents over all relevant documents. F-score is defined in Eq 1. The reciprocal rank is calculated from the first relevant document in the retrieved documents. Top-ten scores each query is correct if relevant documents are within the top ten results.

$$F_\beta = (\beta^2 + 1) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \tag{1}$$

IV. Experiments

This section details our experiment setup for evaluating query correctors on the bTIR task. The query corrector produces corrected candidate queries, which is then fed to a traditional search engine. The overall process is illustrated in Fig 1. The dictionary is built from the preprocessed documents and cleaned manually. The preprocessing stage is detailed in Section IV-A. The query corrector and the search engine are detailed in Section IV-B and Section IV-C, respectively.

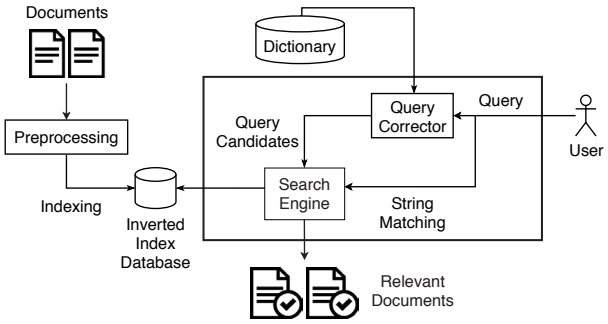


Figure 1. Overview of our proposed search engine

A. Preprocessing

The preprocessing stage is used to clean both the input queries and the documents. First, non-language related characters (i.e., special characters) are stripped. Second, the text is segmented into a sequence of words using a neural-based word tokenizer trained on Thai UGWC [12].

[1]Apache Lucene supports many scoring methods including vector space model (TF-IDF) and Okapi BM25, which are based on bag-of-words, see https://lucene.apache.org/core/8_6_0/core/org/apache/lucene/search/package-summary.html#scoring

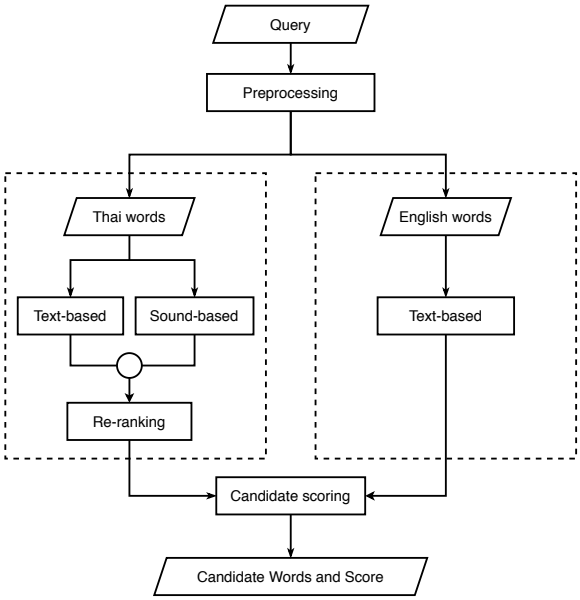


Figure 2. Query Corrector

For the reason that Thai UGWC [12] improves resilience against spelling errors Finally, stop words[2] are removed.

B. Query Corrector

The query corrector can be split into preprocessing, word correction, and query generation. This is to accommodate our bTIR task (detailed in Section III-A), which contains multiple words per query. This framework allows us to evaluate existing query correction methods (e.g., ThaiQCor 2.0 [3]), which presumed that a query contains a single keyword. This paper will refer to these single-keyword query correctors as word correctors. The overview of the query corrector with multiple Thai word corrector is shown in Fig 2. The input query text is preprocessed (see Section IV-A) into a sequence of words. Each word is corrected independently by the word corrector to produce candidate words. The word corrector is detailed later below. The combination of each candidate word is combined to create new queries. Although the number of combinations could grow rather quickly, the average combination of queries produced is 1.14 on the test set (with a minimum of 1 query and a maximum of 4 queries). There are two re-ranking methods for merging the sequence of correction candidates for Thai pipelines with multiple word correctors: ThaiQCor 2.0's and ours (detailed under Re-ranking in this section).

After computing the word candidates of each preprocessed word, each candidate is scored according to Eq 2. Next, the word candidates, along with the scores, are sent to the search engine. Where N is the total number of candidates, and $rank_w$ is the rank of the candidate word w.

$$CandidateScore_{(w)} = \frac{N - rank_w + 1}{N} \tag{2}$$

Word corrector: The word corrector is responsible for producing a list of candidate words for each word in the preprocessed query. The word corrector is utilized two in separate pipelines depending on the language (see Fig 2). The Thai pipeline either only contains a text-based spelling corrector or both text-based and sound-based corrector. On the other hand, the English pipeline only features a text-based spelling corrector is presented.

We primarily evaluated two works on word correctors: SymSpell [4], and ThaiQCor 2.0 [3]. SymSpell a general text-based spelling corrector that can efficiently perform full dictionary searches. Whereas ThaiQCor 2.0 (TQC2) is the most recent publication on Thai query correction. TQC2 proposed two word correctors: a spelling corrector using word approximation and a sound-based corrector.

Four word corrector are experimented in this paper. They are composed of ThaiQCor 2.0 [3], ThaiQCor 2.0 with tone and canceled letters removal [1], ThaiQCor 2.0 with Symspell, and SymSpell. Due to a lack of public availability of the G2P module mentioned in ThaiQCor 2.0 [3], the Thai Language Toolkit's G2P module[3] was chosen as a replacement. The hyperparameters of all methods are tuned for the highest F_2 on the validation set (as listed in Appendix A).

Re-ranking: Since the Thai pipeline may contain multiple word correctors, multiple lists of candidates must be merged before candidate scoring. The word candidates are re-ranked by their re-rank scores (Eq 3) in the descending order. N_{spell} is the total number of candidates produced by the spelling-based corrector, N_{sound} is the total number of candidates produced by the sound-based corrector. $rank_{w,C}$ is the rank of the candidate word w produced by the corrector C. If the corrector C did not produce the candidate word w, then $rank_{w,C} = N_C$

$$RerankScore_{(w)} = \frac{(\frac{N_{spell} - rank_{spell}}{N_{spell}}) + (\frac{N_{sound} - rank_{sound}}{N_{sound}})}{2} \tag{3}$$

C. Search Engine

The search engine looks for the relevant documents D_R by scoring the documents D on a set of query combinations Q. Each query $q \in Q$ and each document $d \in D$ are vectorized into $q^{(emb)} \in \mathbb{R}^v$ and $d^{(emb)} \in \mathbb{R}^v$, using a modified term frequency–inverse document frequency (tf-idf) as defined in Eq 4 and Eq 5. Where v is the size of the vocabulary (dictionary), $f_{w,x}$ is the frequency of the word w in the sequence of words x, $f_{w,D}$ is the frequency of the

[2]Thai stop words from https://github.com/PyThaiNLP/pythainlp/blob/dev/pythainlp/corpus/stopwords_th.txt, English stop words from https://www.nltk.org/nltk_data/.

[3]tltk.nlp.g2p from https://pypi.org/project/tltk/

word w in all documents D, $q_i^{(score)}$ is the candidate score of the word q_i, $Count(x)$ is the word count of sequence x, and $Count(D)$ is the total number of documents. Each document d is scored against the set of query combinations Q as detailed in Eq 6.

$$tfidf(w, x, D) = \frac{f_{w,x}}{Count(x)} \cdot \log(\frac{Count(D)}{f_{w,D}}))$$

$$q_i^{(emb)} = tfidf(q_i, q, D) \cdot q_i^{(score)} \quad (4)$$

$$d_i^{(emb)} = tfidf(d_i, d, D) \quad (5)$$

$$score(d, Q) = \max_{q \in Q}(\frac{q^{(emb)} \cdot d^{(emb)}}{|q^{(emb)}| \cdot |d^{(emb)}|}) \quad (6)$$

A cut-off threshold is used to filter out irrelevant results. The threshold is calculated based on the highest scoring document, as defined in Eq 7. Where σ is a hyper-parameter, and D_R is a set of relevant documents. A retrieved document $d \in D_R$ is ranked higher if the document contains a sub-string that exactly matches the search query. After that, the documents are sorted by their respective $score(d, Q)$ in descending order.

$$threshold = \sigma \cdot \max_{d \in D}(score(d, Q))$$
$$D_R = \{d \in D \wedge score(d, Q) \geq threshold\} \quad (7)$$

V. Results & Discussion

The outline and result discussion are presented in this section. The results are shown in Table II and Table III. Bear in mind that the cut-off threshold does not affect the top-ten evaluation since the cut-off is always the first ten search results. The interesting points from the results are as follows.

First, SymSpell alone outperformed all other query correction methods. Further investigation showed that our task does rarely feature an erroneous query that specifically requires the sound-based correction. Thus, the minimal improvements in recall are outweighed by the reduced precision from the additional correction candidates (comparing SymSpell against TQC2 + T, R + SymSpell).

Second, the re-introduction of tone and canceled letters removal (previously proposed by [1]) showed significant search performance improvement (denoted as T in Table II). However, this might be caused by a lack of exact G2P module, as published in the original paper [3].

Third, changing from ThaiQCor 2.0's word approximation to SymSpell yield another significant jump in search performance (denoted as SymSp in Table II).

Lastly, our proposed re-ranking method improves end-to-end search performance for merging ThaiQCor 2.0's sound-based correction with both ThaiQCor 2.0's word approximation as well as SymSpell (denoted as R in Table II and Table III).

Table II
End-to-end search performance evaluation of various query correction methods on the bTIR task

Approaches	*Prec*	*Recall*	F_1	F_2	MRR
No correction	0.384	0.533	0.388	0.436	0.538
ThaiQCor 2.0	0.312	0.453	0.326	0.375	0.479
ThaiQCor 2.0 + T	0.370	0.522	0.373	0.422	0.558
ThaiQCor 2.0 + T, R	0.409	0.487	0.397	0.428	0.553
TQC2 + T + SymSp	0.421	0.533	0.411	0.451	0.580
TQC2 + T, R + SymSp	0.441	**0.537**	0.431	0.469	0.585
SymSpell	**0.447**	0.531	**0.433**	**0.470**	**0.590**

[T] Tone and canceled letters removal [1].
[R] Our proposed RerankScore from Eq 3. Otherwise, candidates are rank using ThaiQCor 2.0's *Score* [3].

Table III
Top-ten performance evaluation of various query correction methods on the bTIR task

Approaches	Top-10
No correction	0.661
TQC2 + T + SymSp	0.686
TQC2 + T, R + SymSp	0.695
SymSpell	**0.703**

[T] Tone and canceled letters removal [1].
[R] Our proposed RerankScore from Eq 3. Otherwise, candidates are rank using ThaiQCor 2.0's *Score* [3].

VI. Conclusion & Future work

In this paper, we developed the business-oriented Thai Information Retrieval task (bTIR) and evaluated how a multitude of query correctors affect information retrieval performance on bTIR. There are three important findings. First, that sound-based correction provides very little to no benefit on bTIR. Second, a full dictionary search spelling correction is crucial to improving search performance. Third, we proposed a method for merging and re-ranking query candidates for systems with multiple sub-correctors.

In addition, we introduce an IR framework for utilizing and evaluating existing single-keyword query correction methods. We also found that ThaiQCor 2.0 sound-based corrector can be further improved by methods previously proposed by ThaiQCor 1.0 [1] (i.e., tone and canceled letters removal).

This study substituted the ThaiQCor 2.0's word approximation with symmetric delete indexing, implemented by SymSpell [4]. This enables the query corrector to perform a full dictionary search for each correction. However, symmetric delete indexing could potentially be adapted for sound-based correction.

Other than this, we found that about 10% of the queries contains transliterations of proper names. For example, the English words for Coronavirus, Siam Paragon, KPlus are โคโรน่าไวรัส, สยามพารากอน and เคพลัส in Thai, respectively. Since transliteration words have similar pronunciation with their counterparts, transliterate-capable query corrector could also further improve search performance.

Table IV
Consolidated list of hyperparameters of the tuned query correctors

Hyperparameter	Value
SymSpell	
σ	0.55
English SymSpell's max edit distance	2
Thai SymSpell's max edit distance	2
Word candidates limit	5
ThaiQCor 2.0 + T, R	
σ	0.55
English text-based corrector minimun IOU	0.4
English text-based max edit distance	3
Thai sound-based corrector minimun IOU	0.5
Thai sound-based max edit distance	1
Thai text-based corrector minimun IOU	0.4
Thai text-based max edit distance	1
Word candidates limit	5
TQC2 + T, R + SymSp	
σ	0.55
English SymSpell's max edit distance	2
Thai sound-based corrector minimun IOU	0.4
Thai sound-based max edit distance	2
Thai SymSpell's max edit distance	2
Word candidates limit	5

VII. Acknowledgments

This research was supported by Kasikorn Business—Technology Group (KBTG).

Appendix A
Hyperparameters

A consolidated list of hyperparameters of the various query correction models reported in Section V is shown in Table IV.

Appendix B
Re-ranking

This section will compare our proposed method (detailed in Section IV-B) for re-ranking candidates produced from multiple correctors against ThaiQCor 2.0's (TQC2) [3]. We will examine two situations: the correctors produce a balanced number of candidates, and the correctors produce a significantly different number of candidates. The scenarios are shown in Table V and Table V, respectively. In the ranking columns, 1/5 denotes that that corrector proposed that candidate as the first from a total of five candidates. While −/5 denotes that that corrector did not propose the candidate. The scoring columns list the score of each candidate when scored with a particular method. The re-ranking process sorts each candidate by the score in descending order.

Table V details the score of each candidate when both correctors produced five candidates. Both ours and TQC2 prefer candidates highly ranked by both correctors (see rows 1, 2). When a candidate is already highly ranked by one corrector, another corrector's proposal is always preferred by our method. However, TQC2 prefers unranked candidates over lowly ranked candidates (see rows 1, 4).

Table V
Re-ranking example with balanced number of candidates from two query correctors

#	Candidate	Ranking		Scoring	
		Cor1	Cor2	TQC2 [3]	Ours
1	High-rank only in 1 corrector	1/5	-/5	4	0.4
2	High-rank in 2 corrector	1/5	1/5	5	0.7
3	Ranked by 2 correctors	3/5	3/5	3	0.4
4	High-rank in 1 corrector	1/5	4/5	3.5	0.5

Table VI
Re-ranking example with imbalanced number of candidates from two query correctors

#	Candidate	Ranking		Scoring	
		Cor1	Cor2	TQC2 [3]	Ours
1	High-rank only in small	1/3	-/8	2	0.33
2	High-rank only in large	-/3	1/8	7	0.44
3	Ranked by large correctors	-/3	6/8	2	0.13
4	High-rank in 1 corrector	1/3	6/8	3	0.46

Table VI details the candidates' scores when the two correctors proposed a significantly different number of candidates. By comparing row 1 against 2 and 3, TQC2 shows a preference over candidates from correctors that produced more candidates. On the other hand, our re-rank scores are in line with scores from the previous situation (a balanced number of candidates).

References

[1] N. Angkawattanawit, C. Haruechaiyasak, and S. Marukatat. Thai q-cor: Integrating word approximation and soundex for thai query correction. In *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 1, pages 121–124, 2008.

[2] Witoon Kanlayanawat and Somchai Prasitjutrakul. Automatic indexing for thai text with unknown words using trie structure.

[3] S. Thaiprayoon, A. Kongthon, and C. Haruechaiyasak. Thaiqcor 2.0: Thai query correction via soundex and word approximation. In *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pages 113–117, 2018.

[4] wolfgarbe. wolfgarbe/SymSpell. https://github.com/wolfgarbe/SymSpell, Nov 2019.

[5] Todsanai Chumwatana. A survey of automatic indexing techniques for thai text documents. *Information Technology Journal*, 9(1):81–91, 2013.

[6] Choochart Haruechaiyasak, Chaianun Damrongrat, Chatchawal Sangkeettrakarn, Sarawoot Kongyoung, and Niran Angkawattanawit. Sansarn look!: A platform for developing thai-language information retrieval systems. In *Proc. of the 21st International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2006)*, pages 85–88, 2006.

[7] T. Chumwatana, K. W. Wong, and H. Xie. An automatic indexing technique for thai texts using frequent max substring. In *2009 Eighth International Symposium on Natural Language Processing*, pages 67–72, 2009.

[8] Bhaskar Mitra and Nick Craswell. Neural models for information retrieval. *CoRR*, abs/1705.01509, 2017.

[9] Charles L Clarke, Nick Craswell, and Ian Soboroff. Overview of the trec 2009 web track. Technical report, WATERLOO UNIV (ONTARIO), 2009.

[10] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, Daniel Campos, and Ellen M Voorhees. Overview of the trec 2019 deep learning track. *arXiv preprint arXiv:2003.07820*, 2020.

[11] Michael Taylor, Hugo Zaragoza, Nick Craswell, Stephen Robertson, and Chris Burges. Optimisation methods for ranking functions with multiple parameters. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 585–593, 2006.

[12] A. Lertpiya, T. Chaiwachirasak, N. Maharattanamalai, T. Lapjaturapit, T. Chalothorn, N. Tirasaroj, and E. Chuangsuwanich. A preliminary study on fundamental Thai NLP tasks for User-generated Web Content. In *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pages 1–8, 2018.

Utilization-Weighted Algorithm for Spreading Factor Assignment in LoRaWAN

Kasama Kamonkusonman
Electronic and Telecommunication Engineering
King Mongkut's University of Technology Thonburi
Bangkok, Thailand.
kasama.kamon@mail.kmutt.ac.th

Rardchawadee Silapunt
Electronic and Telecommunication Engineering
King Mongkut's University of Technology Thonburi
Bangkok, Thailand.
rardchawadee.sil@mail.kmutt.ac.th

***Abstract*— Long Range Wide Area Network (LoRaWAN) is one of the leading low power wireless networks that can support thousands of Internet of Things (IoT) devices. To enhance the scalability of LoRaWAN, this paper proposes the Utilization-Weighted (UW) algorithm, which is the spreading factor management algorithm designed based on the M/D/1 queue theory. The main concept of this algorithm is channel utilization balancing that helps form groups of nodes assigned with different spreading factors (SFs). The simulations are performed under two scenarios that are similar and various uplink time interval among SFs. The results show that our UW algorithm can outperform the traditional Min-airtime method in both scenarios. The packet received rate (PRR) of the UW algorithm is clearly higher than that of the Min-airtime method for all number of nodes and time intervals. Especially in the various time interval simulation of the networks of 120, 600, and 1,200 nodes, the maximum PRR improvements occur at 1, 3, and 5 times of the minimum time interval between uplinks, $T_{off,i}$, respectively, and are around 34%, 36%, and 35%, respectively.**

Keywords*— *Internet of Things, LoRaWAN, scalability, spreading factor assignment, time interval

I. INTRODUCTION

LoRaWAN is one of the leading Low Power Wide Area Network (LPWAN) technologies. The characteristics of LoRaWAN are suitable for Internet of Things (IoT) deployment due to low power consumption, wide area coverage, and low cost [1]. According to the LoRaWAN specification [2], LoRaWAN is a media access control (MAC) protocol designed on top of LoRa modulation physical layer which is an open standard to connect LoRa low-power devices to the internet. The architecture of LoRaWAN consists of 4 main components: nodes, gateway, network server, and application server. The nodes connect to the LoRaWAN gateway in the star of stars topology. For uplink, the gateway is responsible for translating LoRa modulation packets to internet protocol (IP) and forwarding the received packets from the nodes to the servers. On the hand, the received downlink IP packets are also translated to LoRa protocol and forwarded to the nodes by the gateway.

For LoRaWAN, the nodes can choose to use one of three standard classes, which consists of class A, class B and class C, depending on their applications. However, most of previous works focus on class A due to its lowest power consumption. The class A nodes autonomously select time to send uplink data, whereas downlink can be sent to the nodes only in one of two windows that are opened after the uplink. The nodes will then sleep for the rest of the time. As same as the other classes, the standard communication protocol of LoRaWAN class A uplink is the Additive Links On-line Hawaii Area (ALOHA), which is a random access protocol. In ALOHA, nodes send packets without any collision control mechanism such as the channel sensing. Especially in the single gateway-single channel scenario, the collisions can occur dramatically [1], thus limiting the scalability of the network.

To support thousands of IoT devices, scalability of LoRaWAN is one of the key issues. Several methods have been used to improve scalability such as implementing other proposed protocols such as slotted-aloha [3], Carrier-Sense Multiple Access (CSMA) [4-5], adding multiple gateways [1], spreading factor (SF) assignment and allocation [6-14], etc. So far, the SF assignment and allocation has been the most popular technique to improve LoRaWAN scalability. Previous studies were focused on how to manage SFs proportion in a single physical channel gateway whose number of virtual channels are, by the LoRa modulation principle, equal to the number of SFs. Note that, packets sent with different SFs are theoretically orthogonal to each other.

Different SFs have different capacity of data transmission. For example, a lower SF device can send data at higher rate. Also, due to the duty cycle limitation, the lower SF device can send higher number of data packets. The number of sent packets also depends on the inter-arrival time, which is time interval between packets. Therefore, the variation in the inter-arrival time can have an impact on the collision of packets especially in the network with large number of devices. Unfortunately, most of prior research does not consider the impact of time interval between packets among SFs [6-14].

This paper proposes the SF assignment algorithm for LoRaWAN class A based on the Utilization-Weighted (UW) concept of channels. More specifically, we interest in the time interval between uplinks that has an impact on channel utilization and thus SF assignment. Two scenarios of interest include i) same uplink time interval among SFs and ii) various uplink time intervals among SFs. To evaluate the performance of our algorithm, packet received rate (PRR) or the ratio of successful received packets over sent packets, will be compared with the Min-airtime method, the default algorithm of LoRaWAN that assigns SF based on minimum air time, which is also used in [1].

This paper has been supported by Petch Pra Jom Klao Master's Degree Research Scholarship from King Mongkut's University of Technology Thonburi. (Grant No. 20-2560)

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

II. SPREADING FACTOR ASSIGNMENT TECHNIQUES

For LoRaWAN protocol, the transmission between node and gateway consists of physical layer parameters, MAC layer commands, and security. The LoRa physical layer uses chirp spread spectrum (CSS) modulation scheme which has 5 important parameters: SF, bandwidth (BW), coding rate (CR), transmission power, and center frequency. Ideally, transmissions with different chirp rates, which is defined as (1), are always orthogonal to each other,

$$Chirp\ Rate = \frac{BW^2}{2^{SF}}. \quad (1)$$

In this sense, assigning different SFs or BWs can lead to orthogonality between packets. However, the receiver has different sensitivity values to receive packet with different SFs, as example shown in Table I for SX1301 chip. Thus, to ensure successful packets transmissions to the gateway, the common technique is assigning the lowest possible SF that has lower sensitivity value comparing to Received Signal Strength Indicator (RSSI) at the receiver. This technique is so called the Min-airtime since the lower SF would have lower airtime as calculated from (2),

$$T_s = \frac{2^{SF}}{BW}, \quad (2)$$

where T_s is symbol time (second).

However, as shown in [1], the Min-airtime technique is still insufficient to support connection of thousands of IoT nodes. Thus, previous studies further proposed enhanced SF allocation techniques.

F. Cuomo *et al.* [6] suggested to assign SF assignment by simply equalizing number of nodes in each SF group, or so called EXPLoRA-SF. Even though EXPLoRA-SF can provide higher PRR than Min-airtime technique, this method is still not suitable for a highly loaded network since the higher SF packets tend to have higher probability for collision.

The airtime equalization was then pursued [6-8]. In these studies, the concept is to equalize node airtimes in each SF group in a single channel gateway to improve PRR since the technique helps reducing channel overload. The difference is that in [6], the water filling technique, which is more sophisticated equalization algorithm, was employed. The work in [6] was further extended by including channel capture effects and inter-SF interference to help balancing the number of nodes among SFs [8]. For the imperfect orthogonality, [8] suggested that less proportion of SF11 and SF12 nodes can help increasing PRR.

The SF assignment by fixing the area of each SF group was also reported [9]. Although the method can give a better performance than the network with randomly positioned nodes, the location of nodes, in practice, might not be able to fit in such optimal area.

In addition to the SFs proportion adjustment, other performance improvement techniques including tuning node parameters such as transmission power [10-11] and BW and CR [12], and carrier frequency management [13-14] were also proposed. Note that, in all techniques mentioned above, the time interval between packets is kept the same to all nodes [6-14].

TABLE I. SX1301 SENSTIVITIES VALUES [16]

SF	*7*	*8*	*9*	*10*	*11*	*12*
Sensitivity for 125 KHz bandwidth (dB)	-126.5	-129.0	-131.5	-134.0	-136.5	-139.5

III. UTILIZATION-WEIGHTED SF ASSIGNMENT ALGORITHM

The utilization-weighted (UW) algorithm is designed based on the queuing theory. According to [15], the LoRaWAN class A device with a single gateway exhibits operating characteristic of the M/D/1 queuing system where arrival packets are sent with the Poisson arrival rate λ and serving with the fixed service rate, μ. Since the nodes with different SF assignments can send data with different data rates and are ideally orthogonal to each other, we will virtually have six independent channels for SF7 to SF12. The service rate, μ_i, of each SF is defined as the duty cycle over the transmission time as

$$\mu_i = \frac{D_j}{T_i}, \quad (3)$$

where D_j is duty cycle of the physical channel (percent), T_i is total transmission time of a packet sent by a node with SF *i* (second), *i* is a value between 1 and 6 for SF7 to SF12 respectively, and *j* is value between 1 and c for c channels.

The arrival rate λ_i for a node can be defined as

$$\lambda_i = \frac{1}{P_i}, \quad (4)$$

where P_i is time interval between two uplink packet generations (second).

A link utilization of a node, ρ_i, or busy ratio of the channel for each SF is defined as

$$\rho_i = \frac{\lambda_i}{\mu_i}. \quad (5)$$

For multiple channels, the imbalance utilization might cause some channels to be overloaded while others are underloaded. The overloaded channel is prone to more packet collisions. To reduce packet collisions, we propose to equalize the utilizations of the virtual channels by

$$n_i\rho_i = n_{i+1}\rho_{i+1} = n_{i+2}\rho_{i+2} = \ldots = n_{i+5}\rho_{i+5}, \quad (6)$$

where n_i is the number of nodes holding the SF in group *i*.

This means that the channel with lower link utilization of a node can serve more nodes. Thus, the weight for number of nodes of each SF is defined as the inverse of the utilization as

$$w_i = \frac{1}{\rho_i}. \quad (7)$$

The number of nodes for each SF can then be calculated by

$$n_i = \frac{w_i}{\left(\sum_{i=1}^{6} w_i\right)} \times N \,, \tag{8}$$

where N is the total number of assigned nodes.

Algorithm 1 UW

Input: N: Total number of nodes, **RSSI:** corresponding Received Signal Strength Indicator of each node, n_i **:** optimal number of nodes for each SF from (8)

Output: $NODE_{SF}$ **:** Set of optimal spreading factors of all nodes

Spreading Factor Assignment

1. $Gateway_{SENSI} \leftarrow$ Sensitivity values of gateway
2. **for** 1 $\leftarrow$ 1 to N **do**
3. **while** $NODE_{RSSI}[\mathrm{i}] < Gateway_{SENSI}[SF]$ and $NODE_{SF}[\mathrm{i}] < 12$ **do**
4. $NODE_{SF}[\mathrm{i}] \leftarrow SF + 1$
5. $SF \leftarrow SF + 1$
6. **end while**
7. **while** $No_{SF}[SF] > n_i[SF]$ and $NODE_{SF}[\mathrm{i}] < 12$ **do**
8. $NODE_{SF}[\mathrm{i}] \leftarrow NODE_{SF}[\mathrm{i}] + 1$
9. $No_{SF}[SF] \leftarrow No_{SF}[SF] + 1$
10. **end while**
11. **end for**

The pseudo code of our UW algorithm is described as Algorithm 1. First, N is defined as the total number of nodes to be implemented in a network. Then, we set a [1 × N] $NODE_{RSSI}$ vector to collect corresponding RSSI of each node, a [1 × N] $NODE_{SF}$ vector to store optimal SF for each node, a [1 × M] No_{SF} vector to store the number of nodes for each SF, and a [1 × M] $Gateway_{SENSI}$ vector to contain gateway sensitivity values for each SF. For our algorithm, M is equal to 6 since SF can be ranged from SF7 to SF12. When the node sends uplink messages to gateway, the signal strength will naturally be attenuated along the path. To ensure no packet loss from path attenuation, the RSSI of the packets must be greater than the sensitivity of the receiver. In the case of LoRaWAN, the higher SF has lower sensitivity value as shown in Table I. For an immobile node, each node ideally has the same RSSI for all uplink messages. Thus, when the node is connected to the network, the system would first check $NODE_{RSSI}$ of each node and compare to the sensitivity of the gateway. Then, the lowest possible SF that has the sensitivity greater than RSSI is first assigned to the node as same as the Min-airtime technique [1]. However, the SF assignment based only on the sensitivity might not give the optimal number of nodes for each SF group or No_{SF} to resolve collision issue. Our system will further check and assign higher SF to the node. This process aims to optimize the No_{SF} as the n_i calculated from (8). Note that, the node can be reassigned with higher SF only.

IV. Network Simulation and Results

In this work, LoRaSim [1], which is a LoRa simulator based on Simply, is employed to simulate the LoRaWAN class A network. For simplicity, only uplink traffics are considered. The simulation parameters are defined as shown in Table II. The nodes are generated randomly by the default Min-airtime method. They are distributed around the single gateway that is located at coordinate (0,0) with radius of 8 km. We model the gateway to mimic the characteristics of the SX1301 chip [16]. The simulations of the network utilizing the Min-airtime method and the proposed UW algorithm are performed under two main scenarios that are same uplink time interval among SFs and various uplink time intervals among SFs. The network performances for both algorithms are evaluated and compared in terms of PRR.

TABLE II. Simulation Parameters

Parameter	Value
Carrier frequency (MHz)	923.2
Bandwidth (kHz)	125
Code Rate (CR)	4/5
Preamble length (symbol)	8
Packet length (byte)	10
Time interval between uplinks (second)	Exponential distribution with mean of 150 s
Number of gateways	1
Captuer effect	No
Number of nodes	120 – 6,000 nodes
Pathloss model	Hata suburban model [17] with Transmitting power = 14 dBm, Transmitter antenna gain = 3 dBi, Receiver antenna gain = 6 dBi, Transmitter antenna height = 1.5 m., Receiver antenna height = 10 m.

A. Same Uplink Time Interval among SFs

For the first scenario, we simulate all uplink traffics with the same time interval between two uplink packets for each SF (SF7 – SF12) on one physical channel. The PRRs are measured as a function of number of nodes, from 120 to 6,000 nodes. The SF distributions of 1,200 nodes using the Min-airtime method and the proposed UW algorithm are shown in Fig. 1 and Fig. 2, respectively. From our choice of the Hata suburban model [17] and assigned node power, SF7 – SF12 nodes can be assigned within the radius of 6.87 km, 7.84 km, 8.94 km, 10.19 km, 11.63 km, and 13.62 km respectively, as shown in Fig. 3.

Thus, within 8 km around the gateway, the Min-airtime method assigns only SF7, SF8, and SF9. We also notice that most nodes are assigned with SF7. These too many SF7 packets will overload the SF7 virtual channel while SF10 – SF12 virtual channels are idle. The SF7 virtual channel is thus the most congested channel, which can lead to low PRR. On the other hand, our algorithm distributes SF7 – SF12 nodes

such that all virtual channels can possess the same utilization ratio by using (8), which means all the virtual channels are exploited equally. As shown in Fig. 4, the PRRs determined by both algorithms of the 10 bytes packet and 150 seconds time interval decrease when the number of nodes increases. The PRR of the proposed algorithm is overall higher than that of the Min-airtime method. The maximum improvement is around 9.2% at 3,600 nodes, as shown in Table III.

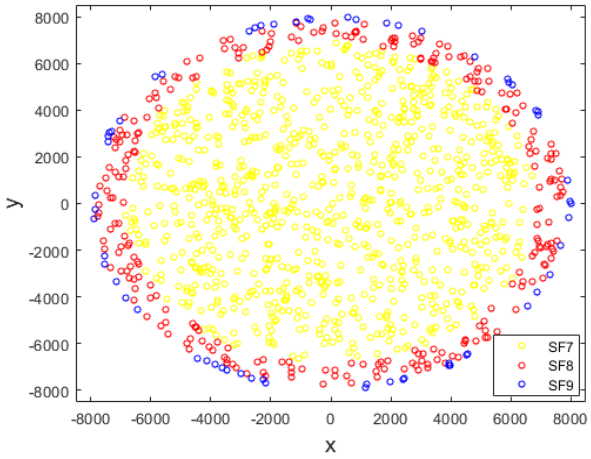


Fig. 1. SF assignment for the Min-airtime method.

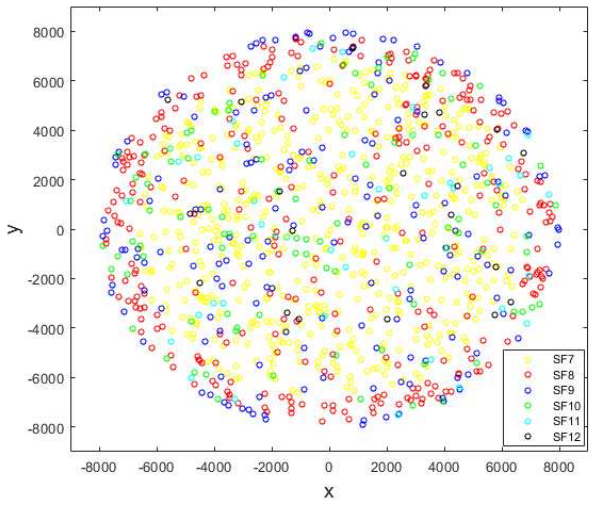


Fig. 2. SF assignment for the UW algorithm.

TABLE III. PRR IMPROVEMENT BY THE UW ALGORITHM

PRR Improvement (%)	Number of Nodes					
	120	*600*	*1,200*	*2,400*	*3,600*	*6,000*
	1.36	4.78	6.83	8.53	9.20	7.99

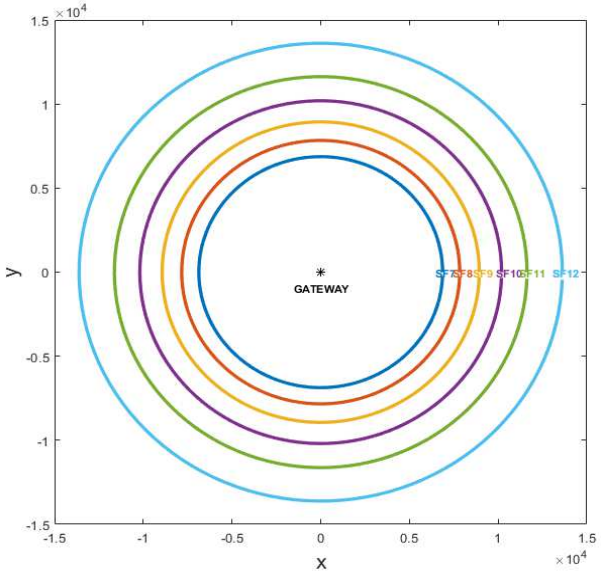


Fig. 3. Assigned area for SF7 – SF12 nodes.

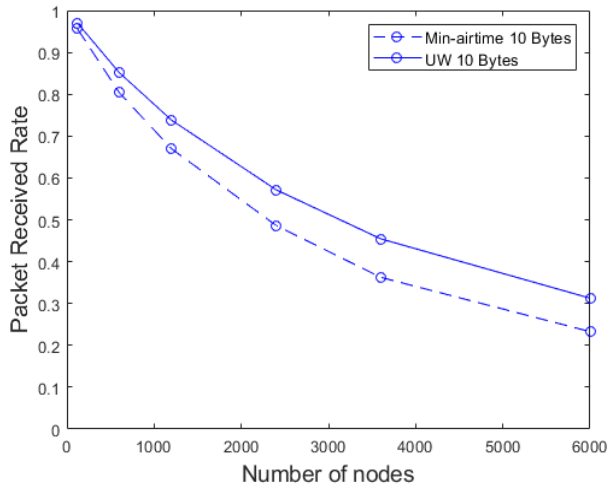


Fig. 4. PRR of the 10 bytes packet with the 150 seconds time interval among SFs.

B. Various Uplink Time Intervals among SFs

For this second scenario, we consider the concept of duty cycle of the LoRaWAN network that limits the time spent on a channel. For example, 1% of duty cycle means that the node can only send data for a total of 36 seconds in 1 hour. In this sense, each SF virtual channel has different capacity to send packets due to differences of data rates. With the same length of payload, SF7 packets can be sent with the fastest rate, so we can send higher number of packets with least time interval between uplinks. On the other hand, SF12 packets can be sent at the slowest rate, so we can send lower number of packets, and with the longest time interval between uplinks.

From the mentioned idea, we simulate uplink traffics with similar settings to the first scenario, except the time intervals between uplinks that are varied differently for each SF. Number of nodes are varied at 120, 600, and 1,200 nodes. The minimum time interval between uplinks, $T_{off,i}$, is used as the

reference time for each SF. The $T_{off,i}$ for each SF is calculated from

$$T_{off,i} = \frac{T_i}{D_j}. \quad (9)$$

For the first sub-experiment, we use the reference time $T_{off,i}$ for each SF and measure the total PRRs for both algorithms. Then, we vary the time interval between uplinks up to 5 times of $T_{off,i}$ for each SF. For the Min-airtime method, SF7 is still mostly assigned to the nodes, similar to the first scenario. Thus, the SF7 virtual channel is still the most congested channel. On the other hand, when the imbalance is found, our UW algorithm will suggest moving packets to other virtual channels that have lower utilization or higher SF channel. Although changing to higher SF may lead to longer time interval between packets, the total PRR increases significantly. As shown in Fig. 5, the PRR increases with the time interval and decreases with the number of nodes, for both algorithms. The performance of our UW algorithm is, however, significantly higher than the Min-airtime method for all time intervals and number of nodes. For the networks with 120, 600, and 1200 nodes, the maximum PRR improvements over the Min-airtime method occur at the reference time intervals of 1, 3, and 5 x $T_{off,i}$ respectively, and are about 34%, 36%, and 35%, respectively.

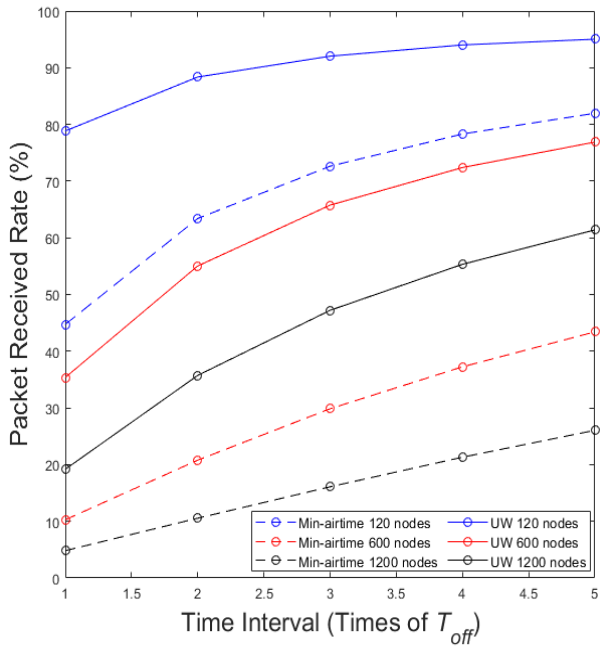


Fig. 5. PRR of the 10 bytes packet with various time interval among SFs.

V. Conclusions

In this paper, we propose the UW algorithm for SF management of LoRaWAN based on fair utilization of 6 virtual M/D/1 channels. Our concept is to equalize packet transmissions on each virtual channel (SF7 – SF12) to reduce packet collision on a channel with higher load. The results from simulation show that our UW algorithm has better performance when comparing to the Min-airtime method for both scenarios. Particularly, the PRR of our algorithm is improved up to 36% over the Min-airtime method when using the various uplink time intervals among SFs scenario. Future works include consideration of imperfect orthogonality of SFs, multichannel, presence of downlink and other parameter adjustments.

References

[1] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso, "Do LoRa low-power wide-area networks scale?," in Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2016, pp. 59-67.

[2] L. Alliance, "LoRaWAN 1.0. 3 specification," lora-alliance. org, vol. 1, pp. 2018-07, 2018.

[3] T. Polonelli, D. Brunelli, A. Marzocchi, and L. Benini, "Slotted aloha on lorawan-design, analysis, and deployment," Sensors, vol. 19, no. 4, p. 838, 2019.

[4] T.-H. To and A. Duda, "Simulation of lora in ns-3: Improving lora performance with csma," in 2018 IEEE International Conference on Communications (ICC), 2018: IEEE, pp. 1-7.

[5] N. Kouvelas, V. Rao, and R. R. V. Prasad, "Employing p-CSMA on aLoRa network simulator," CoRR, vol. abs/1805.12263, 2018.

[6] F. Cuomo, M. Campo, A. Caponi, G. Bianchi, G. Rossini, and P. Pisani, "EXPLoRa: Extending the performance of LoRa by suitable spreading factor allocations," in 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2017: IEEE, pp. 1-8.

[7] A. Tiurlikova, N. Stepanov, and K. Mikhaylov, "Method of assigning spreading factor to improve the scalability of the LoRaWAN wide area network," in 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2018: IEEE, pp. 1-4.

[8] G. Bianchi and F. Cuomo and D. Garlisi and I. Tinnirello "Capture Aware Sequential Waterfilling for LoraWAN Adaptive Data Rate",arXiv:1907.12360 [cs.NI], Nov 2019.

[9] J.-T. Lim and Y. Han, "Spreading factor allocation for massive connectivity in LoRa systems," IEEE Communications Letters, vol. 22, no. 4, pp. 800-803, 2018.

[10] K. Q. Abdelfadeel, V. Cionca, and D. Pesch, "Fair adaptive data rate allocation and power control in LoRaWAN," in 2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM), 2018: IEEE, pp. 14-15.

[11] M. Slabicki, G. Premsankar, and M. Di Francesco, "Adaptive configuration of LoRa networks for dense IoT deployments," in NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium, 2018: IEEE, pp. 1-9.

[12] H. Xie, X. Yuan, Z. Jia, and Z. Wang, "Spreading Factor Allocation for Large-Scale Deployment In LoRaWAN Network," in 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020: IEEE, pp. 880-885.

[13] E. Sallum, N. Pereira, M. Alves, and M. Santos, "Improving Quality-of-Service in LoRa Low-Power Wide-Area Networks through Optimized Radio Resource Management," Journal of Sensor and Actuator Networks, vol. 9, no. 1, p. 10, 2020.

[14] Q. Zhou, J. Xing, L. Hou, R. Xu, and K. Zheng, "A Novel Rate and Channel Control Scheme Based on Data Extraction Rate for LoRa Networks," in 2019 IEEE Wireless Communications and Networking Conference (WCNC), 2019: IEEE, pp. 1-6.

[15] R. B. Sørensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of latency and MAC-layer performance for class A LoRaWAN," IEEE Wireless Communications Letters, vol. 6, no. 5, pp. 566-569, 2017.

[16] "Semtech. SX1301 Datasheet" . 2017. Accessed on: Mar. 30, 2020. [Online]. https://semtech.my.salesforce.com/sfc/p/#E0000000JelG/a/44000000MDnR/Et1KWLCuNDI6MDagfSPAvqqp.Y869Flgs1LleWyfjDY

[17] M. Hata, "Empirical formula for propagation loss in land mobile radio services," IEEE transactions on Vehicular Technology, vol. 29, no. 3, pp. 317-325, 1980.

Development a home electrical equipment control device via voice commands for elderly assistance

Narumol Chumuang
Dep. of Digital Media Technology,
Muban Chombueng Rajabhat University,
Ratchaburi, Thailand
lecho20@hotmail.com

Mahasak Ketcham
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
mahasak.k@it.kmutnb.ac.th

Sakchai Tangwannawit
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
sakchai.t@it.kmutnb.ac.th

Worawut Yimyam
Dep. of Business Information Management, Phetchaburi Rajabhat University
Phetchaburi, Thailand
worawut_yimyam@hotmail.com

Sansanee Hiranchan
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
s6207021910051@kmutnb.ac.th

Montean Rattanasiriwongwut
Dep. of Management Information System, King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
montean.r@it.kmutnb.ac.th

Patiyuth Pramkeaw
Media Technology Programe,King Mongkut's University of Technology Thonburi,
Bangkok, Thailand
patiyuth.pra@kmutt.ac.th

***Abstract*— The paper develops a home electrical equipment control device via voice commands for elderly assistance. This study offers a processing Natural Language Processing (NLP) to extract the text from human voice command in case of the turn on and off home's light switch. First, to analyze and design the structure of sound from the mobile phone which has the elderly support system installed then compare the words with learned database. Secondly, to measure the accuracy of command voice with the confusion matrix model. Finally, to verify order voice between users and systems for matching controlling communication. The experiment result shows that 30 messages extracted voice command from sample group age between 55-60 years old accuracy 85 per cent and able to develop an application using voice commands for old people.**

Keywords— Automatic speech recognition, Natural language processing, Elderly

IV. INTRODUCTION

The population structure of Thailand has entered "Aging Society" since 2005, with a proportion of the population aged 60 years and over more than 10 per cent. It is expected that by 2021, Thailand will have up to 20% of the elderly and 2031 will have 28% of the elderly, leading to a high-level society.[1]

Table I. A number and proportion of the elderly population Classified by age group [8]

year	2020		2030		2040	
	number	Percent	number	Percent	number	Percent
Total Elderly	12,621	100	17,578	100	20,519	100
Age 60 - 69	7,255	57.5	9,260	52.7	8,958	43.7
Age 70 - 79	3,676	29.1	5,897	33.6	7,639	37.2
Age over 80	1,689	13.4	2,420	13.8	7,639	19.1

In 2018, Thailand will have one in five elderly population comparing children according to the Foundation for Research and Development of the Elderly Found that Thailand will enter the aging society completely in the year 2018 [2] In addition to the rankings of Asian countries with a rapidly increasing proportion of the elderly population today. Thailand is ranked second only to Singapore, so having a fast-growing proportion of the elderly population shows that Thailand is necessary prepared to support the aging society because the elderly are considered a high-risk group for various diseases.

The Ministry of Public Health reported that Thai society will have more elderly people in the future and may take result in the elderly burden rate increasing which is linked to illness. However, if divided the elderly according to their potential found that 78% self-care, 20 % family care needed and 2% Bedridden. Therefore, the important process to help Thai elderly has a better quality of life requires a participates from the community to help each other by trying to push the elderly group at home move into the social group [3] and to help the elderly help themselves or solve the problem of being alone that causes depression Therefore bringing more technology to get in contact with the elderly.

Due to the current popularity of voice control technology. [4] There are many electrical devices created for human convenience. The advancement of the Internet of

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

Things (IoT) is an environment that consists communicate and connect via the wired and wireless Communication Protocol for various things. [5] There is a way to identify, recognize the context of the environment and to interact, interact and work together. The communication capabilities of this thing will lead to many new innovations and services such as voice commands on/off the light switches in different rooms without human physical control.

The purpose of this study is to determine developing a system to assist the elderly by using voice commands. Which is a branch of artificial intelligence Which can use Thai And English in conversation To request usage in the system Therefore, there is an idea to apply human speech to adapt with IoT to control various electrical appliances. Inside the house through sounds That will be a tool to help facilitate the use of electrical appliances To help facilitate the elderly who are at home alone Or without someone to take care of, such as voice commands on / off the light switch Order voice calls in an emergency Order to open different websites, open music or watch videos In order to solve loneliness and relieve stress, such as YouTube Application, etc., and can use voice commands to activate the applications installed on the mobile phones of the elderly.

V. LITERATURE SURVEY

Speech Recognition is a computer program system that can convert audio files into text and humans able to speak into the microphone or other devices. The program understands accuracy word depend on the size of vocabulary groups, volume and pronunciation characteristics of the speaker. The system will listen to the sound and decide what sound is heard.[6]

The important technology in ASR is called the Hidden Markov Model (HMM). This technology is able to understand words. From the classification of differences and estimates of the possibility of the components of the unit that is the basis of adjacent sounds based on the principle that each sound has different signal boundaries and unique characteristics.

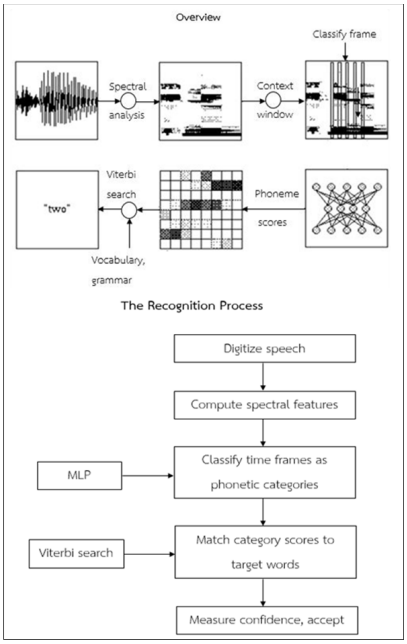


Fig. 1 Speech Recognition [9]

Initial, speech recognition will transform voice wave to number as a requirement and calculate into Spectral symbol. This domain will calculate in each 10 msec and in each part called "Frame" and used to classify number into phonetic-based categories for deploy to Multi-Layer Perceptron which has a function for calculating the output of nodes from the previous layer in hidden layer. The Activation Function is hiding floor has an important function which tries to convert the information into that layer to be able to distinguish it using a single straight line (Linearly Separable) and before the data will be sent to the Output Layer sometimes More than one layer is needed to transform data into Linearly Separable form and use the Viterbi search algorithm to match the neural network output scores with the desired vocabulary, as shown in Figure 2 Viterbi Search uses The continuation of the base unit of sound from the highest possibility calculated through the Gaussian Mixture Model. The advantage of Viterbi Search is data processing in Real-Time. The disadvantage is that the system chooses to Path Prob too low and unable to take into consideration again, although sometimes the Path may have a higher probability that the Path when search the last word in the end of the sentence.

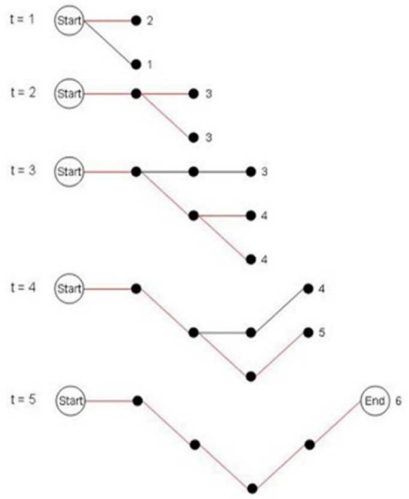


Fig. 2 Viterbi Search

Responding to computers through the use of a mouse and keyboard able to understand the operating instructions of the computer system getting human access modern technology for many years. However, The Automatic Speech Recognition (ASR) has change technological barrier by making computer operations become easier for all users level even unskilled. New technology performance support user is able to access or control computer system through speech command.

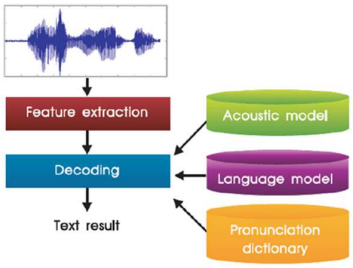


Fig. 3 Automatic Speech Recognition (ASR) Structure System [10]

Speech recognition technology is a system that helps convert speech into text that the computer system can use for further processing This technology was developed within the NECTEC research unit. To help reduce restrictions on access to modern technology By using the natural language processing system to connect users to various types of computer services In the speech recognition process The system will create a Word Network based on the Language Model. Each word consists of the Acoustic Model Hidden Markov Model (HMM) of each sound unit that Assembled into words After that, the system will enter the vector of Speech Feature into the word network. In order to calculate the probability of the incoming sound that will match which path in the word network and will choose the best answer as the route with the highest probability With all steps being decoding

D. Sound Analysis

Sound Analysis is an extraction sound input for additional speech recognition information. Normally, sound passes through our ears (Vibration of the eardrum) and convert into electrical signals sent to a brain. Although, limited of human hearing competency affected some part of the sound and only necessary sound received to the brain that is the same process of sound analysis will select a spot of demand sound wave and convert to text later.

E. Natural Language Processing [12]

Natural Language Processing (NLP) define as a field of Artificial Intelligence which making computer understand human language in purpose communication and analysis data as a text. The reason of NLP implement come from human needed computer able to transmitting with human, supporting reading ability and listening ability. Overview of NLP process summary are :

- Morphological Level : remove words into letters, consonants, vowels or spelling
- Lexical Level : understand each word, mix letters and find the meaning
- Syntactic Level : advert remembered word and structure level from standard rule or expertise
- Semantic Level : understand the context and meaning of words in structure or unstructured sentence standard.
- Discourse Level : understand the connection of sentences , impact of previous and current meaning sentences in difference condition.
- Pragmatic Level : understand the meaning of words and sentences based on situations or prior knowledge. Which may not be included in the content So that it can be interpreted as close to a human being that can connect new information with prior knowledge at all times

F. Decision Tree [31]

Decision trees is techniques providing results in the form of tree structures, which is considered one of the most commonly used learning methods in machine learning. Classifying data into various classes using the attributes of the data to classify decision-making trees obtained from learning, to know which properties of the data determine the classification and each of the characteristics of the data. How important it is, how different it is that it is helpful for users to analyze data and make more accurate decisions

Decision tree method include :

- internal node : The various properties of data are used to decide where the data will go, with the internal node being the initial node called the root node.
- branch, link : The property value or condition of the property on the node used to classify the data which the internal node will branch into equal amount of the property value of that internal
- leaf node : The other result in data classification

At present, there is development of the algorithm for creating decision trees. Which mainly comes from one basic method, which is a top-down greedy search called ID3 , which begins by selecting the best property and creating the root node. The root node, according to the root node's properties, will then have the best properties of the separated data to continue to create the child nodes of that root node

and continue to create child nodes and sub-trees of each branch indefinitely. Until the divided data is grouped together or the amount of divided data in a given branch is less than the specified value.

The construction of a C4.5 decision tree is similar to the ID3 algorithm but with further development, i.e. the use of standard values in deciding which properties to use as the root or node in the tree. Has the highest gain as the root or node. This gain is calculated using knowledge from information theory, that is, information values of data depend on the probability of data which can be measured in the form of bits according to the formula.

$$I(M) = \sum_i^n - P(mi)\, log2\ \Pi(\mu\iota) \qquad (1)$$

If set of M consist possible value {m1 , m2 ,…,mn} as possible value mi equal P(mi)

VI. ANALYZE AND DESIGN THE STRUCTURE OF THE SYSTEM

In terms of designing the structure of the system For the development of assistance systems for the elderly using voice commands Which will determine the work order of this special problem The working process will pick up the sound from the mobile phone equipped with the assistance system for the elderly. After that, separate the words Then compare the words with the database that is used to learn the information After that, send the result to Arduino, receive the charge to the electrical device and send the command to open the application. The study has designed the overall system operation as shown in figure 4.

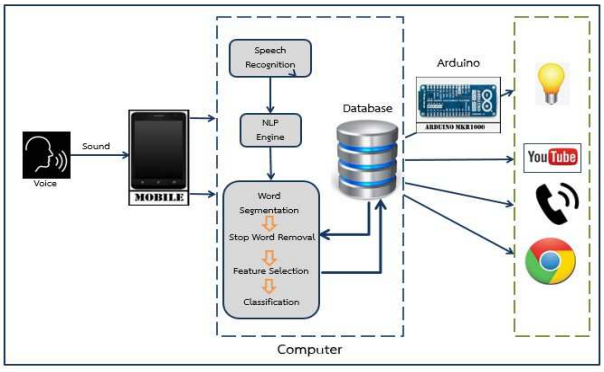


Fig. 4 overall system operation design

System development and device operation For the development of assistance systems for the elderly by using voice commands Show picture results as follows.

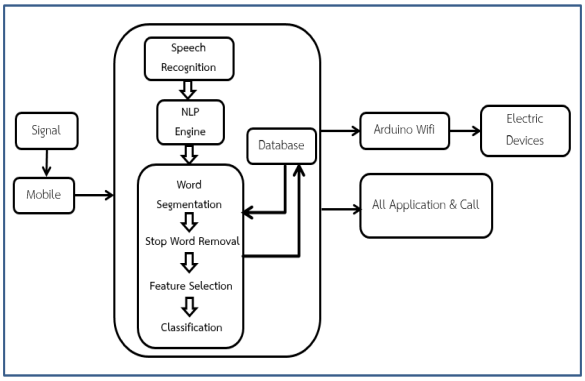


Fig. 5 The step operation of the voice command system to turn on-off the light

STEP 1 Voice command "turn on – off the light" on mobile application where is voice transferred to signal.

STEP 2 When system received signal will recoding input into speech recognition process then word segmentation source based on database convert it to text. After word removal stopped, started feature selection and classification specific demanded word then send to database.

STEP 3 Information will communicate ip to Arduino board via wifi with application in same time, then sending command to electric device working follow user's voice.

VII. FEATURE SELECTION

Selecting features for reducing data size But must lose the least important characteristics of the data and the accuracy of the results to a minimum By considering the weight of the key word By choosing the technique of TF-IDF (Term Frequency-Inverse Document Frequency) by replacing documents and search data And counting the frequency of words that appear in documents (TF) and the inverse frequency with all documents (IDF) which can calculate the inverse frequency of all documents as follows

$$TF - IDF = TF \times log(n/DF) \qquad (3)$$

TF : A weight of number of words that appear in the document

N : All sentences

DF : All documents word appearance

D. System Validation

Model simulation describes the model's performance using Confusion Matrix as a tool for explanation. And calculate the prediction results of the model And compare the results The definition of the Confusion Matrix is explained as follows.

- True Positive (TP) is the prediction correct result according to the expected value (Actual).

- True Negative (TN) is the predictable incorrect results according to the expected value (Actual).

- False Positive (FP) is the predictions incorrect according to the expected value of fallback intent negative.

- False Negative (FN) is the prediction incorrect results but the prediction of that is not the case. This is Actual answers drawn from fallback intent to answer with the correct intent and then test using the generated data set. After that, the above results are calculated for the accuracy as formula below :

$$\text{Correct Mapping (CM)} = TP + TN \quad (4)$$

$$\text{Incorrect Mapping (IM)} = FP + FN \quad (5)$$

$$\text{Accuracy} = CM / (CM + IM) \quad (6)$$

If CM + IM = 100 in case of accuracy data

Model simulation describes the model's performance using Confusion Matrix as a measure of the quality criteria in the system to recheck the correctness of the learning voice commands.

In terms of accuracy, it is determined by the accuracy and appropriateness, which can be divided into 5 levels as follows :

- Accuracy of 90 to 100 percent means the efficiency is very good.
- Accuracy of 80 percent means the efficiency is good.
- Accuracy of 70 percent means the efficiency is average.
- Accuracy of 60 percent means the efficiency is less.
- Accuracy less than 60 percent means not efficiency.

VIII. Experiment result

The results of the test of voice commands from the answers used to test the sample between the ages of 55-60 years, 30 people to issue voice commands, turn on the lights, turn off the lights, amount 30 times. From the test results showed that 85% accuracy shows. To show that it can really help the elderly by using voice commands The results of the experiment according to the table :

Table II Test results voice command

No.	Voice command sentence	Repeat time
1	Turn on Bed Room's light	4
2	Turn on living room's light	4
3	Turn on every house's light	4
4	Turn off Bed Room's light	4
5	Turn off living room's light	4
6	Turn off every house's light	4
7	Call my son	2
8	Call ambulance	2
9	Open Line application	2
	Total	30

Table III System Validation result

Variable	Results
True Positive (TP)	63
True Negative (TN)	22
False Positive (FP)	8
False Negative (FN)	7
Correct Mapping (CM)	85
Incorrect Mapping (IM)	15
Accuracy	**85**

IX. Conclusion and Dicussion

A development voice commands for elderly assistance focus on a home electrical equipment control device via mobile application. The research found that the producer can develop a system to help the elderly to open applications installed on users' phones and turn on-off the light by voice commands as you learn the data set in the database by the researchers using voice recognition, natural language processing (Natural Language Processing), Arduino wifi board, Arduino program, Android studio program and Vs code program in further development and implementation with the help of Google APIs has made the elderly voice commands more efficient in voice commands.

The benefits of the development able to help the elderly system to control the on-off of the lights and can be extended to increase convenience for the elderly for greater convenience when alone at home and also being able to provide patients and disabled people.

Another point able to supporting elderly ability controlling mobile application for facilitate using voice commands calling other or open applications immediately

Development of a system to assist the elderly by using voice commands Can discuss that From the research objectives to develop a system to assist the elderly by using voice commands And assess the accuracy of the elderly help system by using voice commands The results of the research showed that The test number 30 messages 30 testers, the testers aged 55-60 years, gave an accuracy of 85 percent (Accuracy). The researchers applied Automatic Speech Recognition Technology and Natural Language Processing to use with voice commands that can help the elderly achieving goals efficiently.

The researchers evaluated the test by using the model describing the effectiveness of the Confusion Matrix model as the quality criteria in the system. Come to measure efficiency, accuracy But there are still limitations with the sound being interrupted, resulting in slow processing. And may cause audio errors that may cause the message to go wrong as well Speaking too fast until the system is unable to transcribe the spoken voice into the correct text. And internet speeds that affect slower performance related to Theerapong and Sittikom reserach [13] that provide the same recommendations Should increase the set of voice commands to be more diverse Easy to operate And increasing the amount of system learning set Therefore will increase the efficiency of using voice commands.

From the development of a system to assist the elderly by using voice commands Can be developed in the form of a relay (Relay) which is consistent with the research of Chomnuch that provides the same suggestions.[32]

E. Limitation

A development voice commands for elderly assistance have limitation and future work such as :

a) The system cannot operate voice commands other than the learning kit that has been learned for the elderly with voice commands.

b) The use of voice to operate the system should be clear to match the system's instructions.

F. Future work

c) Developing a system to assist the elderly by using voice commands can be applied to the model in the development that can be applied to the development of real-house lights on and off by using relay circuits for further development.

d) Increasing the amount of imported learning sets Will increase the efficiency of using voice commands.

REFERENCE

[1] Puchapong Nodethaisong. "6th Elderly Generation Survey year 2017". the Foundation for Research and Development National Statistic.(2560). [Online]. Available from: https://www.matichon.co.th/news-monitor/news_911854.

[2] Office of Health Promotion Fund. "Thailand growing to Aging Society". (2561). [Online]. Available from: https://www.thaihealth.or.th.

[3] Office of Health Promotion Fund. "Aged Socity with careness". (2560). [Online] . Available from: https://www.thaihealth.or.th

[4] Nunta Khunpakdee. "Voice Recognition". (2561). [Online]. Available from: https://medium.com

[5] Theerapong Prawanno, Sittikhum Prasert. "Development Android Application voice control direct marketing via cloud". (2559). [Online]. Available from: https://sites.google.com/site/priyyaniphnth/home

[6] Nattawut Aueypornlert. " Speech Recording".(2561). [Online]. Available from: https://biometricskmit.wordpress.com

[7] Kraisorn Seepbun. "Understand Board ESP8266". (2561). [Online]. Available from: http://dtecesp8266arduino.blogspot.com/2017/12/esp8266-esp8266-esp8266-wifi-3.html

[8] Foundation For Older Persons' Development. "Population Situation 2561". [Online]. Available from: https://fopdev.or.th

[9] Pongpissanu Noinang. "SPEECH RECOGNITION".(2561). [Online]. Available from: https://sites.google.com/site/pongpisanunoinang/

[10] Chai Wutiwiwatchai, Sadaoki Furui. " Voice Recognition". (2561). [Online]. Available from: http://digitized-thailand.org/index.php/2011-01-10-02-39-58-96/2011-01-10-03-49-40

[11] Nattanut Prompa. " Arduino" (2561). [Online]. Available from: https://www.arduinospro.com

[12] Paisal Kiattananan.Natural Language Processing (NLP). (2561). [Online] . Available from: http://dv.co.th/blog-en/get-to-know-natural-language-processing-nlp/

[13] Teerapong Prawanno, Sittikom Prasert. "Application Android control electricity via sound cloud". Ratchamongkol Lanna Technology University.2559

[14] Tanil Moungpol, Aueychai Intarasombat, Anut Maneechoti. "Voice trasmittion Text Application for Deaf people via Google API". RatchapatNakornpathom University. 2552.

[15] Tanill Mougpoll, Aueychai Intarasombat. "Voice Controller Electronicity with Android and Micro-Controller". RatchapatNakornpathom University.2560.

[16] Jakkrit Sanemahut. "Travel Information Voice Search Engine". Nareunsuan. 2557.

[17] Preedakom Kaewtachart, Vorapoj Keawtachart. "Voice control lighting System". Mahasarakham Technical College.2560.

[18] Burasskorn Yusook, Kongtep Boobmee, Worapan Sarasureeporn et al,. " Voice command application to enter English lessons". Rajamangala University of Technology, 2558

[19] Ekkachai Noawanich. "Thai consonants training program for hearing impaired people using artificial neural networks" King's Mongkut University of Technology North Bangkok, 2550.

[20] Olan Daowien, Boonchareon Wongkittisuksa, Sawit Tuntanuch et al,. "Design of speech-dependent speech recognition system for stroke patients". Prince of Songkla University, 2553.

[21] Jessda Kajornrit, Piyanuch Chaipornkaew, Nuengrutai Aengchoin. "Application of Internet of Things technology to control lighting systems for smart homes". Dhurakij Pundit University, 2560.

[22] Chutamanee Tungkatoch. "Creating dialogue content from the website for use in chat robots". Chulalomgkorn University. 2553.

[23] Paul Jasmin Rani, Jason Bakthakumar, Praveen Kumaar.B, Praveen Kumaar.U and Santhosh Kumar. VOICE CONTROLLED HOME AUTOMATION

SYSTEM USING NATURAL LANGUAGE PROCESSING (NLP) AND INTERNET OF THINGS (IoT). [Retrieved December 4, 2018]. Third International Conference on Science Technology Engineering & Management (ICONSTEM) from https://ieeexplore.ieee.org/document/8261311/

[24] Cyril Joe Baby, Faizan Ayyub Khan and Swathi J. N. Home Automation using IoT and a Chatbot using Natural Language Processing. [Retrieved December 4, 2018].School of Electronics Engineering School of Computer Science and Engineering VIT University. From https://ieeexplore.ieee.org/document/8245185//

[25] Ly Pichponreay, Chi-Hwan Choi, Jin-Hyuk Kim, Kyung-Hee Lee andWan-Sup Cho. " Smart Answering Chatbot based on OCR and Overgenerating Transformations and Ranking". [Retrieved December 4, 2018]. Chungbuk National University. From https://ieeexplore.ieee.org/document/7536948/

[26] Dongkeon Lee, Kyo-Joong Oh, Ho-Jin Choi Korea Advanced. "The ChatBot Feels You – A Counseling Service Using Emotional Response Generation". [Retrieved December 4, 2018]. School of Computing, Korea Advanced Institute of Science and Technology. From https://ieeexplore.ieee.org/document/7881752/

[27] Sorawit Saengkyongam. "Term Frequency". 2562[Online]. Available from: https://lukkid dd.com/tf-idf

[28] Napong Wanitchayapong. "Confusion Matrix" 2562. [Online]. Available from: https://plagad.wordpress.com/2010/08/26/confusion-matrix/

[29] Anoon Morsuwan. "Arduino Article". (2561). [Online]. Available from: https://www.thaieasyelec.com/article-wiki/basic-electronics.articles-arduino-what-start-arduino - arduino.html

[30] Arthit Piromkit. "Board Arduino Uno R3". (2561). [Online]. Available from: http://dd4toew.blogspot.com/2017/05/arduino-uno-r3.html

[31] Mitchell Tom. "Decision Tree" (2561). [Online]. Available from: https://th.wikipedia.org/wiki/decisiontree

[32] Chompunuch Sukpol. " Development of an electrical control system for switching on and off using Thai spoken voice commands By using the Kenier Reserber technique (k-NN)". Rajaphat Nakornpathom University, 2552.

Intelligent Medicine Identification System Using a Combination of Image Recognition and Optical Character Recognition

Nagorn Maitrichit
Image Information and Intelligence Laboratory
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakhon Pathom, Thailand
nagorn.mat@student.mahidol.ac.th

Narit Hnoohom
Image Information and Intelligence Laboratory
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakhon Pathom, Thailand
narit.hno@mahidol.ac.th

Abstract—This research aims to develop an automatic verification system with deep learning techniques to verify prescription dispensing accuracy. The proposed method will be able to help pharmacies to reduce errors that lead to patients receiving the wrong medicine to patients. The system consists of two models: image classification and text classification. The image classification model uses raw medicine blister pack images, then removes the background to interpret the features based on the pattern recognition for Histograms of Oriented Gradients (HOG) of the model. It is composed of Convolution Neural Network (CNN), Linear Regression, and Logistic Regression. The text classification model uses text extraction to obtain imprints appearing on the blister package then matches the words to a bag of word. The dataset collected two-hundred types of medicine blister packs images inside plastic zip bags as a dataset. It includes 300 high-quality images of front-side medicine blister packages for each type of package in light-controlled conditions with a black background, which are used for training the model. The automatic verification system uses the majority vote based on the confidence of the two models. Experimental results, indicate that the image classification model of CNN with HOG feature extraction has the highest accuracy at 95.83 percent. In-text classification results show that the method using Character Region Awareness For Text detection (CRAFT), Keras-OCR, and text correction gave the highest accuracy at 92 percent. Overall accuracy was 94.23 percent.

Keywords—image classification, bag of words, text extraction, deep learning, medicine blister pack

I. INTRODUCTION

A general hospital uses numerous medicines to treat illnesses. The drug dispensing statistics from the Rajavithi Hospital reveal that the drug dispensing process has errors of 3.8 percent [1]. In the dispensing process, a pharmacy looks up the medication order in a Medication Management System to know the type of medicine, quality, and doctor information. The pharmacist receives the medicine from a drug store and prints off the counseling documents to provide to the patient with more information on the medication in the drug, which are put into a plastic zip bag for the patient. During this process, pharmacists can be distracted and make errors due to multiple reasons, which include a large volume of patients, doctor's handwriting, stress, heavy workload, long hours of working, and a low number of pharmacists. The issue becomes serious especially when it involves children or the elderly. The ultimate objective of this research is to design and develop an intelligent medicine identification system using a combination of image recognition and optical character recognition models to verify prescription dispensing accuracy that can recognize medications inside plastic zip bag. The purpose is to ensure that the brand, quantity, and drug in the form of blister packages to be dispensed, exactly match the information specified in the prescription.

II. RELATED WORK

In the literature, most of the organizations conduct the pill identification of pills, tablets, and capsules by using shapes, colors, and imprints. Image processing and machine learning have been applied to extract relevant features with a trained classifier to accomplish tablet or capsule identification. Y. Lee and U. Park [2] proposed an automatic method to match drug pill images by using imprints appearing on the pills with edge localization and invariant moments features for classification. The experiments showed 76.74 percent matching accuracy. Chen, R.-C and Chan [3] proposed a neural network, fuzzy method, and relevant feedback for classifying pill shapes by using five features of drug images including colors, shapes, ratio, magnitudes, textures, and dynamic weight. The dataset included 2,116 images with 822 types of drug pills were downloaded from multiple web sites. Experiment results showed a 92.60 percent matching accuracy. S. Suntronsuk [4] proposed extracting texts from pill images with a technique using Imprint Area and Kasar for processing edge masks of the imprint. Experiment results showed that the text imprint extraction method has 56.67 percent accuracy. Charlene Tay [5] proposed using feature extraction of the image using Hu Moments and other shape descriptors to build a classification model. Experiment results showed that using a shape classifier with the decision method has 81 percent accuracy. Overall, organizations classify the pills, except for Jing- Syuan Wang [6], who proposed the identification of blister packages. Specifically, Highlighted Deep Learning (HDL) uses automatic detection and segmentation to remove the background to process raw blister packages images, then builds a ResNet CNN model to classify the correct blister packages types. The dataset was from an experiment at the adult lozenge dispensing of MacKay Memorial Hospital. It consisted of 272 types of blister packages, with 65 images for each type (font and back) for image training and 7 images for validation, a total of 39,168 images. The experiments built

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

three models consisting of the cropped front view, cropped back view, and signature template (both-side view). The experiments showed almost 100 percent accuracy by the signature template approach. W. Chang [7] proposed ST-Med-Box assist chronic patients to avoid in taking the wrong medications based on deep learning techniques. Experiment results showed that accuracy reaches 96.6 percent for classify eight types of different medicines. Although many solutions have been published to help the patients forget take the wrong medication, their solution is focused on pill, tablets, and capsules that are not focused on blister packages in zip bags.

III. System Overview

This research project aims to develop a model for identification of blister packages in zip bags by using machine learning techniques. It is expected that the model can be used in a prototype machine to help pharmacies reduce errors that cause patients to receive the wrong medicine while less computation time and guarantees reliability. Figure 1 shows the research design.

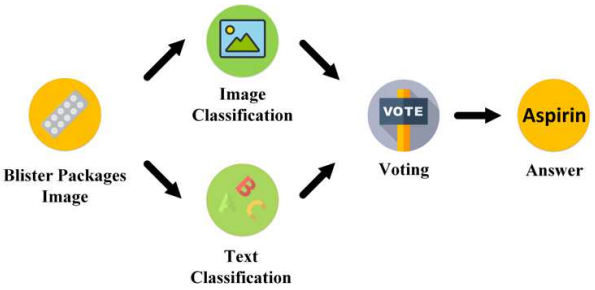


Fig. 1. System overview of the research design.

The Intelligent Medicine Identification System is divided into five steps that have been taken towards processing the image for blister packages identification. The blister packages images dataset is processed into two models, image classification and text classification. Following that, image classification is performed to separate the blister package from the background through processes for image classification. Otherwise, text classification for matching text appears in the blister package to a bag of words. Result from two model processes into results based on the confidence of the answer for each model. The four components of the system: blister image dataset, image classification, text classification, and voting, are described.

A. Blister Package Image

Fig. 2. Example of blister package with rectangle shape.

Fig. 3. Example of blister package with rounded rectangle shape.

The challenge dataset consists of 200 different types of blister packages from the outpatient dispensing station at Banphaeo General Hospital. In the dataset for machine learning training, researchers included 170 images for each type of package inside plastic zip bags with controlled environment lighting and positioning, totaling 34,000 images. Each blister package has recognition information on front side view which is the existing imprint information that contains the name of the manufacturer, trade name, expiry date, prohibitions and precautions, registration number, and warnings of dangerous drugs or special controlled drugs. The input images are the main rectangle or rounded rectangle shapes, as shown in Fig. 2 and Fig. 3.

B. Image Classification

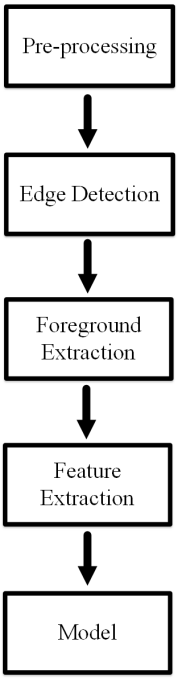


Fig. 4. Flow chart of the image classification.

The image classification is a process in computer vision that classifies the blister package type. This step is divided into five parts: pre-processing, edge detection, foreground extraction, feature extraction, and model, as shown in Fig. 4.

1) Pre-processing

(a) Input Image (b) Segmentation (c) Smoothing

Fig. 5. Schematic of pre-processing.

All the blister package images were already dataset imaged on a uniform background. The segmentation process resized the images to 300 × 300 pixels, converted them from RGB to the HSV model, and performed gaussian blur to reduce the high frequency, as shown in Fig. 5.

2) Edge Detection

(a) Candy Edge Detector (b) Morphological Operation (c) Contours

Fig. 6. Schematic of edge detection.

The edge detection is a process to find the boundary of the blister package from the blister package image. It is performed using canny edge detection [8], morphological operation [9], and contours [10], as shown in Fig. 6.

3) Foreground Extraction

(a) Corner Location (b) Reorientation

Fig. 7. Schematic of foreground extraction.

Foreground extraction is a process to extract the foreground from the image. It consists of a corner location and reorientation. The corner location is a process to approximates a polygonal curve from the contours of the image with a specified precision. It returns four polygonal points in images.

Reorientation is an image transform that corrects for the effects of perspective. It takes the four points representing the outline of the images and applies a perspective transform to adjust the orientation adjustment in a top-down or birds eye view of the image in order to ensure that blister package images have an upright orientation, as shown in Fig. 7.

4) Feature Extraction

(a) Input (b) Result

Fig. 8. Blister package processed by feature extraction.

After the process removes the background in the foreground extraction step, the HOG is used to extract the characteristics of local shape or features of a blister package that extracted location of a dense grid on an image. It is divided into small cells and the histogram of the border that occurs in a cell is stored by calculating the size and direction in horizontal and vertical borders, then a classification model is created to classify the combined features. Although the HOG gives an accurate description of the contour of blister packages images, it requires a large computational time. Figure 8 shows the results of blister package images process by HOG.

C. Text Classification

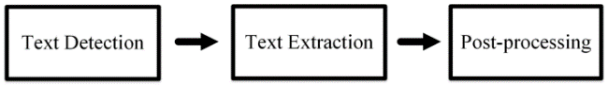


Fig. 9. The flow chart of the text classification.

The text classification is a process using imprint information on the blister package that classifies the blister package type. This step is divided into three parts: text detection, text extraction, and post-processing, as shown in Fig. 9.

1) Text Detection

(a) Input Image (b) Text Detection (c) Sub-Images

Fig. 10. Schematic of pre-processing.

Text detection is the method of locating areas in an image wherever the text is present. It marks the imprint area as separate from the blister package with a red rectangular bounding box around a sentence and creates a subdivided image from a rectangular bounding box into multiple images. In one image, there can be multiple rectangular bounding boxes, as shown in Fig. 10.

2) Text Extraction

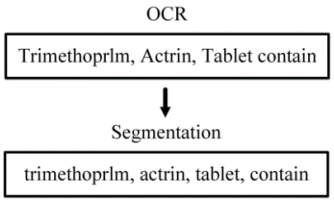


Fig. 11. Schematic of text extraction.

Text extraction is the process that makes the text from different types of documents, such as images, available for processing in computers. Segmentation is a process of changing sentences from OCR into new words using lowercase letters such as "Tablet contains" to "tablet" and "contain", as shown in the example in Fig. 11.

3) Post-processing

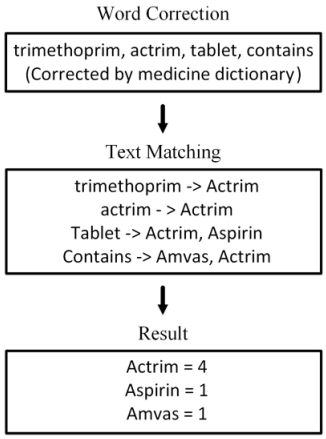


Fig. 12. Schematic of post-processing.

Post-processing is a process to increase the accuracy of the text classification. It consists of word correction and text matching. Word correction is a process that allows correction of misspelled words that are read from OCR. It corrects text by looking up a similar word in a medical dictionary that is closest to the candidate. Text matching is a process that identifies the names of medicine by comparing words and sentences with a bag of words in which the medicine names that are most similar to that of the drug are found and a result is returned, as shown in Fig. 12.

IV. Experimental Results

A. Image Classification

In this experiment, the researchers created three types of models by using 200 types of blister packages, each of which consists of Linear Regression, Logistic Regression, and CNN model. Table I shows the experiment results on image classification.

TABLE I. EXPERIMENT RESULTS OF IMAGE CLASSIFICATION MODEL

Trained model	200 types of Medicine		
	Logistic Regression	***Linear Regression***	***CNN[a]***
Train-test (80:20)	91.06 %	92.18 %	98.13 %
10-Fold	90.84 %	91.13 %	95.83%

[a.] Convolutional Neural Network (CNN)

The image classification performance of the CNN model was compared with those of other model, such as Logistic Regression, Linear Regression, and CNN. The overall accuracies of the methods that proposed methods in the image classification of blister package datasets. The result shown in Table I demonstrated that the CNN achieves the highest classification accuracy (91.06 percent, 92.18 percent, and 98.13 percent) in train-test (80:20) and (90.84 percent, 91.13 percent, and 95.83 percent) in 10-Fold for datasets. The CNN has achieved the highest image classification accuracy because CNN is the operation of finding patterns. It has kernel that basically scan an image and place where kernel have one hundred percent matched.

B. Text Classification

The researchers used the two pre-trained models for text detection Tesseract OCR [11] and Keras-OCR (CRAFT [12] and Keras CRNN [13]) by using 200 types of blister packages for testing the accuracy of the models. Table II shows the results of the text classification models.

TABLE II. EXPERIMENT RESULTS OF TEXT CLASSIFICATION MODEL

Pre-trained model	200 types of Medicine	
	Without Word Correction	***With Word Correction***
Tesseract OCR[b]	52 %	78 %
Keras-OCR (CRAFT[b] + Keras RCNN)	83 %	92 %

[b.] Optical Characteristic Recognition (OCR)

[c.] Character-Region Awareness For Text Detection (*CRAFT*)

The accuracy of the models is improved by using word correction in the post-processing approach in Tesseract OCR and Keras-OCR because without word correction some characters will be read incorrectly. This may provide another word that does not exist in the database, causing the text-matching process for the comparison of word sentences in a bag of word from drug names to be mismatched.

C. Voting

Voting is a technique that combines the models to produce improvement of the results from text and images-based for the confidence answer. If any model of prediction is more accurate than the other model, it is used as a final prediction. Table III shows the overall accuracy of increase by ensemble CNN with Keras-OCR.

TABLE III. EXPERIMENT RESULTS OF VOTING

200 types of Medicine	
Model	***CNN + Keras-OCR[d]***
Accuracy (Train-test 80:20)	94.23%

[d.] Optical Characteristic Recognition

V. Conclusion

In this paper, the researchers proposed an image identification system for medicines using a combination of image classification and text classification. In this system pattern recognition and imprint of blister package inside zip

bags are extracted. The overall accuracy rate is 94.23 percent. The experiment confirms that this method is feasible and effective, especially for high contrast blister packages. However, the image classification and text classification techniques are still imperfect due to the reflection of the plastic zip bags, similar colors, magnitude, and texture. In future work, the researchers would explore the use of another type of image segment from image processing to machine learning in image segmentation to increase the accuracy of the blister package identification of the system.

ACKNOWLEDGEMENT

This research was supported by Thammasat University research grant TU2017015 for the project “Digital Platform for Sustainable Digital Economy Development”, based on the RUN Digital Cluster collaboration scheme and supported by Banphaeo General Hospital on medical knowledge and research materials. This research was also supported by the Department of Computer Engineering, Faculty of Engineering at Mahidol University.

REFERENCES

[1] Kristina, S., Supapsophon, P., and Sooksriwong, C., 2020. Dispensing Errors: Preventable Medication Errors By Pharmacists In Outpatient Department, A University Hospital, Bangkok, Thailand. [online] Aasic.org. Available at: <http://aasic.org/proc/aasic/article/view/90> [Accessed 21 August 2020].

[2] Y. Lee, U. Park, and A. K. Jain, "PILL-ID: Matching and Retrieval of Drug Pill Imprint Images," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 2632-2635, doi: 10.1109/ICPR.2010.645.

[3] Chen, R.-C & Chan, Yung-Kuan & Chen, Y.-H & Bau, C.-T. (2012). An automatic drug image identification system based on multiple image features and dynamic weights. International Journal of Innovative Computing, Information, and Control. 8. 2995-3013.

[4] S. Suntronsuk and S. Ratanotayanon, "Pill image binarization for detecting text imprints," 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, 2016, pp. 1-6, doi: 10.1109/JCSSE.2016.7748859.

[5] Charlene Tay. Pharmaceutical Pill Recognition Using Computer Vision Techniques. Spring 2016.

[6] Y. Lee, U. Park, and A. K. Jain, "PILL-ID: Matching and Retrieval of Drug Pill Imprint Images," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 2632-2635, doi: 10.1109/ICPR.

[7] W. Chang, L. Chen, C. Hsu, C. Lin and T. Yang, "A Deep Learning-Based Intelligent Medicine Recognition System for Chronic Patients," in IEEE Access, vol. 7, pp. 44441-44458, 2019, doi: 10.1109/ACCESS.2019.2908843.

[8] Canny, J. 1986. A computational approach to edge detection. *IEEE Trans. Pat. Anal. Mach. Intell.*, 8(6):679–698.

[9] P. Maragos And L. Pessoa “Morphological Filtering For Image Enhancement And Detection,” Chapter For The Image And Video Processing Handbook, Acad. Press, March 1999.

[10] Elder, J., and Zucker, S. 1996. Computing contour closures. In Proc. Euro. Conf. Computer Vision, Vol. I, Cambridge, England, pp. 399–412.

[11] Qiuhau Lin Tong Zhang and Derek Ma. Mobile camera based text detection and translation. IEEE, 7(2):87–105, 2010.

[12] Baek, Youngmin et al. “Character Region Awareness for Text Detection.” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019): 9357-9366. Shi, B., Bai, X., & Yao, C. (2017).

[13] An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Medicine Identification System on Mobile Devices for the Elderly

Pitchaya Chotivatunyu
Image Information and Intelligence Laboratory
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakhon Pathom, Thailand
pitchaya.cht@student.mahidol.ac.th

Narit Hnoohom
Image Information and Intelligence Laboratory
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakhon Pathom, Thailand
narit.hno@mahidol.ac.th

***Abstract*—This research develops an application that helps the elderly to identify medicine from a mobile image, to reduce confusion in taking medication, and thus to reduce the rate of medication errors. The data used in this research are collected from the medicine blister packs for the elderly consisting of 14 types of medicine, which are taken with the smartphone cameras and amounting to a total of 56,000 single medicine blister pack images for image classification model training. For object detection model training, there are a total of 21,000 single medicine blister pack images with added multiple medicine blister pack images amounting to 120 images from the image dataset. Text recognition is used to identify the medicine type using Keras-OCR. For all experimental results in the image classification model experiments reveal that the MobileNet V2 with 14-class detection has the highest accuracy at 93.79 percent. The object detection model is the MobileNet V1 with the highest mAP of 0.875 with the Average Precision with 0.5 IoU and 0.75 IoU at 0.998 and 0.91, respectively.**

Keywords—medicine blister packs, image processing, image classification, deep learning, object detection, text recognition

I. Introduction

In recent years, the elderly rate has been rising more than 20 percent, while on the other hand, the birth rate has been decreasing, which results in an older society in Thailand. The research from the Department of Older Persons [1] reports that in February 2019, there were 11 million elderly in Thailand, with females accounting for 4.9 million people and males for 6.1 million people. Due to the rapid rise in the numbers of the elderly, there should be a preparation plan for the upcoming elderly society. When entering old age, there is an inevitable deterioration in natural abilities including the onset of diseases which begin in middle age, placing the elderly as a group at risk of physical and mental problems, mental disorders, and natural diseases, often including Alzheimer's, diabetes, high cholesterol, hypertension, heart disease, prostate cancer, osteoporosis, and mental and social conditions. The elderly also experience problems of intellectual deterioration, such as confusion and depression, which may cause dangerous behavior such as forgetting to take medication, repeatedly taking medication, or forgetting doctor appointments, resulting in an inability to handle the symptoms and complications of therapeutics.

To address the aforementioned concerns on older society in Thailand, this research presents an application that helps the elderly to identify medicine from a mobile image, and thus to reduce the rate of medication errors. At present, the application can recognize the medicine packages of the 14 most common types of medication that cover the 7 most common illnesses or conditions usually found in patients, which are Alzheimer's, diabetes, hypertension, high cholesterol, heart disease, prostate cancer, and osteoporosis. To avoid taking the wrong medications, the application will provide the treatment of the medicine, the manufacturer, and generic name on medicine packs for each of the medicines the application can recognize. Furthermore, the application will be used as a medium between caretakers and the elderly to notify and monitor the medicine taking of the elderly.

II. Related Work

Deep learning is an arithmetic method created for machine learning based on the Convolutional Neural Network (CNN). Deep learning is motivated by the architecture and functionality of the human brain by interconnecting several neurons. The computer vision systems, natural language processing systems, audio recognition systems, bioinformatics systems, speech recognition systems, and image analytics systems, all of which now make extensive use of deep learning algorithms. This research is mainly interested in image classification, object detection, and text recognition methods.

Several recent research works have leveraged image processing and deep learning algorithms to classify medicine images [2–8]. Dongsun Kim and Junchul Chun [2] proposed retrieval of the content-based images which are used on 873 pill image cases with 18 different types of shapes in the National Medical Library (NML) database. The shape classification uses the k-mean clustering algorithm and the Canny edge detector. The authors can prove that the method works efficiently when the database has many similar shaped images. Yaniv et al. [3] proposed a competition on pill image recognition on the RxIMAGE collection. The dataset includes 1,000 classes of pills. A training dataset of 7,000 images: 2,000 high-resolution images and 5,000 low-resolution quality images and the testing dataset consisting of 6,486 low-resolution quality images of identical shape and color distributions were used. The top 3 competitors were "nhatuntsev", "castelo", and "msumpf" using feature space distances, which are defined as distances weighted sum in color and shape space, CNN from Tensorflow open-source, and deep learning on features obtained from the Scale-Invariant Feature Transform (SIFT) descriptor method, correspondingly. The top 3 winners got mAP scores of 0.08, 0.09, and 0.27, correspondingly. Wang et al. [4] proposed a highlighted deep learning-based medicine blister package identification approach on 272 classes of medicine blister packages. The dataset contains images with front and back for each type of packages of 72 images, of which 65 images of each side were split for training the model, and 7 images of each side were split for testing the model. Then the conventional Deep Learning (DL) approach was followed by using ResNet classifiers on normal images and the

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

highlighted deep learning approach by using ResNet classifier on images with a juxtaposed package. The results from the conventional DL approach give 92.68 percent and 91.63 percent from the testing images are correctly identified. From the highlighted DL approach the result was 99.84 percent of the testing images. Chang et al. [5] proposed an ST-Med-Box, which is a medicine recognition system based on deep learning techniques. The system included medicine recognition devices, mobile applications, and cloud-based management platforms. The dataset includes 4,000 single drug images and 1,000 multiple drug images on 8 classes of drugs. The recognition accuracy reaches up to 99.5 percent for each of the 8 types of medicine. Magalhães et al. [6] proposed a three-staged approach on medicine box recognition which uses text recognition, barcode recognition, and feature matching techniques for each stage. Image features are used to identify the medicine box through a set of distinctive features such as Speeded Up Robust Features (SURF), SIFT, or oriented Features from Accelerated Segment Test (FAST), and rotated Binary Robust Independent Elementary Features (BRIEF). The results of barcode recognition, text recognition, and feature matching techniques are 80 percent, 80 percent, and 100 percent success rates, correspondingly. Ou et al. [7] proposed an automatic pill image detection approach on the convolutional neural network including detection and classification methods. The dataset contains 131 classes of pills, which compose of 1,680 images with 3,144 annotations for detection and 470,000 images for classification. The pill detection uses Feature Pyramid Networks (FPN) with Resnet-50 and the pill classification method uses Xception. For testing, the 400 images with 2,825 annotations were used and obtained test accuracy of Top 1, Top 3, and Top 5 of 79.4 percent, 88.3 percent, and 91.8 percent correspondingly. Taran et al. [8] purposed a fine-grained recognition of pharmaceutical packages from the mobile image. The PharmaPack dataset used contains 1,000 classes of pharmaceutical packages using consumer mobile phones. The dataset is tested on SIFT and aKaZe descriptors with recognition based on the RANdom SAmple Consensus (RANSAC) matching using a pre-defined number of descriptors to be 300, 500, and 1,000. Therefore, the SIFT descriptor gives a better result than the aKaZe descriptor for all use parameters due to the SIFT descriptor providing more informative representations.

III. Proposed Methods

The medicine packs will be identified by using the deep learning algorithms which are object detection, image classification, and text recognition. This research will cover the processes as shown in Fig. 1. The medicine identification process will use object detection, image classification, and text recognition to find the type of medicine blister package image.

A. The Image Acquisition

The 14 types of medicine blister packs images are taken from cameras and smartphones. The whole dataset contains a total of 14,000 images with a variety of background, perspective, and light conditions in each image. The dataset is composed of 2 sets: the dataset for the classification model and the detection model. The classification model uses a single medicine blister pack containing 14 classes of medicine and consisting of 4,000 images per each class.

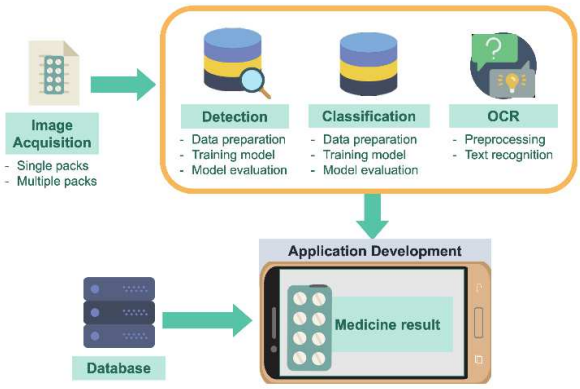


Fig. 1. Research Development Processes.

The detection model uses a single medicine blister pack dataset consisting of 3,000 images per class and the multiple medicine blister packs dataset consisting of 120 images. The image contains multiple types of medicine and places in a random position and background as shown in Fig. 2 and Fig. 3.

Fig. 2. Example of a single medicine blister pack image.

(a) Multiple medicine packs image I (b) Multiple medicine packs image II

Fig. 3. Example of multiple medicine blister packs dataset for object detection.

1) Data preparation

This process is the process of creating useful information datasets by preprocessing images or cleaning datasets to provide the best input for training models. In this research, the dataset is divided into 2 sets: a single medicine blister pack image dataset and a multiple medicine blister packs image dataset. For a single medicine blister pack dataset with a total of 21,000 images, a training set and a testing set are divided to the ratio of 8:2 which are 1,200 images and 300 images per class. In the multiple medicine blister pack dataset, 120 images of multiple blister pack images are added to the previous training set.

The images are then labeled and an XML file is created keeping information of the image such as filename, the bounding box of the interest object, image size, and

annotation of the object. The annotation of the object in this research consists of the 14 medicine names.

2) Training models

The object detection model uses a pre-trained model from Google TensorFlow posted on Github called model zoo [9]. The Single Shot Detector (SSD) MobileNet V1 [10], SSD MobileNet V2 [11], and Faster Regional Convolution Neural Network (Faster RCNN) Inception V2 [12] architecture are used as a pre-trained model in this research. The object detection will return the predicted class and the bounding box of the object, with the standard loss function for classification of prediction class and localization of the bounding box. The evaluation metric for the object detection model is using mAP values.

B. Classification

1) Data preparation

In classification, a single medicine blister pack dataset is used for model training. The dataset is divided into a training dataset and a testing dataset of 3,200 and 800 images per class, correspondingly, which are in the ratio of 8:2. The images are placed into a subdirectory of each class of medicine in the training set and test set.

2) Training models

The image classification model used in this research is from the Google TensorFlow pre-trained model. The pre-trained models that will be used in this research are Inception V3 [13], Inception V4 [14], and MobileNet V2 [15]. When running the training script the bottleneck is created to analyze and caches from the training set. The model will be training with a learning rate of 0.1, 0.01, and 0.001 for a total of 500 epochs. The classification model result is returning the predicted class from the image using the standard loss function of prediction class from the Tensorflow pre-trained image classification Application Programming Interface (API).

The evaluation metric for the classification model is using accuracy level calculated from the confusion matrix which is a widely used metric for classification.

C. Optical Character Recognition (OCR)

1) Preprocessing

To detect text on an image, image preprocessing is an important step to achieve better text recognition results. The object detection model is used to crop the medicine image then the OpenCV Library [16] is used to preprocess the image.

Once the OpenCV library is successfully integrated, the application can preprocess the input image by cropping the image to get the interesting text area on the medicine pack. Each medicine image will undergo the OCR process for extract medicine names from the text in the detected medicine pack image.

2) Text Recognition

The text recognition method is using the Keras-OCR Library [17]. The Keras-OCR is the compose of the Keras Convolutional Recurrent Neural Network (CRNN) implementation for text recognition and the Character-Region Awareness For Text detection (CRAFT) text detection model, which provides a high-level API for training a text detection and OCR pipeline. The input image from the device camera as a bitmap then passes to the Keras-OCR API and return text result and bounding box of text in the image.

IV. Experimental Results

The object class was categorized by medicine blister package consist of 14 classes, which are "Ambes", "Amlopine", "Anapril", "Apolets", "Calciferol", "Fosamax", "Harnal Ocas", "Madiplot", "Miformin", "Millimed", "Prenolol", "Sandoz", "Vitabion", and "Zimmex", respectively. These medicines are used for the training model in this research.

A. Experiments on Object Detection

In the experiment of object detection, the input images for training, validation, and evaluation processes are generated by the image annotation tool in the data preparation process under the object detection section. There are a total of 21,000 single medicine blister pack images in the dataset to be input data, the 80 percent of the dataset is used for training and 20 percent for testing. The number of images in each dataset amounts to 16,800 images for training and 4,200 images for evaluation with multiple medicine blister pack images amounting to 120 images added for experimenting on detection model comparison.

This research experiments on model development with different scales of input images used; the experiment is divided into 5 experiments which are 14-class detection model training on 1,000 single medicine pack images per class, 1,000 single medicine pack images per class with 120 multiple medicine pack images, 1,500 single medicine pack images per class, 1,500 single medicine pack images per class with 120 multiple medicine pack images, and 3,000 single medicine pack images per class. The single medicine pack images and the multiple medicine pack images in which the images are used for training are as shown in Fig. 2 and Fig. 3, correspondingly.

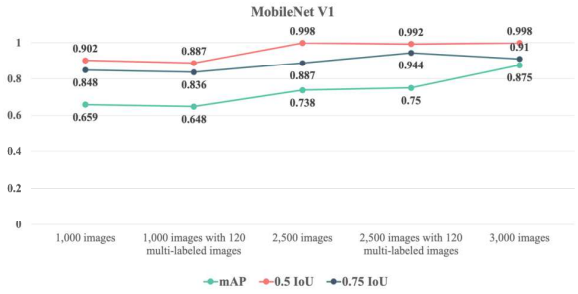


Fig. 4. Model evaluation comparison of MobileNet V1.

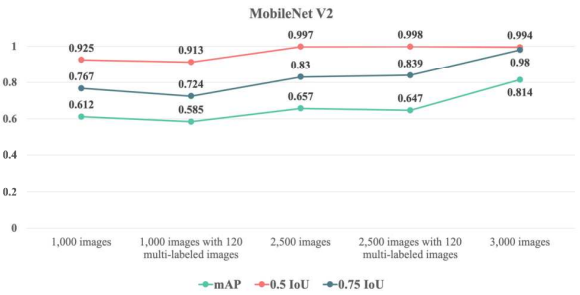


Fig. 5. Model evaluation comparison of MobileNet V2.

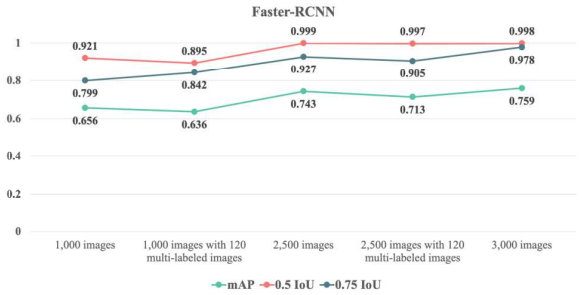


Fig. 6. Model evaluation comparison of Faster RCNN Inception V2.

According to the high performance and well-developed neural network models of TensorFlow object detection techniques, we only experiment in the aspect of model compatibility for this research. Participating models in this experiment comprise SSD MobileNet V1, SSD MobileNet V2, and Faster R-CNN Inception V2.

For all experimental results of object detection methods are demonstrated in Fig. 4 and Fig. 6 reveals that the detection model using higher images per class for training gives better results. From all the results from Table I, the Faster R-CNN Inception V2 with 14-class detection trained with 3,000 images per class has the highest mAP of 0.875 with the Average Precision of 0.5 IoU and 0.75 IoU at 0.998 and 0.91, respectively.

TABLE I. EXPERIMENTAL RESULTS OF OBJECT DETECTION MODELS

Architecture	Dataset	mAP	0.5 IoU	0.75 IoU
MobileNet V1	1000 images	0.659	0.902	0.848
	1000 images with 120 multi-labeled dataset[a]	0.648	0.887	0.836
	2500 images	0.738	0.998	0.887
	2500 images with 120 multi-labeled dataset[a]	0.75	0.992	0.944
	3000 images	0.875	0.998	0.91
MobileNet V2	1000 images	0.612	0.925	0.767
	1000 images with 120 multi-labeled dataset[a]	0.585	0.913	0.724
	2500 images	0.657	0.997	0.83
	2500 images with 120 multi-labeled dataset[a]	0.647	0.998	0.839
	3000 images	0.814	0.994	0.98
Faster RCNN Inception V2	1000 images	0.656	0.921	0.799
	1000 images with 120 multi-labeled dataset[a]	0.636	0.895	0.842
	2500 images	0.743	0.999	0.927
	2500 images with 120 multi-labeled dataset[a]	0.713	0.997	0.905
	3000 images	0.759	0.998	0.978

[a] Multiple medicine pack images dataset, which contains 120 images

B. *Experiments on Image Classification*

In the experiment of image classification, the input images for the training, validation, and evaluation process consist of the same 14 classes from the object detection dataset. There are a total of 56,000 single medicine blister pack images from the image dataset to be input data, 80 percent of the dataset is selected to be the training and testing dataset. The number of images in each dataset amount to 44,800 images and 11,200 images of training and evaluation datasets, respectively.

As in object detection experiments, we only experiment in the aspect of model compatibility for this research. Participating models in this experiment comprise MobileNet V2, Inception V3, and Inception V4. All experimental results in this experiment demonstrated in Table II reveals that the MobileNet V2 with 14 class detection has the highest accuracy at 93.79 percent.

TABLE II. EVALUATION RESULTS OF IMAGE CLASSIFICATION MODELS

Class	MobileNet V1	Inception V3	Inception V4
14 Classes (4,000 images)	93.79%	92.75%	94.85%

C. *Comparing Results*

From the previous section, the object detection model and classification model can both classify the medicine image. This section concerns on comparisons of the efficiency of the classification's weights.

For the classification model, the selected model was MobileNet V2 architecture. The model achieved an accuracy level on the evaluation dataset of 93.79 percent. The confusion matrix values are illustrated in Table III which shows the model can predict relatively accurate due to the overall accuracy, precision, and recall values. The bold character indicates that the "Sandoz" class gives the highest sensitivity, specificity, and precision of 1.00. The "Sandoz" class predictions give zero false negative values and 70 true positive value which is the total number of test images of the class. However, other classes have fairly high values of precision, sensitivity, and specificity.

TABLE III. CONFUSION MATRIX OF CLASSIFICATION MODEL

Name	TP	FP	TN	FN
Ambes	87	12	987	59
Amlopine	87	0	987	71
Anapril	86	1	988	70
Apolets	80	6	994	65
Calciferol	85	3	989	68
Fosamax	88	1	986	70
Harnal Ocas	80	3	994	68
Madiplot	46	15	1,028	56
Miformin	83	1	991	70
Millimed	56	14	1,018	57
Prenolol	83	4	991	67
Sandoz	**70**	**0**	**1,004**	**71**
Vitabion	53	7	1,021	64
Zimmex	90	4	984	67

For the object detection model, the selected model was MobileNet V1 architecture. The model achieved very precise localization of medicine packs, but for the classification, it achieves an accuracy level on the evaluation dataset of 82.34 percent. The confusion matrix is illustrated in Table IV showing the model can predict relatively accurate due to the overall precision and recall values. The bold character

indicates that the "Madiplot" class gives the highest sensitivity, specificity, and precision of 1.00. The "Madiplot" class predictions give zero false negative values and 249 true positive values which is the total number of test images of "Madiplot" class, while other classes have false-negative predictions which result in lower values of precision, sensitivity, and specificity.

TABLE IV. CONFUSION MATRIX OF OBJECT DETECTION

Name	TP	FP	TN	FN
Ambes	250	9	2,632	609
Amlopine	249	125	2,633	493
Anapril	248	87	2,634	531
Apolets	184	20	2,698	598
Calciferol	250	163	2,632	455
Fosamax	250	5	2,632	613
Harnal Ocas	250	192	2,632	426
Madiplot	**249**	**0**	**2,698**	**618**
Miformin	126	0	2,756	618
Millimed	247	11	2,635	607
Prenolol	220	0	2,662	618
Sandoz	184	0	2,698	618
Vitabion	93	3	2,789	615
Zimmex	82	3	2,800	615

TABLE V. CONFUSION MATRIX OF CLASSIFICATION ON DETECTED IMAGES

Name	TP	FP	TN	FN
Ambes	250	95	1,998	1,157
Amlopine	166	6	2,082	1,246
Anapril	**232**	**15**	**2,016**	**1,237**
Apolets	240	273	2,008	979
Calciferol	183	139	2,065	1,113
Fosamax	137	210	2,111	1,042
Harnal Ocas	31	0	2,217	1,252
Madiplot	52	0	2,196	1,252
Miformin	148	41	2,100	1,211
Millimed	228	180	2,020	1,072
Prenolol	53	12	2,195	1,240
Sandoz	209	18	2,039	1,234
Vitabion	72	11	2,176	1,241
Zimmex	247	252	2,001	1,000

For the experiment on enhancing the classification method, the detection model was used for localization then sending a cropped medicine image to the classification model. The accuracy of classification dropped to 64.22 percent, as shown in the confusion matrix in Table V. The values show the model can predict quite accurately due to the overall accuracy, precision, and recall values. The bold character indicates that the "Anapril" class gives the highest sensitivity, specificity, and precision of 0.93, 0.99, and 0.94, correspondingly. The "Anapril" class predictions give 15 false negative values and 232 of true positive value which is almost the total number of test images of the class. However, some of the other classes have fairly low values of precision, sensitivity, and specificity.

D. Experiments on OCR

1) Medicine name recognition

In the experiment of OCR, the input images for the evaluation process consist of the 14 medicine blister pack classes which are the same as in the object detection and image classification dataset.

TABLE VI. CONFUSION MATRIX OF TEXT RECOGNITION USING KERAS-OCR

Name	TP	FP	TN	FN
Ambes	25	0	324	1
Amlopine	25	0	324	1
Anapril	25	1	324	0
Apolets	25	0	324	1
Calciferol	25	0	324	1
Fosamax	25	0	324	1
Harnal Ocas	25	0	324	1
Madiplot	**24**	**0**	**325**	**1**
Miformin	25	0	324	1
Millimed	25	0	324	1
Prenolol	25	0	324	1
Sandoz	25	0	324	1
Vitabion	25	0	324	1
Zimmex	25	0	324	1

The experimental results reveal that the Keras-OCR with 14 class text recognition with a total of 25 single medicine blister pack images from the image dataset to be input data, has the highest accuracy at 99.71 percent. From Table VI, the confusion matrix demonstrates the evaluation of each class on Keras-OCR. It shows that only one of the images was classified incorrectly which is from the "Madiplot" class.

V. DISCUSSION AND CONCLUSIONS

To identify medicine blister packages by an efficient system, this paper proposes a medicine identification system on mobile devices for the elderly, where the image was taken undergoes 3 processes which are object detection, image classification, and text recognition. The final result will be given by weighting between the three results. The overall identification performance of each process, when evaluated on the test data is accurate. Hence, the associated four sides of the medicine can be efficiently and precisely located by the detection model. For the image classification model, the results were fairly high due to the model being trained with the whole images without cropping out the background, along with the high performance and well-developed neural network models of TensorFlow image classification algorithms. Lastly, the text recognition gave almost perfectly accurate results due to the input image being quite clear to read, hence it was small in size. The Keras-OCR did an excellent work of detecting and recognizing text, along with

the preprocessing function of the image before extracting the text on the images. The proposed system can effectively reduce the problem of medication errors caused by taking incorrect medicine, thus giving the elderly a safe medication environment.

In future work, the number of images in the dataset should be added to train image classification and object detection models to yield higher test accuracy. The expiry date can be extracted using the OCR method to caution the elderly when taking medicine to avoid consuming expired medicine.

ACKNOWLEDGMENTS

This research was supported by Thammasat University research grant TU2017015 for the project "Digital Platform for Sustainable Digital Economy Development", based on the RUN Digital Cluster collaboration scheme and supported by the Department of Computer Engineering, Faculty of Engineering at Mahidol University. This research was also supported by Banphaeo General Hospital on medical knowledge and research materials.

REFERENCES

[1] dop.go.th, "Situation and Trends of the Thai Elderly Society in 2013 - 2030 B.E." http://www.dop.go.th/th/know/1/47 (accessed Aug. 22, 2019).

[2] D. Kim and J. Chun, "Drug Image Retrieval by Shape and Color Similarity of the Medication," in *2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering*, May 2011, pp. 387–390, doi: 10.1109/CNSI.2011.40.

[3] Z. Yaniv *et al.*, "The national library of medicine pill image recognition challenge: An initial report," in *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Oct. 2016, pp. 1–9, doi: 10.1109/AIPR.2016.8010584.

[4] J.-S. Wang, A. Ambikapathi, Y. Han, S.-L. Chung, H.-W. Ting, and C.-F. Chen, "Highlighted Deep Learning based Identification of Pharmaceutical Blister Packages," in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sep. 2018, vol. 1, pp. 638–645, doi: 10.1109/ETFA.2018.8502488.

[5] W.-J. Chang, L.-B. Chen, C.-H. Hsu, C.-P. Lin, and T.-C. Yang, "A Deep Learning-Based Intelligent Medicine Recognition System for Chronic Patients," *IEEE Access*, vol. 7, pp. 44441–44458, 2019, doi: 10.1109/ACCESS.2019.2908843.

[6] L. Magalhães, B. Ribeiro, N. Alves, and M. Guevara, "A three-staged approach to medicine box recognition," in *2017 24º Encontro Português de Computação Gráfica e Interação (EPCGI)*, Oct. 2017, pp. 1–7, doi: 10.1109/EPCGI.2017.8124317.

[7] Y.-Y. Ou, A.-C. Tsai, J.-F. Wang, and J. Lin, "Automatic Drug Pills Detection based on Convolution Neural Network," in *2018 International Conference on Orange Technologies (ICOT)*, Oct. 2018, pp. 1–4, doi: 10.1109/ICOT.2018.8705849.

[8] O. Taran, S. Rezaeifar, O. Dabrowski, J. Schlechten, T. Holotyak, and S. Voloshynovskiy, "PharmaPack: Mobile fine-grained recognition of pharma packages," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug. 2017, pp. 1917–1921, doi: 10.23919/EUSIPCO.2017.8081543.

[9] "Tensorflow/models," *GitHub*. https://github.com/tensorflow/models (accessed Mar. 09, 2020).

[10] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *ArXiv151202325 Cs*, vol. 9905, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *ArXiv180104381 Cs*, Mar. 2019, Accessed: Mar. 10, 2020. [Online]. Available: http://arxiv.org/abs/1801.04381.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *ArXiv150601497 Cs*, Jan. 2016, Accessed: Mar. 10, 2020. [Online]. Available: http://arxiv.org/abs/1506.01497.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *ArXiv151200567 Cs*, Dec. 2015, Accessed: Mar. 10, 2020. [Online]. Available: http://arxiv.org/abs/1512.00567.

[14] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *ArXiv160207261 Cs*, Aug. 2016, Accessed: Mar. 10, 2020. [Online]. Available: http://arxiv.org/abs/1602.07261.

[15] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *ArXiv170404861 Cs*, Apr. 2017, Accessed: Mar. 10, 2020. [Online]. Available: http://arxiv.org/abs/1704.04861.

[16] "Object detection | TensorFlow Lite," *TensorFlow*. https://www.tensorflow.org/lite/models/object_detection/overview?hl=th (accessed Mar. 10, 2020).

[17] F. Morales, *faustomorales/keras-ocr*. 2020.

A Conversational Agent for Database Query: A Use Case for Thai People Map and Analytics Platform

Thikamporn Simud[1,2], Somchoke Ruengittinun[1], Navaporn Surasvadi[2],
Nuttapong Sanglerdsinlapachai[2] and Anon Plangprasopchok[2]
[1]Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand
[2]National Electronics and Computer Technology Center, Pathum Thani, Thailand
thikamporn.s@ku.th, somchoke.r@gmail.com,
{navaporn.surasvadi, nuttapong.sanglerdsinlapachai, anon.plangprasopchok}@nectec.or.th

***Abstract*—Since 2018, Thai People Map and Analytics Platform (TPMAP) has been developed with the aims of supporting government officials and policy makers with integrated household and community data to analyze strategic plans, implement policies and decisions to alleviate poverty. However, to acquire complex information from the platform, non-technical users with no database background have to ask a programmer or a data scientist to query data for them. Such a process is time-consuming and might result in inaccurate information retrieved due to miscommunication between non-technical and technical users. In this paper, we have developed a Thai conversational agent on top of TPMAP to support self-service data analytics on complex queries. Users can simply use natural language to fetch information from our chatbot and the query results are presented to users in easy-to-use formats such as statistics and charts. The proposed conversational agent retrieves and transforms natural language queries into query representations with relevant entities, query intentions, and output formats of the query. We employ Rasa, an open-source conversational AI engine, for agent development. The results show that our system yields F1-score of 0.9747 for intent classification and 0.7163 for entity extraction. The obtained intents and entities are then used for query target information from a graph database. Finally, our system achieves end-to-end performance with accuracies ranging from 57.5%-80.0%, depending on query message complexity. The generated answers are then returned to users through a messaging channel.**

***Index Terms*—Conversational agent, Chatbot, Data analytics, Data visualization, Poverty alleviation**

I. Introduction

Nowadays, most organizations analyze their big data for making quick and flexible decisions to stay competitive in the business. For example, in healthcare and clinical research, advanced analysis on both structured and unstructured clinical data enables intelligent solutions to improve speed and efficiency in providing life-saving diagnosis and treatment. In manufacturing industry, new cost-saving and revenue opportunities may be discovered. In commercial business, firms use big data analytics to predict the market trends for making investment decisions on new products or boosting profitability. Big data analytics also has an enormous impact on a government to improve efficiency in public management [1]. The government can make vital decisions, strategic plans, and public policies that affect millions of people using big data analytics. Those obtained products are used for developing necessary infrastructure, e.g., transportation, healthcare, education, and agriculture, and also solving the poverty problem. By solving the poverty problem, the government can maintain domestic tranquility, achieve sustainable development, secure citizens' basic rights, and promote the general welfare and economic growth [2].

Thai People Map and Analytics Platform (TPMAP) [3] is a tool that supports government organizations to discover more efficient and innovative policies to decrease poverty problem. Due to data complexity, the present version of TPMAP does not have a flexible user interface and it requires certain technical skills to query the data. Only programmers or data scientists can query and summarize the data. Policy makers, on the other hand, need to request technical users to query the data for them. With several parties involved in the process, the time from making a data request to obtaining results in a ready-to-use format is then stretched. Furthermore, the data obtained from the technical users may not exactly match the required one as a result of miscommunication between them.

A conversational agent or chatbot, which has been developed and used as a personal assistant in various fields, is a solution for addressing the lack of technical skill issue. Chatbot has become extremely popular for not only in customer service domain, but also in education, e-commerce, or healthcare domains [4]. Advantages of chatbots include increasing convenience, reducing the costs for services and supports, enabling one-to-one marketing, and also creating new methods for collecting data [5]. The factor that made chatbots widely available is recent rapid developments of artificial intelligence (AI), especially natural language understanding (NLU). Many AI platforms are equipped with building blocks for users to easily and quickly build a chatbot system, e.g., Facebook Bots for Messenger[1], Google Dialogflow[2], IBM Watson Assistant[3], Microsoft Azure Bot Framework[4], and Rasa[5].

In this paper, we developed a chatbot by applying Rasa, an open-source framework for building the conversational pipeline. The main advantages of the framework are that

[1]https://wit.ai
[2]https://dialogflow.cloud.google.com
[3]https://www.ibm.com/cloud/watson-assistant
[4]https://www.luis.ai
[5]https://rasa.com

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

it is flexible to be modified and the entire chatbot can be deployed on premises [6]. Basically, Rasa consists of two key components: Rasa Core and Rasa NLU [6], [7]. The former module is for managing conversation dialog and the latter is for recognizing intent and entities from an input message. Both of them can run separately or sequentially in a data processing pipeline. The intent and entities extracted from Rasa NLU are then transformed into a graph database query for acquiring information from TPMAP. Currently, the system supports two types of responses: text messages and chart images.

The remainder of this paper is organized as follows: Section II provides a summary of related works. Section III describes our developed conversational agent. Section IV presents experimental setting including all resources used in this study. Section V shows the experimental results and discusses about them. Section VI concludes the study.

II. Related Work

Table I summarizes conversational agents developed as a key component of self-service data analytics systems in many domains. Their tasks are ranging from fundamental ones, e.g., storing data and supporting basic inquiries, to more advanced ones, e.g., helping users to understand things. Many chatbots, e.g., AgronomoBot [11], WeightMentor [15], and the work of Choi et al. [9], have been used for not only answering questions, but also showing summary reports.

Regarding data visualization in chatbots, the work in [18] surveyed appropriateness of data presentation in chatbots and found that more than 50% of questions concerning data comparisons or trends are preferred to be responded with charts or graphs instead of text messages. Several recent works have implemented chatbots that can generate graphs and charts. GameBot [17], for example, can generate a chart that summarizes statistics for a sport game or an individual player. Lauren et al. [13] also developed a chatbot to plot a historical graph showing openning, closing, high and low stock prices. Additionally, Bieliauskas et al. [8] developed a chatbot to simplify complex visualizations.

For Thai conversational agent, many studies have focused on "AI-based" chatbots. For example, Robloke et al. [14] developed a chatbot to intelligently collect data from users. Muangkammuen et al. [12] also developed a chatbot for answering Thai FAQs using deep learning technique. To the best of our knowledge, however, there has been no Thai chatbot that is able to visualize complex query results from a database for the purpose of self-service data analytics yet. Hence, we developed the first Thai chatbot for self-service data analytics on a graph database which can produce responses with charts for visualizing complex information. Our chatbot is partially similar to some previous works, e.g., question-answering systems [10] and the chatbot for book recommendation [16] that applied machine learning techniques to extract relevant entities and mapped appropriate answers to questions from data. Nevertheless, our work differs from those in the way that we developed a system to learn how to process data from a graph database and transform output results into appropriate visualizations.

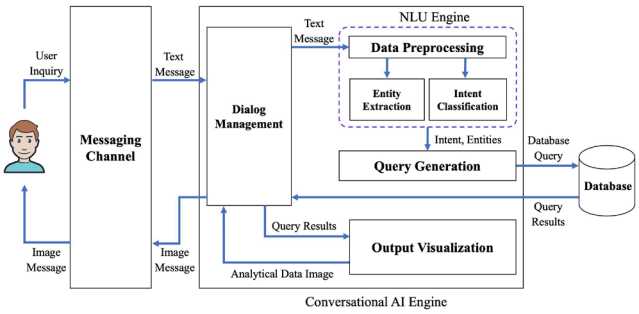


Fig. 1: Architecture of conversational agent

III. Chatbot Design and Development

Fig. 1 illustrates the architecture of our conversational agent. The chatbot consists of two essential modules, i.e., messaging channel and conversational AI engine, which are described in this section as follows.

A. Messaging Channel

The messaging channel acts as a medium for interfacing and exchanging messages between a human user and a conversational agent. In our study, LINE messaging platform[6] is chosen to be the messaging channel. The platform is connected to our conversational AI engine through its messaging API[7]. When a message (in JSON format) is sent over HTTPS protocol to LINE messaging platform, a webhook event is triggered to push the message to our conversational engine. After processing, the engine then sends the output back to the user by calling LINE platform API.

B. Conversational AI Engine

This module is the core component of our chatbot to process an input, recognize a user request, query a database, and then generate an output in a specific format. The module consists of four sub-modules, i.e., natural language understanding (NLU) engine, dialog management, query generation, and output visualization.

1) NLU Engine: Our chatbot agent uses NLU engine to understand a text message from a user. The engine has 3 steps to operate, which are explained as follows.

a) Data Preprocessing: An input message is converted into a suitable format for the later tasks [19]. Basically, the message is firstly separated into a list of words or tokens using spaCy tokenizer, however, it has not supported Thai language yet. So we employed a Thai pre-trained language model from fastText [20], which consisted of 300-dimension word embeddings trained using CBOW on Common Crawl and Wikipedia data, to be used in spaCy tokenizer. Next,

[6]https://line.me
[7]https://developers.line.biz/en/docs/messaging-api/overview

TABLE I: List of related works

Name of Chatbot	Domain	Year	Objective	Output Data Type
Bieliauskas et al. [8]	Human-Computer Interaction (HCI)	2017	To interact with a visualization software and illustrate specific visualized data from a conversation with a human	Image
Choi et al. [9]	Electric Appliance	2017	To interactively instruct users how to use complicate functions of some appliances	Text, Image
Decha et al. [10]	Question Answering	2017	To develop a Thai question answering system to receive an input question, analyze an expected type of answer, and search for all possible answers from source documents	Text
AgronomoBot [11]	Agriculture	2018	To search and display the data acquired from a Wireless Sensor Network in farm management systems	Text, Image, File (PDF)
Muangkammuen et al. [12]	Question Answering	2018	To develop a chatbot to answer FAQs, using Long Short-Term Memory (LSTM), for recognizing Thai textual questions and applying an information retrieval model for searching mostly relevant correct answers	Text
Lauren et al. [13]	Finance	2019	To provide required information for stock analysis, e.g., stock prices, historical price graphs, stock prediction, financial news, and stock sentiment	Text, Image
Robloke et al. [14]	Restaurant Booking	2019	To suggest restaurants from user specified information, e.g., food categories and locations, through automatic-developed dialog	Text
WeightMentor [15]	Healthcare	2019	To support weight-control activities of a user by giving personal feedback and motivational dialogs, or encouraging the user to submit self-assessment reports	Text, Image
Chaiwong et al. [16]	Book Recommendation	2020	To recommend books and answer questions regarding to books	Text
GameBot [17]	Sport	2020	To provide information, e.g., an individual player statistic, a team seasonal score, and a game overview, for sport fans	Text, Image

PyThaiNLP[8] is used for annotating part-of-speech (POS) tag for each token. Pre-determined patterns of regular expressions are also used for finding some entity patterns, e.g., genders and areas. Subsequenly, spaCy featurizer converts all tokens into numeric attribute vectors, which are used as features for the next two steps that learn and extract related entities and their attributes as well as classify user intentions.

b) Entity Extraction: Entities appearing in a query message generally reflect relevant pieces of information, needed to be acquired from a database. In our case, those entities along with their attributes, which appear in TPMAP, are about people and their attributes and locations. Age, gender, place name, educational level and household income are examples of attributes of interest.

Fig. 2 illustrates the entities extracted from an example of a Thai inquiry, which is translated to English as "What are the differences in the amounts of the people having different educational levels, higher than high-school, in the houses containing a person older than 30 years?". Under Rasa framework, we used Conditional Random Fields (CRFs) to recognize entities and their attributes. The parameters of CRFs were "True" for BILOU flag, the maximum iteration number of 50 for algorithm optimization, and the weight of 0.1 for L1 and L2 regularization. The features used for CRFs were "low" and "title" features for a word before and after a token, and "bias", "prefix2", "prefix5", "suffix5", "title", "digit", and "pattern" features for the token itself. We also added six additional features to a list of features for the token, i.e., "prefix3", "prefix4", "suffix4", "suffix6", "suffix7", and "part-of-speech type". Moreover, the synonyms of entities were also pre-detected and added to the entity dictionary.

Message	อยากทราบข้อมูลความแตกต่างเปรียบเทียบจำนวนคนที่จบการศึกษาสูงกว่า ม.ปลายในบ้านที่มีคนที่อายุมากกว่า 30 ปี
(The message translated to English)	What are the differences in the amounts of the people having different educational levels, higher than high-school, in the houses containing a person older than 30 years?
Entities	['house', 'person', 'condition', 'age', 'person', 'condition', 'education']

Fig. 2: Inquiry message and extracted entities

c) Intent Classification: An intent of a user, e.g., comparing or listing a series of data, is another important piece of information. Particularly, it instructs the system how to "cook" all relevant information and produce a proper response corresponding to the user intention. A linear support vector machine (SVM)[9] was applied for classifying an intent on word embeddings, generated from the terms appearing in the input message. Training parameters, i.e., types of word embeddings and a choice of word-embedding aggregation method [21], were tuned up using a grid search technique. Other SVM

[8]https://github.com/PyThaiNLP/pythainlp

[9]In this work, scikit-learn (https://scikit-learn.org/stable) was employed for SVM implementation.

parameters were set as the default values, i.e., values of 1, 2, 5, 10, 20, 100 for C parameter, 0.1 for gamma parameter, a linear kernel for the training kernel, the maximum fold number of 5 for cross validation, and weighted F1-score for evaluating the hyper parameters. For a particular message, all possible intents are ranked according their scores from the SVM classifier.

2) Dialog Management: The task of this sub-module is to ask for clarification if a message from a user containing an unclear intent or missing entities of interest. The dialog management, which is not our current focus, can be implemented using an open-source machine learning framework. By providing conversational flows for training a model and a conversation history, the model can select a suitable action with the highest calculated probability value [6].

3) Query Generation: The intents and entities obtained from the NLU engine module are fed into this module to generate a database query. Since the TPMAP data are stored in a graph database, relevant nodes and their relationships related to the input entities are considered for generating the database query. For example, Fig. 3 shows the nodes and their relationships in the database, corresponding to the input query depicted in Fig. 2 (the entities in the query highlighted in different colors). The generated query string is then used to acquire some results from the database. Fig. 4 illustrates an example of the query results.

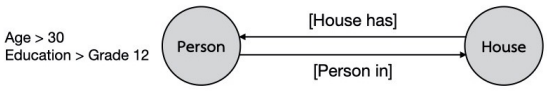


Fig. 3: Nodes and relations in TPMAP graph database related to an example of an inquiry message

Message	อยากทราบข้อมูลความแตกต่างเปรียบเทียบจำนวนคนที่จบการศึกษาสูงกว่า ม.ปลายในบ้านที่มีคนที่อายุมากกว่า 30 ปี
(The message translated to English)	What are the differences in the amounts of the people having different educational levels, higher than high-school, in the houses containing a person older than 30 years?
Intent	ask_compared_data
Entities	['house', 'person', 'condition', 'age', 'person', 'condition', 'education']
Result	[{ "education_main": "ป.ตรี หรือเทียบเท่า", "amount": 1116196 }, { "education_main": "สูงกว่าปริญญาตรี", "amount": 49963 }, { "education_main": "อนุปริญญา หรือเทียบเท่า หรือ ปวส.", "amount": 530565 }]

Fig. 4: An example of a user query with its intent, entities, and its corresponding query result visualized as a simple bar chart

4) Output Visualization: The query result from the previous module is sent to this module in order to present them in an easy-to-use format, e.g., a bar chart for data comparison. To generate graphs, we employ the Plotly[10] python library for this task. Currently, our system supports two types of bar chart plots, i.e., a simple bar chart and a grouped bar chart. An example of a simple bar chart generated by our work is shown in Fig. 4 and that of a grouped bar chart is in Fig. 5.

Message	ระหว่างคนที่เป็นหลานของหัวหน้าครัวเรือนที่จบการศึกษาต่ำกว่าประถมกับสูงกว่าประถม จะเป็นชายเทียบกับหญิงต่างกันอย่างไร
(The message translated to English)	Between the grandchildren of the household who completed lower or above elementary school, what about amount of men compared to women?
Intent	ask_compared_data
Entities	['person', 'hh_relation', 'household', 'condition', 'education', 'condition', 'education', 'gender', 'gender']
Result	[{ "education_main": "ต่ำกว่าชั้นประถมฯ (ป.4, ป.7, ป.6)", "gender_sub": "ช", "amount": 60638 }, { "education_main": "ต่ำกว่าชั้นประถมฯ (ป.4, ป.7, ป.6)", "gender_sub": "ญ", "amount": 55623 }, ...]

Fig. 5: An example of a user query with its intent, entities, and its corresponding query result visualized as a grouped bar chart

IV. Experimental Setting

A. Database

Fig. 6 shows the graph structure of TPMAP data. TPMAP data, which is stored in a graph database, consists of seven types of nodes, i.e., Person, House, Tumbol, Amphur, Province, Region, and Poverty Indicator. An individual node has its properties; edges from the node represent relations to the other entities. Data stored in TPMAP came from two sources: (1) Community Development Department of the Ministry of Interior and (2) the Ministry of Finance [3]. The data from the former source is the census-based Basic Minimum Needs (BMNs) yearly surveyed at a nationwide scale. The data covers 36.89 million individuals and 12.96 million households, distributed in 84,116 villages, 7,585 tumbols (sub-districts), 878 amphurs (districts), and 76 provinces. The data from the latter source are Ministry of Finance's welfare registration data, covering approximately 14.59 million individuals nationwide.

B. Query Message Dataset

To train and test the developed conversational agent, we generated a set of query messages relating to entities, their attributes and relations in TPMAP. Basically, a given message was generated by choosing an intent from a set of intents

[10]https://plotly.com

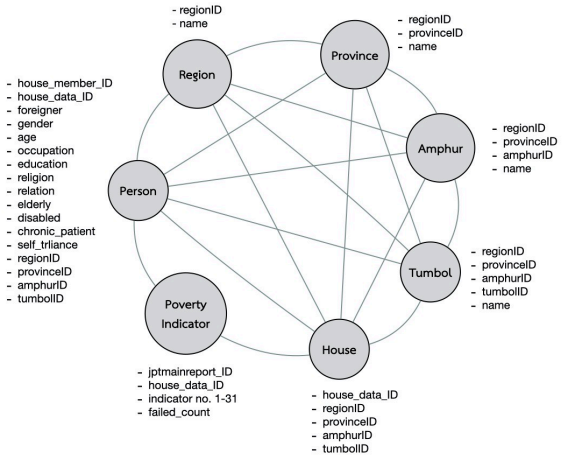


Fig. 6: TPMAP graph database

and related entities from all possible options of entities. Three types of intents relevant to database queries include listing, summarizing and comparing data. Other intents, e.g., greeting and affirmation, were added to the set in order to respond the conversation smoothly. There were approximately 15 message patterns for each intent. All combinations of relevant entities were filled into each pattern to get complete query messages, which are then used for train and test the agent. There were totally 8,970 query messages generated.

C. Evaluation Scheme and Measurement

Performance evaluation is conducted at two levels, i.e., NLU level and overall level. At the NLU level, the generated query message dataset was splited into 80% training set (7,171 messages) and 20% test set (1,799 messages) for intent classification and entity extraction. The test set contained totally 15,388 entities to be extracted. F1-score along with precision and recall were assessed as the measurements for these trials. Those measurements were calculated using the following equations.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$F1\text{-}Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

At the overall level, we tested whether the developed chatbot returned an appropriate output. In order to evaluate how well our agent performed at different levels of query complexity, the generated query messages were divided into four different levels by two factors, i.e., existence of entities related to the area nodes (e.g., Tumbol or Amphur) and existence of attributes of Person and House nodes. While the most simple messages did not contain any entities related to the area nodes or attributes of Person and House nodes, e.g., "Which amphur has the most number of residents?", the most complex messages contained both of them, e.g., "Could you show a comparison of the number of women having 34-64 year old in Amphur Phutthamonthon and Amphur Mueang Nakhon Pathom, Nakhon Pathom province with the number of those in Amphur Khuan Kalong and Amphur Thung Wa, Satun province?". An accuracy was employed as the measurement for this overall evaluation.

V. Results and Discussion

Table II shows the precision, recall, and F1-score for intent classification and entity extraction. All values are the macro average values among all intent and entity classes. The developed NLU module achieves F1-score of 0.9747 for intent classification and 0.7163 for entity extraction. With a correct intent and complete entity retrieval, the developed chatbot responds to a user inquiry correctly.

TABLE II: Performance evaluation of intent classification and entity extraction on NLU level

Process	**Measurement**		
	Precision	**Recall**	**F1-Score**
Intent Classification	0.9697	0.9824	0.9747
Entity Extraction	0.8108	0.7134	0.7163

Table III shows accuracy obtained from the evaluation at the overall level, i.e., the process from retrieving an inquiry message until replying a final output to a user. The developed chatbot yields accuracy of 57.5%-80.0% according to the complexity of input messages. Compared to the performance on the most simple cases, the chatbot yields lower accuracy of 5% on inquiry messages having entities related to the area nodes, and 17.5% on those having attributes of Person and House nodes. Thus, the complexity of query messages, in terms of existence of entities related to the area nodes and attributes of the Person or House nodes, obviously affects the overall performance of the developed conversational agent.

TABLE III: Accuracy from evaluation on overall level separated by complexity of inquiry messages

Factors of message complexity		**Attributes of Person/House**	
		not existing	existing
Entities related to area nodes	not existing	80.0%	62.5%
	existing	75.0%	57.5%

VI. Conclusion

In this paper, a Thai conversational agent is developed on top of TPMAP to support self-service data analytics by providing information in easy-to-use formats corresponding with natural language inquiries from users. Rasa, an open-source conversational AI engine, is employed to implement the conversational agent. An evaluation of the NLU module shows

that our developed agent performs intent classification with F1-score of 0.9747 and entity extraction with F1-score of 0.7163. For an evaluation of the overall process, the chatbot yields accuracy from 57.5% to 80% according to the complexity of query messages. If a correct intent is classified and all entities are retrieved, the chatbot can get all relevant data from the TPMAP graph database and then return the information, in form of either a text message or a bar chart, to the user correctly.

ACKNOWLEDGMENT

The first author was financially supported by Young Scientist and Technologist Program (YSTP), which is a scholarship awarded by National Science and Technology Development Agency (NSTDA). This work was also partly supported by Thai People Map and Analytics Platform (TPMAP), a joint project between the office of National Economic and Social Development Council (NESDC) and National Electronics and Computer Technology Center (NECTEC). The authors wish to thank Natthapon Chinarangsi for graph database implementation.

REFERENCES

[1] R. Joseph. (2019) Big data analytics in government: How the public sector leverages data insights. Accessed on: 15 July 2020. [Online]. Available: https://www.intellectyx.com/blog/big-data-analytics-in-government-how-the-public-sector-leverages-data-insights/

[2] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Communications of the ACM*, vol. 57, no. 3, pp. 78–85, 2014.

[3] N. Surasvadi, P. Ruchikachorn, C. Siripanpornchana, S. Thajchayapong, and A. Plangprasopchok, "Tpmap: a data analytics and visualization platform to support thailand target poverty alleviation programs," in *2019 IEEE Pacific Visualization Symposium (PacificVis)*. Los Alamitos, CA, USA: IEEE Computer Society, apr 2019, pp. 325–326.

[4] M. Carisi, A. Albarelli, and F. L. Luccio, "Design and implementation of an airport chatbot," in *Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good*, 2019, pp. 49–54.

[5] F. P. Putri, H. Meidia, and D. Gunawan, "Designing intelligent personalized chatbot for hotel services," in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, ser. ACAI 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 468–472.

[6] A. Harshita, F. Robin, Becker snd Mehnoor, and R. Lucian, "Development of an artificial conversation entity for continuous learning and adaptation to user's preferences and behavior," Technical University of Munich, p. 13, 2019.

[7] T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol, "Rasa: Open source language understanding and dialogue management," *arXiv preprint arXiv:1712.05181*, 2017.

[8] S. Bieliauskas and A. Schreiber, "A conversational user interface for software visualization," in *2017 ieee working conference on software visualization (vissoft)*. IEEE, 2017, pp. 139–143.

[9] H. Choi, T. Hamanaka, and K. Matsui, "Design and implementation of interactive product manual system using chatbot and sensed data," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2017, pp. 1–5.

[10] H. Decha and K. Patanukhom, "Development of thai question answering system," in *Proceedings of the 3rd International Conference on Communication and Information Processing*, 2017, pp. 124–128.

[11] G. M. Mostaco, I. R. C. De Souza, L. B. Campos, and C. E. Cugnasca, "Agronomobot: a smart answering chatbot applied to agricultural sensor networks," in *14th international conference on precision agriculture*, vol. 24, 2018, pp. 1–13.

[12] P. Muangkammuen, N. Intiruk, and K. R. Saikaew, "Automated thai-faq chatbot using rnn-lstm," in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2018, pp. 1–4.

[13] P. Lauren and P. Watta, "A conversational user interface for stock analysis," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5298–5305.

[14] R. Robloke and B. Kijsirikul, "A task-oriented dialogue bot using long short-term memory with attention for thai language," in *Proceedings of the International Conference on Advanced Information Science and System*, 2019, pp. 1–6.

[15] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. McTear, "Weightmentor, bespoke chatbot for weight loss maintenance: Needs assessment & development," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 2845–2851.

[16] J. Chaiwong, N. Prajugjit, and K. Sookhanaphibarn, "Book recommendation website with chatbot," in *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 2020, pp. 193–195.

[17] Q. Zhi and R. Metoyer, "Gamebot: A visualization-augmented chatbot for sports game," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–7.

[18] M. Hearst and M. Tory, "Would you like a chart with that? incorporating visualizations into conversational interfaces," in *IEEE Visualization Conference (VIS)*. IEEE, 2019, pp. 1–5.

[19] B. Ravi. (2018) Nlp behind chatbots. Accessed on: 15 July 2020. [Online]. Available: https://medium.com/bhavaniravi/deploying-a-chatbot-nlp-model-demystifying-rasanlu-2-server-7704afc74d1f

[20] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[21] J. Schuurmans and F. Frasincar, "Intent classification for dialogue utterances," *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 82–88, 2019.

Real-life Human Activity Recognition with Tri-axial Accelerometer Data from Smartphone using Hybrid Long Short-Term Memory Networks

Narit Hnoohom[1], Anuchit Jitpattanakul[2] and Sakorn Mekruksavanich[3]
[1]*Image, Information and Intelligence Laboratory, Department of Computer Engineering*
Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand
narit.hno@mahidol.ac.th
[2]*Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics*
Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th
[3]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th

***Abstract*—Human activity recognition (HAR) has an enthusiastic research field in time-series classification due to its variation of successful applications in various domains. The availability of affordable wearable devices have provided many challenging and interesting research HAR problems. Current researches suggest that deep learning approaches are suited to automated feature extraction from raw sensor data, instead of conventional machine learning approaches that reply on hand-crafted features. Based on the recent success of Long Short-Term Memory (LSTM) networks for HAR domains, this work proposes a generic framework for accelerometer data based on LSTM networks for real-life HAR. Four hybrid LSTM networks have been comparatively studied on a public available real-life HAR dataset. Moreover, we take advantage of Bayesian optimization techniques for tuning hyperparameter of each LSTM networks. The experimental results indicate that the CNN-LSTM network surpasses other hybrid LSTM networks.**

***Keywords*—LSTM, human activity recognition, deep learning, time-series data, smartphone, real-life activity**

I. Introduction

Since wearable equipment and portable computing devices are affordable, there is a large amount of collected data comprising of locations, motions, human movements' physiological signals of human movement and environmental information. A research field, human activity recognition (HAR), aims at studying the developments of human behaviors by understanding the attributes obtained from this data. Smartphones that are new technologies have many additional features comprising of multitasking and various sensors. The mobile devices are being integrated in daily lives quickly. It is also expected that the devices can efficiently track people's activities and help them making their decisions. Data are massively collected from the people using the wearable devices because of the inertial sensors embedded in the devices. Physiological signals data including sensor signals from accelerometer, heart rate monitors, thermometers and gyroscopes are collected.

Most smartphones have accelerometers. The accelerometers are sensors measuring objects' accelerations along referred axes. The sensors can effectively monitor the actions involving repetitive motions; for example, standing, walking, running, sitting and climbing. Data obtained from the sensors may be processed in order to identify movements' sudden changes. Gyroscopes are other standard sensors for smartphones. These sensors can measure orientation with gravity. Signals from the sensors can be processed in order to detect the devices' positions and alignments [1].

Human Activity Recognition (HAR) has attracted the machine learning community [2] due to availability of real-word applications such as fall detection in elderly healthcare monitoring [3], exercise monitoring and tracking in sport science [4] and preventing office work syndrome [5]. Currently, HAR become one of the challenging research topics due to availability of sensors in wearable devices (e.g., smartphone, smart watch, and etc.) that are low cost and less power consumption, including live streaming of time-series data 6. Recent researches in HAR, both dynamic and static human activities, are studied by sensor data collected from wearable devices to achieve better understanding of relationship between health and behavioral biometric information. The HAR approaches can be categorized broadly into two categories according to data sources that are visual-based recognition and sensor-based recognition. In visual-based HAR, video/image data are recorded and processed with computer vision techniques [7]. The second HAR category is the sensor-based HAR that works on time-series data captured from a wide of various sensors embedded in wearable devices [8]. Due to the advancement of sensor technology and pervasive computing, sensor-based HAR is becoming more widely used in smart devices as smartphones with privacy well protected [9]. Therefore, this work focuses on the smartphone sensor-based human activity recognition.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

Smartphones have various sensors (e.g. accelerometers, gyroscopes Bluetooth and ambient sensors) that are used for recognizing human activities. The sensor-based HAR system of the smartphones is a machine learning (ML) model that can continually monitor users' activities while the smartphones are attached to the users' bodies. The conventional approaches for recognizing the activities adopt machine learning algorithms (e.g. decision tree, naïve Bayes, support vector machine, and artificial neural network) [10]. Unfortunately, these approaches have limitations caused by human domain knowledge. Therefore, the approaches have limited performances in the terms of classification accuracy and other evaluation metrics.

Deep learning (DL) approaches are used in order to overcome the mentioned limitations in this work. The features of raw data can be learned automatically by using more than one hidden layer instead of manually extracting the data according to the human domain knowledge. These approaches' deep architecture can extract high-level deep features that are appropriate for complicated problems (e.g. HAR). A smartphone-based HAR is build by employing the deep learning approaches [11], [12].

So, this work studies LSTM-based human activity recognition with sensor data of smartphone. Four hybrid LSTM networks have been comparatively researched to impact of using different kinds of smartphone sensor data from a public dataset called "a public domain dataset for real-life human activity recognition using smartphone sensors" or shortly called RL-HAR dataset. Moreover, we use Bayesian optimization to config hyperparameters of the LSTM models.

The structure paper is organized as follows. Section II provides foundations and background theory used in this study. Section III proposes the proposed HAR framework for smartphone sensor data. The experimental setting and conducted results show in section IV. Finally, Section V concludes the results.

II. Theoretical Background

A. Human Activity Recognition from Sensor Data

HAR focuses to explain human behaviors which enable the computing systems to proactively support users based on their requirement [13]. Human activities, such as walking, sitting, working, and running, can be defined as a set of actions performed by the user over a period in a given protocol. Formally, suppose a user is operating some types of activities applying to a predefined activity set *A* [14]:

$$A = \{a_1, a_2, a_3, ..., a_m\} \tag{1}$$

where *m* stands for the number of activity categories. There also is a sequence of sensor data (*s*) gathering that apprehends the activity information.

$$s = \{d_1, d_2, ..., d_t, ..., d_n\} \tag{2}$$

where d_t stands for the sensor data reading at time *t* and *n* stands for any sequences and $n \geq m$.

The HAR assignment is to build a recognition function *F* to predict the activity sequence depended on sensor data reading *s*. The function *F* can be determined as:

$$F(s) = \{a_1^{'}, a_2^{'}, a_3^{'}, ..., a_n^{'}\}, a_n^{'} \in A \tag{3}$$

While the real activity sequence is expressed as:

$$F(s) = \{a_1^{*}, a_2^{*}, a_3^{*}, ..., a_n^{*}\}, a_n^{*} \in A \tag{4}$$

Commonly, implementing such as a HAR system is operated in five fundamental steps: Data collection, Segmentation, Feature extraction, Model training, and Classification as shown in Fig. 1.

Deep learning tends to overcome the limitation of feature extraction problem in conventional machine learning [15]. Deep learning performances for HAR with different kinds of networks can be shown in Fig. 2. Feature extraction and model training processes are simultaneously operated in the deep learning approach. The features can be automatically learned through the network instead of being manually hand-crafted as conventional machine learning approaches [16].

III. Proposed Methodology

The framework LSTM-based human activity recognition framework, which is proposed in this work, enables the sensor data captured from smartphone sensor to classify the activity that the smartphone user has performed. Figure 3 illustrates the overall methodology used in this study to achieve the research goal.

A. RL-HAR Smartphone Dataset

The activity public dataset used for the proposed framework is RL-HAR [17]. The RL-HAR dataset contains activity information that was collected from 19 participants of different ages, genders, heights and weights. The subjects were doing real-life activities by carrying a smartphone. There are 4 activities performed by each person (i.e. Active, Inactive Walking, and Driving). The smartphone integrated tri-axial accelerometer, gyroscope, magnetometer and GPS was used to record sensor data while each of the subjects was performing the four real life activities. Tri-axial linear acceleration data were acquired at different rates. Some samples are shown in Fig. 4. The available dataset contains 17,378,634 samples. This dataset is an imbalanced dataset as shown in Fig. 5.

B. LSTM Architectures

The LSTM networks' architectures differently used in this paper include Vanilla LSTM, Stacked LSTM, Bidir-LSTM, and CNN-LSTM. The Vanilla or original LSTM model has a single hidden LSTM layer and a standard feedforward output layer [18]. This model has the Stacked LSTM as an extension having multiple hidden LSTM layers containing multiple memory cells. The stacked LSTM hidden layers increase the depth and accuracy of the model [19]. Currently, the stacked LSTMs are a stable technique for solving sequence prediction problems. A stacked LSTM architecture can be considered as an LSTM model having multiple LSTM layers. Each LSTM

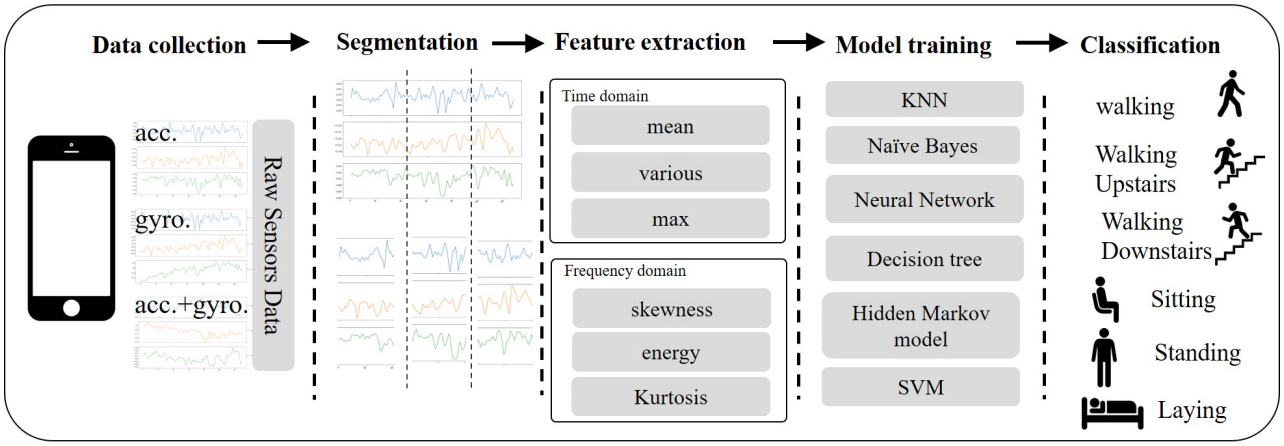


Fig. 1. Sensor-based HAR using conventaional ML approaches

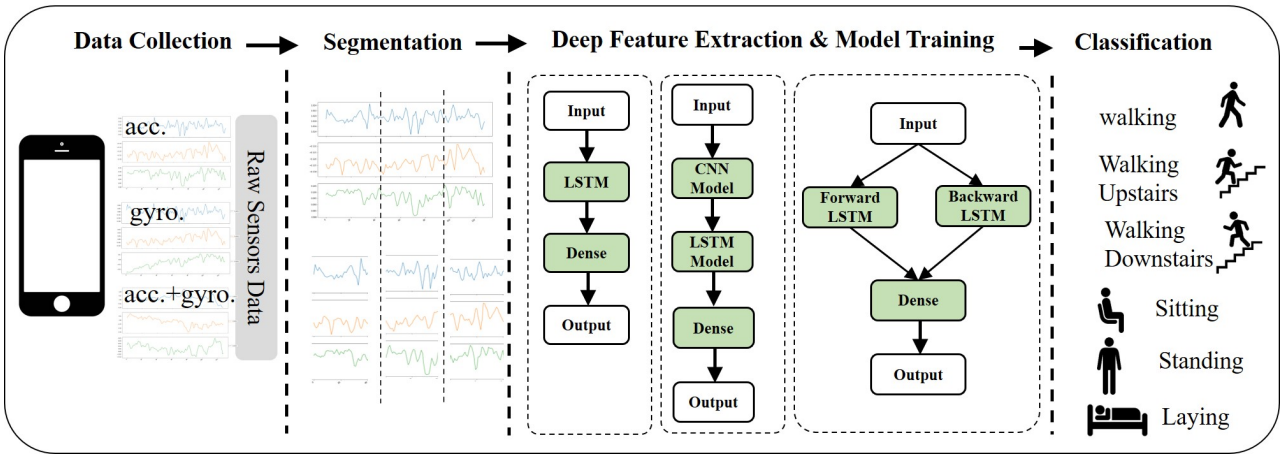


Fig. 2. Sensor-based HAR using conventaional DL approaches

layer has a sequence output instead of a single value output to the LSTM layer below. Instead of one output time step for all input time steps, it is one output per input time step.

The CNN-LSTM architecture utilizes Convolutional Neural Network (CNN) layers for extracting features on input data combined with LSTMs in order to improve sequence prediction [20]. CNN-LSTMs were developed in order to predict visual time series and describe sequences of images such as videos. It is suitable for solving problems having temporal structures in inputs; for example, the orders of images in videos or words in texts, or requiring generated output with temporal structures such as words in textual descriptions. A LSTM network that is called CNN-LSTM is presented in this paper in order to improve the performance of the recognition.

C. Tuning Hyperparameter by Bayesian Optimization

For deep learning algorithms, hyperparameters are important because these can directly control the behaviors of training algorithms. These also significantly affect the deep learning models' performances.

To solve functions that are computationally expensive to find the extrema, Bayesian optimization is an effective method because it can solve functions without closed-form expressions and computationally expensive functions, derivatives difficult to be evaluated, or non- convex functions. The optimization goal is to find the maximum value at the sampling point for an unknown function f in this paper.

$$x^{+} = \underset{x \in A}{\operatorname{argmin}} f(x) \tag{5}$$

where A stand for the search space of x.

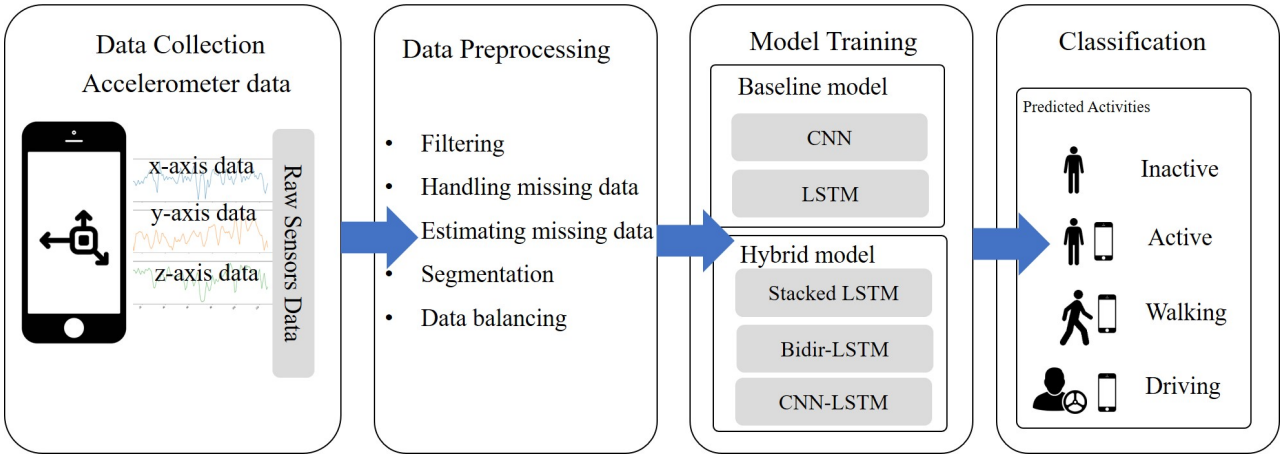
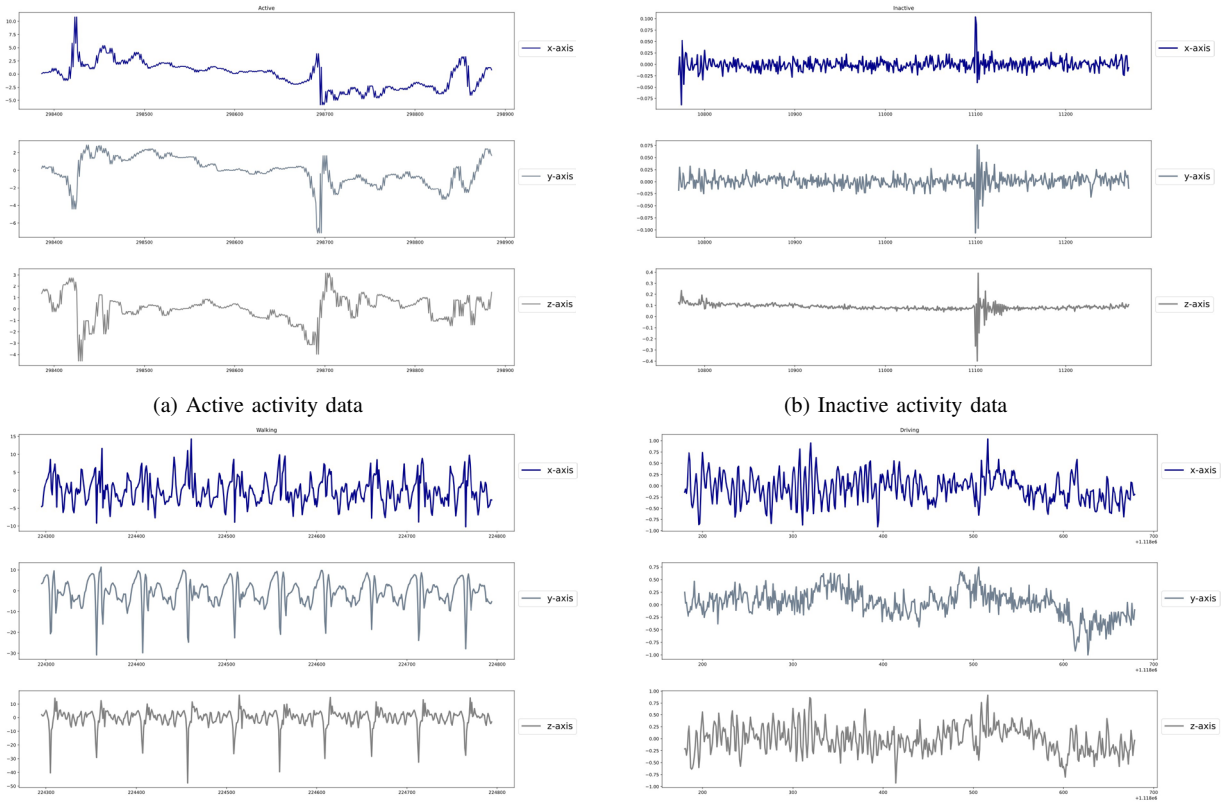


Fig. 3. The proposed framework of LSTM-based HAR



(a) Active activity data

(b) Inactive activity data

(c) Walking activity data

(d) Driving activity data

Fig. 4. Some samples from RL-HAR of activity data (a) Active (b) Inactive (c) Walking and (d) Driving

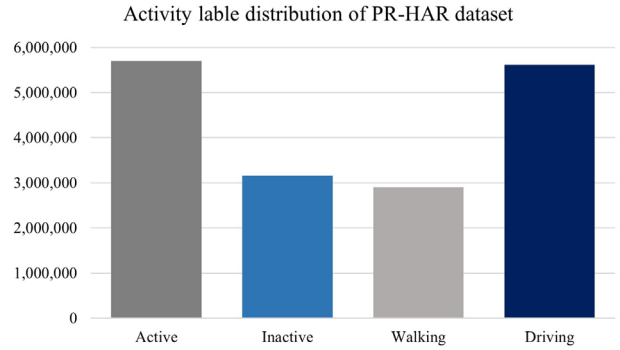


Fig. 5. Activity lable distribution of PR-HAR dataset

IV. Experiments and Results

In this section, we describe experiment setting and experimental results used to evaluate the LSTM networks for human activity recognition.

A. Experiments

To compare the performance of each LSTM networks for the HAR problem, we used variations in the experiments. The first LSTM network, we used a basic LSTM network called Vanilla LSTM that composed of only one LSTM hidden layer working with Dropout and one Dense layer. The second variation, we added one more LSTM hidden layer. This kind of this configuration is called Stacked LSTM. In the same way of addition a LSTM hidden layer, the third variation is Bidir-LSTM. The fourth variation is CNN-LSTM that is a LSTM network combining a convolution layer to a LSTM layer. The summary of these LSTM network architectures is shown in Table I.

B. Experimental Results

This section presents the results obtained by the experiments. The experimental hardware and software configuration were as follows. The LSTM-based HAR models in Table I were implemented by using Python's Scikit-learn, TensorFlow [21] and Keras [22] for each tests consisted of the different LSTM models. All tests were executed on Google Colab platform. We also optimized the LSTM networks hyperparameters by the Bayesian hyperparameter optimization platform by SigOpt [23]. This platform is the standardized and scalable platform that are accompanied by the API generation facilitating of model performances. It also allows the state of being parallelized operations for faster evaluation.

The UCI-HAR dataset was divided into 70% to use as training data and remaining 30% for testing data. This work conducts several experiments for evaluating the recognition performance of the LSTM networks with a variation of metrics such as Accuracy, Precision, Recall, and F1-score. Table II shows the accuracy values obtained from the various LSTM networks which are trained on the RL-HAR dataset.

It can be indicated that the CNN-LSTM networks outperforms all the other networks with 93.51% accuracy as shown in Fig. 6. The performance of all the hybrid LSTM networks are better that baseline LSTM. The confusion matrix for the hybrid LSTM networks is shown in Fig. 6.

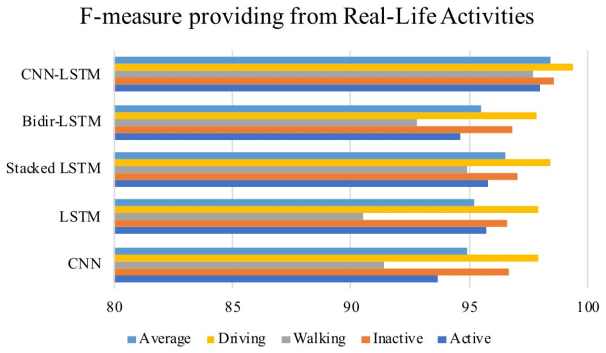


Fig. 6. Accuracy of the different LSTM networks on the RL-HAR dataset

C. Comparative analysis with previous works

The comparison of the accuracy of our proposed model with other LSTM networks is shown in Table III. It is clear that the proposed CNN-LSTM networks outperforms the other previous works. The main reason behind achieving better accuracy is that the spatial feature extraction by layer convolutional neural network that improve the overall accuracy up to 3.20% comparing to the latest work [24].

V. Conclusion and Future Works

In this study, we proposed a framework of LSTM-based human activity recognition that explores the LSTM network worked with high performance for HAR problem. We elected four types of LSTM networks to comparatively study on their recognition performance with a smartphone sensors i.e., tri-axial accelerometer. We evaluated these LSTM networks to a public available dataset called RL-HAR by regarding predictive accuracy and other performance metrics including precision, recall, and F1-score. The experimental results show that the CNN-LSTM networks proposed in this work outperforms the other baseline LSTM networks with high accuracy of 98.51%. Moreover, we also compared the proposed LSTM network to previous work. The CNN-LSTM networks can improve the accuracy up to 3.20%.

For future direction, developing the LSTM models further and evaluating the model with different hyper parameters including the learning rate, batch size, regularization and others will be investigated. Moreover, we arrange to apply the proposed model to more complex activities for tackling other challenges in deep learning and HAR by evaluating it on other public activity datasets, such as OPPORTUNITY and PAMAP2 datasets.

TABLE I. Different architectures of LSTM networks used in this research

DL Models	Layer Types
CNN	Conv1D-Dropout-Maxpooling-Dense-Softmax
LSTM	Lstm-Dropout-Dense-Softmax
Stacked LSTM	Lstm-Dropout-Lstm-Dropout-Dense-Softmax
Bidir-LSTM	BiLstm-Dropout-Dense-Softmax
CNN-LSTM	Conv1D-Conv1D-Dropout-Maxpooling-Lstm-Dropout-Dense-Softmax

TABLE II. Performance metrics of the five LSTM models

	Evaluation metrics			
Network	**Accuracy**	**Precision**	**Recall**	**F1-score**
CNN	95.22	95.40	95.20	95.20
LSTM	95.71	95.70	95.70	95.70
Stacked LSTM	96.69	96.70	96.70	96.70
Bidir_LSTM	95.71	95.70	95.70	95.70
CNN-LSTM	98.51	98.50	98.50	98.50

TABLE III. Performance comparison with the previous work

Ref.	**Method**	**Sensors**	**Accuracy**
Garcia	SVM	Acc.+GPS.	60.10 ±11.43
Gonzalez	SVM	Acc.+Magn.+GPS.	62.66 ±11. 68
[24]	SVM	Acc.+Gyro.Magn.+GPS.	67.22 ±13.13
Proposed method	CNN-LSTM	Only Acc.	98.51

ACKNOWLEDGMENT

This work was supported by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, under Grant No. 642072 and supported by the Department of Computer Engineering, Faculty of Engineering at Mahidol University. Authors also thank the SigOpt team for the provided optimization services.

REFERENCES

[1] S. Mekruksavanich and A. Jitpattanakul, "Classification of gait pattern with wearable sensing data," in *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*, 2019, pp. 137–141.

[2] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698 – 703, 2019.

[3] N. Hnoohom, A. Jitpattanakul, P. Inluergsri, P. Wongbudsri, and W. Ployput, "Multi-sensor-based fall detection and activity daily living classification by using ensemble learning," in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, 2018, pp. 111–115.

[4] S. Mekruksavanich and A. Jitpattanakul, "Exercise activity recognition with surface electromyography sensor using machine learning approach," in *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT NCON)*, 2020, pp. 75–78.

[5] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "Smartwatch-based sitting detection with human activity recognition for office workers syndrome," in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, 2018, pp. 160–164.

[6] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," vol. 6669, 06 2011, pp. 289–296.

[7] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A review on human activity recognition using vision-based method," *Journal of Healthcare Engineering*, vol. 2017, pp. 1–31, 07 2017.

[8] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Human activity recognition using triaxial acceleration data from smartphone and ensemble learning," in *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, 2017, pp. 408–412.

[9] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. Yupapin, "Enhanced hand-oriented activity recognition based on smartwatch sensor data using lstms," *Symmetry*, vol. 12, no. 9, 2020.

[10] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition-a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1254, 03 2018.

[11] B. Almaslukh, J. Muhtadi, and A. Artoli, "A robust convolutional neural network for online smartphone-based human activity recognition," *Journal of Intelligent and Fuzzy Systems*, vol. 35, pp. 1–12, 06 2018.

[12] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915 – 922, 2018.

[13] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys*, vol. 46, 01 2013.

[14] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3 – 11, 2019.

[15] A. Baldominos, A. Cervantes, Y. Saez, and P. Isasi, "A comparison of machine learning and deep learning techniques for activity recognition using mobile devices," *Sensors*, vol. 19, no. 3, 2019.

[16] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," vol. 3001, 04 2004, pp. 1–17.

[17] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[19] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked lstm network for human activity recognition using smartphone data," in *2019 8th European Workshop on Visual Information Processing (EUVIP)*, 2019, pp. 175–180.

[20] R. Mutegeki and D. S. Han, "A cnn-lstm approach to human activity recognition," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 362–366.

[21] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.

[22] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[23] I. Dewancker, M. McCourt, S. Clark, P. Hayes, A. Johnson, and G. Ke, "A stratified analysis of bayesian optimization methods," 03 2016.

[24] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "A public domain dataset for real-life human activity recognition using smartphone sensors," *Sensors*, vol. 20, no. 8, 2020.

Can Tweets predict ICO success? Sentiment Analysis for Success of ICO Whitepaper: evidence from Australia and Singapore Markets

Anchaya Chursook
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
anchaya_c@cmu.ac.th

Nathee Naktnasukanjn
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
nathee@cmuic.net

Somsak Chaimaim
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
somsak.c@cmuic.net

Piyachat Udomwong
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
piyachat.u@cmu.ac.th

Jutharut Chatsirikul
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
jutharut.c@cmuic.net

Nopasit Chakpitak
International College of Digital Innovation
Chiang Mai University
Chiang Mai, Thailand
nopasit@cmuic.net

***Abstract*—Initial coin offerings (ICOs) are a new fundraising method for businesses, companies or entrepreneurs through a smart contract. The ICO fundraising model was one of the most popular FinTechs from 2016 to 2018, as the most supportive tool for startups in countries with limitations on initial public offerings (IPOs). Among ICOs ended in 2018, the leading sectors of funds raised included cryptocurrencies, business platforms, and business service projects. An ICO whitepaper contains primary business data that can assist in decision-making for investment. Market sentiment was used to predict the success of ICOs based in Australia and Singapore markets. Results showed that sentiment analysis on tweets can be used to predict ICO success.**

Keywords—Sentiment Analysis, Initial Coin Offerings, ICO Whitepaper, FinTech, Tweets

I. INTRODUCTION

Financial technology or FinTech has played an important role in the business world. FinTech is used to create a competitive advantage or add value and has resulted in a high degree of disruption to virtually every area of business. FinTech has also played a crucial role in reducing the cost of financial transactions by giving people more opportunities to access standard financial services. Crowd-funding is a method of obtaining funding from the general public through an online platform or website. Any individual or organization with innovative ideas can raise funds in this manner using social networks. An initial coin offering (ICO) is a new fundraising method that relies on technology. The company or business offering an ICO raises funds from investors via an exchange of tokens or assets. ICOs have raised very substantial amounts of funds, particularly in 2017 and 2018. A total of 3,782 ICOs were opened in 2018 and these raised almost $11.4 billion [1].

Entrepreneurs choose ICOs as a quick, inexpensive, and easily accessible method to raise funds by creating awareness among global investors via social media. By contrast, ICOs have an inherent risk for investors since some businesses are startups and their only action plans are the text presented as whitepapers. Therefore, comments pertaining to ICO whitepapers on websites or social media are crucial for investors to consider when making investment decisions.

II. RELATED LITERATURE

A. Initial Coin Offerings (ICOs)

Initial coin offerings (ICOs) as token sales or crowd sales are a fundraising method for businesses, companies or entrepreneurs through a smart contract. ICOs usually relate to blockchain technology and cryptocurrency. Entrepreneurs choose ICOs for fundraising as a quick, inexpensive, and easily accessible method to raise funds. Social media tools are used to create awareness among global investors. The ICO issuers use blockchain technology to create new cryptocurrencies and then sell these new digital tokens, or right of the products or services being developed by the ventures, to investors as collateral to support their projects [2]. Investors who are interested in investing in the business can buy the offering and receive the new cryptocurrency tokens issued by the company. These tokens may have some utility in using the product or service offered by the company, or may just represent a stake in the company or project. In 2017, average funding collected by ICOs was $24.35 million, while in 2018, the average fund amount was $11.52 million [1]. Asia has been a substantial area of investment in ICOs, with Singapore, Hong Kong, and China as the top three countries with economies based on the number of ICOs. In 2019, the trend of using ICOs decreased in the top three global ICO markets of the United States, Singapore, and England. Singapore was rated the most competitive economy in the world, with a score of 84.8 out of a possible 100, based on the World Economic Forum in 2019 [3]. The Government of Singapore oversees regulations and promotes the creation of innovative ecosystems for entrepreneurs through investment support and legislation. The Monetary Authority of Singapore (MAS) is an integrated regulator and supervises financial institutions in Singapore, which has now become the Fintech center in Asia. The FinTech regulatory sandbox provided by the MAS is a safe testing space where financial institutions and other

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

financial service providers can test their new business models without immediately incurring all the normal regulatory consequences. In Australia, Fintech has been significantly developed, with strong support from the government. Australia is an attractive global destination for Fintech investment because the country has a mature, diverse, and internationally connected ecosystem.

B. ICO Whitepaper

ICO whitepapers provide detail and clarity as information sources of the business opportunities offered to ICO investors [4]. An ICO whitepaper presents overall public information of the business and explains the business model, blockchain platform, token sale supply, token sale structure, token distribution, use of funds, team, roadmap, etc. [5]. A prospective entrepreneur sets up and publishes a business canvas or a business model, citing the necessary information of the project on the whitepaper [6]. An ICO whitepaper is a pre-announcement of the project. Popular social media sites that ICOs use for broadcasting the announcement are Telegram, Twitter, Facebook, Reddit, Bitcointalk, Medium, and YouTube [7][8].

An ICO whitepaper is an in-depth report that discloses the company project and credibility, especially on topics that present problems, and also proposes solutions. The whitepaper is an effective way to advertise the quality of the ICO project [4]. More than 80.6% of investors read the ICO whitepaper and attempt to understand its content before investing [8]. An ICO whitepaper helps to protect the interests of ICO investors and ensure the long-term viability of the ICO [2]. Thus, the ICO whitepaper is an essential document, and if adequately written, the ICO project will attract many participants and contributors. Investment in an ICO has a high chance of profitable return if the business is successful; however, the investment also has high risk and the business may be unsuccessful [9] resulting in monetary loss. Consequently, prospective investors should carefully study and understand the investment model before making the decision to invest.

C. Sentiment Analysis

Twitter is a social networking application and a popular information source. Tweets on the Twitter platform can be used to assess sentiment analysis for prediction of market success. Search trends and total number of tweets are both positive signs of fundraising volume. The data of startups on social media and ICO rating websites can be used to determine the success of the ICOs. The relationship between funds raised by startups and Twitter sentiment shows increasing emotionality in tweets toward the success of the ICO [10].

Sentiment analysis is defined as a type of natural language processing (NLP) for tracking public mood that automatically identifies sentiments within the text. Tweets on Twitter are used toward sentiment analysis [10-13]. Sentiment analysis allows businesses to understand the emotions of their customers through social media, customer reviews and surveys [11-13].

Sentiment analysis is the most common text classification tool used for analyzing an incoming message and deciding whether the underlying sentiment is positive, negative or neutral toward an entity such as business products and services, organizations, and experience [14]. Sentiment analysis can provide insights into market trends, thereby creating new opportunities for business growth and product development. Communication of company announcements via Twitter attracts investors' attention and decreases information asymmetry [15]. Social media platforms such as Twitter, Facebook and YouTube are important channels for information dissemination into new markets [16]. Nowadays, big data and NLP approaches are implemented to evaluate social media sentiments and assess the financial success of ICOs.

III. RESEARCH METHODOLOGY

A. Data source and method

Data collection was performed following the steps illustrated in Figure 1.

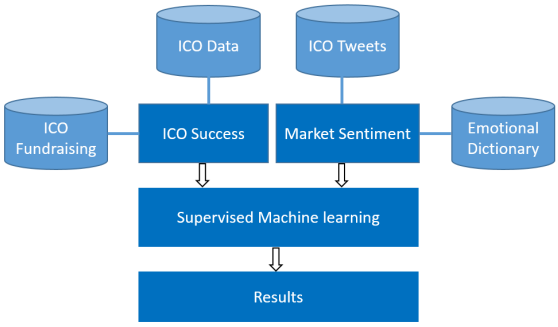


Figure 1. Research Process

First, a dataset of ICOs was extracted from the website www.ico.tokens-economy.com in December 2019. ICOs listed in sectors of business service, cryptocurrency, energy, entertainment, and platform were collected between September 2016 and June 2020, with 185 out of 237 ICOs assembled as datasets in Table I.

TABLE I. NUMBER OF ICOS COLLECTED FROM DATA SAMPLING IN AUSTRALIA AND SINGAPORE BY INDUSTRY SECTOR

Industry sector	Australia		Singapore	
	Number of ICOs	***Data sampling***	***Number of ICOs***	***Data sampling***
Business service	7	5	41	30
Cryptocurrency	19	17	74	53
Energy	2	2	2	2
Entertainment	4	4	12	11
Platform	12	12	64	49
Total	**44**	**40**	**193**	**145**

Second, demographic data of the ICOs used in the study were collected from the website www.ico.tokens-economy.com. The data included name, ticker, sector and country. Additional information was also collated consisting of overviews and related data collected from ICO whitepapers, documents, and company news from www.icobench.io website. ICO data including Hard Cap, Soft Cap, Raised Fund, Type of Platform, Bonus, Coin Accepting, Token Supply, start and end date, ICO Rating and Use of Fund were considered as basic information. This approach enabled a greater understanding of the overall environment that affected market sentiment of ICOs. Total raised funds in comparison to the soft cap were used to define the success or failure of the ICO. The soft cap is the minimum capital amount raised by a crowdsale campaign to

start the project. An ICO that raises funds in excess of its soft cap will be considered as successful. In other words, failure is defined as an ICO that cannot raise sufficient funds to exceed its soft cap. This can be represented as:

$$X = F - S ; X \geq 0 \quad (1)$$

where;

X is the Successful ICO

F is the amount of raised funds

S is the ICO soft cap

To determine successful ICOs, the amount of raised funds is illustrated in equation (1). If the ICO can raise funds (F) of more than its soft cap (S), then the successful ICO (X) is equal to or greater than 0. In this step, we found that 95 out of 185 ICOs raised funds of more than their soft cap. The ICO from the energy sector was removed from the analysis because there was only one successful ICO, and this was not comparable.

Subsequently, Twitter data of 95 ICOs were retrieved and collected using Scrapy, a python scraping framework. Twitter data retrieval was conducted using a keyword search as a short name of ICO. Then, "$" was placed as a prefix to indicate the currency of each token and reduced the number of irrelevant data retrievals. Eight ICOs were excluded from the analysis because of no scraping results from Twitter. Hence, 87 ICOs were included for further analysis. The 87 ICOs comprised 261,975 tweets. Each tweet consisted of (i) ID, (ii) username, (iii) tweet, (iv) URL, (v) number of retweets, (vi) number of favorites, (vii) number of replies, and (viii) date and time.

Sentiment analysis of the ICOs was then implemented on the R program using the "lexicon" package. This package supports sentiment analysis based on the NRC Emotion Lexicon, which is an emotion dictionary. Table II shows an example of a collected tweet in a dataset. This can be described as a user ID (ID) 1165758049073737729 whose name (username) as Bobcat Crypto tweeted (tweet) that he/she had detected a potential MEDIUM strength BUY Signal for $HBZ (Hbz-Coin). Buy Volume Increase of 295%. Price Increase of 16.79%. Current Price BTC 0.00000003 ($0.0003013071) Trade detected on exchange:HitBTC (http://bit.ly/2CEL30A) #HBZ #Crypto #HitBTC #Hbz-Coin. This tweet from Bobcat Crypto shared the ICO's data and possibly persuaded other investors to back the ICO by describing information of increased volume, increased price, and exchanged ICO. Moreover, a hashtag symbol can differentiate the Hbz ICO from other ICOs. The sentiment of the tweet from Bobcat was considered positive.

TABLE II. EXAMPLE OF TWEETS WITHIN A DATA SET

ID	1165758049073737729
Username	Bobcat Crypto
Tweet	Bobcat has detected a potential MEDIUM strength BUY Signal for $HBZ (Hbz-Coin). Buy Volume Increase of 295%. Price Increase of 16.79%. Current Price BTC 0.00000003 ($0.0003013071) Trade detected on exchange: HitBTC (http://bit.ly/2CEL30A) #HBZ #Crypto #HitBTC #Hbz-Coin
URL	/BobcatCrypto/status/1165758049073737729
Number of retweets	0
Number of favorites	0
Number of replies	0
Date and time	8/26/2019 5:49:12 AM

Twitter data were used to analyze social media comments to proceed with the sentiment analysis. The NRC Emotion Lexicon is a list of words and their associations. Eight emotions were selected as anger, anticipation disgust, fear, joy, sadness, surprise, and trust, with two sentiments as negative and positive (Figure 2). When combined, this gave a total of ten real-valued sentiment scores. This lexicon contains a real-valued sentiment score associated with single words and is designed specifically for Twitter. Results of the sentiment analysis from the R program (Table III) show the scores of different sentiments.

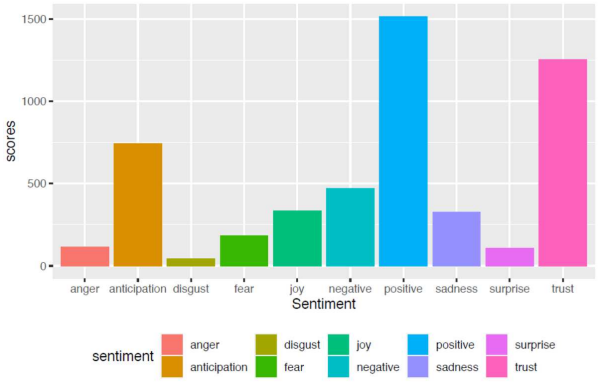


Figure 2. Sentiments Analysis by R Program

TABLE III. EXAMPLE SENTIMENT SCORE

Sentiment	Score
anger	117
anticipation	745
disgust	43
fear	185
joy	336
sadness	328
surprise	107
trust	1253
negative	471
positive	1515

Relationships between market sentiment and successful ICOs were evaluated using the Orange program. This is an open source machine learning and data mining toolkit that includes visualization, exploration, preprocessing and modeling techniques. The Orange is an easy programming tool for machine learning and data mining in Python and C++. Python modules can support data mining tasks from data preprocessing for modeling and evaluation [17]. Here, classification algorithms were applied as a supervised machine learning approach to classify whether tweets concerning ICOs were positive, neutral or negative.

Finally, predictive models were constructed to determine an optimal model to predict successful ICOs using the sentiment data from Twitter. The predictive models were implemented in the Orange program.

IV. Empirical Results

Among the 87 analyzed ICOs, 53 (60.92%) were successful and 34 (39.08%) failed. The business service sector had 10 successful ICOs and 7 failed ICOs. The cryptocurrency sector had 17 successful ICOs and 15 failed ICOs. The entertainment sector had 6 successful ICOs and 2 failed ICOs. The platform sector had 20 successful ICOs and 10 failed ICOs as shown in Figure 3.

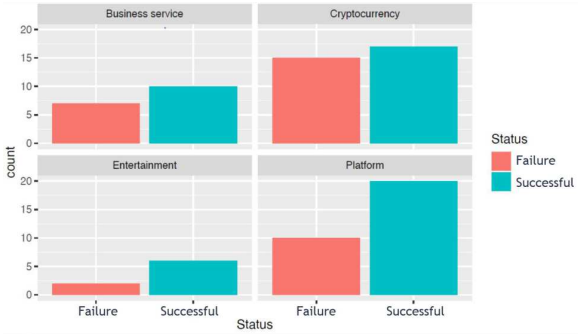


Figure 3. Number of Successful and Failed ICOs by Industry sector

As shown in Table IV, Australia had more successful ICOs than failed ICOs in the Entertainment and Platform sectors, while Singapore had more successful ICOs than failed ICOs in all industry sectors.

TABLE IV. Comparison of Successful and Failed ICOs

Industry sector	Australia		Singapore	
	Successful	***Failed***	***Successful***	***Failed***
Business service	1	2	9	5
Cryptocurrency	1	5	16	10
Entertainment	2	0	4	2
Platform	7	0	13	10
Total	**11**	**7**	**42**	**27**

Sentiment analysis results from Twitter data show that positive tweets were associated with the success of ICOs and referred to fundraising achievement.

The decision tree model is presented in Figure 4. Nodes of the decision tree represent the eight emotional aspects as the variables anger, anticipation, disgust, fear, joy, sadness, surprise, and trust in the tweets, and two sentiments as positive or negative. The attribution "Fear" was randomly selected as the root node. The decision tree model was based on the improved method to decide whether an ICO will be a success or not. In the leaf nodes, the number of branches extracted from this node is two feature attributions. Equal to the possible values, this ICO was found sadness or surprise. The sadness attribute was broken down into branches of sadness and joy. The fear emotion was considered to be the most important factor affecting decision-making by investors to predict ICO success or failure.

The predictive models were generated based on five different classification methods consisting of Decision Tree, Support Vector Machines (SVMs), Random Forest, Logistic Regression and AdaBoost.

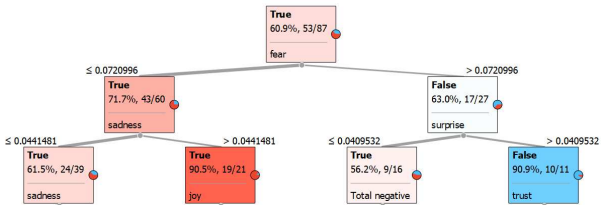


Figure 4. Decision Tree Diagram

Performance improvement is most likely driven by the domain of the data set. Our method was trained on data limitation as a cross-validation setup. The performance of each model was evaluated using a 5-fold cross-validation approach. The data were divided into 5 equal parts. Four parts of the divided data were used for training the predictive model and the remaining part was used to test model performance. The performances of each method are illustrated in the format of confusion matrices as shown in Tables V, VI, VII, VIII and IX comprising Decision Tree, Logistic Regression, SVM, Random Forest, and Adaboost, respectively.

TP (True Positive) and TN (True Negative) referred to the number of correctly detected successful and failed ICOs, respectively. FP (False Positive) denoted failed ICOs that the model falsely classified as successful, whereas FN (False Negative) referred to the number of successful ICOs that were incorrectly classified as failed.

TABLE V. CONFUSION MATRIX OF DECISION TREE

		Predicted		
		False	*True*	∑
Actual	*False*	13	21	34
	True	21	32	53
	∑	34	53	87

TABLE VI. CONFUSION MATRIX OF LOGISTIC REGRESSION

		Predicted		
		False	*True*	∑
Actual	*False*	1	33	34
	True	4	49	53
	∑	5	82	87

TABLE VII. CONFUSION MATRIX OF SVM

		Predicted		
		False	*True*	∑
Actual	*False*	3	31	34
	True	11	42	53
	∑	14	73	87

TABLE VIII. CONFUSION MATRIX OF RANDOM FOREST

		Predicted		
		False	*True*	∑
Actual	*False*	7	27	34
	True	16	37	53
	∑	23	64	87

TABLE IX. CONFUSION MATRIX OF ADABOOST

		Predicted		
		False	*True*	∑
Actual	*False*	13	21	34
	True	22	31	53
	∑	35	52	87

Performance results of each predictive model are shown in Table X. Logistic Regression was considered to be the optimal predictive model for successful ICOs with 57.5% accuracy, 98.1% recall, and an F1 score as the weighted average of precision and recall at 74.8%.

TABLE X. PERFORMANCE EVALUATION RESULTS

Model	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**
Decision Tree	0.527	0.517	0.640	0.604	0.621
SVM	0.543	0.517	0.574	0.736	0.645
Random Forest	0.440	0.506	0.590	0.679	0.632
Logistic Regression	0.593	0.575	0.605	0.981	0.748
AdaBoost	0.524	0.506	0.627	0.698	0.661

Area under the ROC curve (AUC) value was used to indicate the accuracy of the predictor. Results showed close completion between Logistic Regression and Random Forest at 0.904 (Table XI).

TABLE XI. MODEL COMPARISON BY AUC

Model	Decision Tree	SVM	Random Forest	Logistic Regression	AdaBoost
Decision Tree		0.457	0.870	0.336	0.515
SVM	0.543		0.748	0.415	0.545
Random Forest	0.130	0.252		0.096	0.145
Logistic Regression	0.664	0.585	0.904		0.660
AdaBoost	0.485	0.455	0.855	0.340	

V. CONCLUSIONS

An association was determined between Twitter comments relating to the ICO market and the success of ICOs. The fear emotion was considered to be the most important factor affecting decision-making by investors. Consequently, businesses should consider the fear factor when creating an effective marketing policy. Sentiment analysis was determined as useful for ICO issuers to generate an effective whitepaper that attracted investors and, thereby, increased the amount of funds raised.

The machine learning approach was performed to predict the success of ICOs based on tweeted emotions.

In this paper, 87 ICOs were analyzed; however, the time frame of data collection limited the sample size. Future studies should increase the sample size and extend the time frame to achieve greater accuracy and precision in the results.

ACKNOWLEDGMENT

The authors are very grateful acknowledge the support from International College of Digital Innovation, Chiang Mai University and thanks to our supervisor in this research for all support and suggestion for improving the paper.

REFERENCES

[1] ICO Statistics, https://www.fundera.com/resources/ico-statistics

[2] C. Feng, N. Li , M.H.F. Wong and M. Zhang, Initial Coin Offerings, Blockchain Technology, and White Paper Disclosures, 2019.

[3] K. Schwab, "The Global Competitiveness Report 2019" World Economic Forum, Switzerland. pp.13, 2019.

[4] D. Florysiak and A. Schandlbauer, The Information Content of ICO White Papers, 2019.

[5] L. Burns and A. Moro, What Makes an ICO Successful? An Investigation of the Role of ICO Characteristics, Team Quality and Market Sentiment, 2018.

[6] B.Cetingok and B. Deola, "Analysis of success factors for Initial Coin Offerings and Automatisation of Whitepaper Analysis Using Text-Mining Algorithms," thesis, Dept. MSc in Management. Eng., Polytechnic University of Catalonia., Spain, 2018

[7] P. Cerchiello and A. Mirela Toma, ICOs success drivers: a textual and statistical analysis, Dept. MSc in Economics and Management., The University of Pavia., Italy, 2018.

[8] C. Fisch, C. Masiak, S. Vismara and J. Hendrich Block, "Motives and profiles of ICO investors." Journal of Business Research, pp.1-26, 2019.

[9] U.W. Chohan, Initial Coin Offerings (ICOs): Risks, Regulation, and Accountability, Cyberspace Law - Student Authors eJournal, 2017.

[10] S. Albrecht, B. Lutz and D. Neumann, How Sentiment Impacts the Success of Blockchain Startups – An Analysis of Social Media Data and Initial Coin Offerings. Proceedings of the 52nd Hawaii International Conference on System Sciences, Grand Wailea, Maui, USA, pp. 4546–4550, 2019.

[11] A.Agarwal, B.Xie, I.Vovsha, O.Rambow and R. Passonneau, Sentiment analysis of twitter data, in Proceedings of the Workshop on Language in Social Media (Portland, OR), pp.30–38, 2019.

[12] H. Bagheri and M.J. Islam, Sentiment analysis of twitter data, 2017.

[13] K. Jain, A. Singh and A. Yadav, Sentiment analysis using twitter data, International Research Journal of Engineering and Technology (IRJET) Vol.06, pp. 4746-4751, 2019.

[14] S. Park and J. Woo, Gender Classification Using Sentiment Analysis and Deep Learning in a Health Web Forum. Appl. Sci, 2019

[15] M. Prokofieva, Twitter-based dissemination of corporate disclosure and the intervening effects of firms' visibility: Evidence from Australian- listed companies. Journal of Information Systems, Vol.29 pp. 107–136, 2015.

[16] T. Bourveau, E. T. De George, A. Ellahie and D. Macciocchi, Information Intermediaries in the Crypto-Tokens Market, 2019.

[17] J. Demsar, B. Zupan: Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/ orange), Faculty of Computer and Information Science, University of Ljubljana, 2004.

Behavioral Analysis of Transformer Models on Complex Grammatical Structures

Kanyanut Kriengket* Kanchana Saengthongpattana* Peerachet Porkaew*,**
Vorapon Luantangsrisuk* Prachya Boonkwan* Thepchai Supnithi*

* Language and Semantic Technology Laboratory
National Electronics and Computer Technology Center (NECTEC), Pathumthani, Thailand
** Institute of Computing Technology, University of Chinese Academy of Sciences, Beijing, China
{*first.last*}@nectec.or.th vorapon.lua@ncr.nstda.or.th

***Abstract*—State-of-the-art neural MT, e.g. Transformer, yields quite promising translation accuracy. However, these models are easy to be interfered by noises, causing over- and under-translation issues. This paper presents a behavioral analysis of Transformer models in translating complex grammatical structures, i.e. multiple-word expressions and long-distance dependency. Results consistently show that the more complex structures, the less translation accuracy the models yield. We imply that as phrase structures become more complex, the focus patterns learned by the attention mechanism may get erratically sporadic due to the issue of data sparseness. We suggest the use of locality penalty and the increase of attention heads to mitigate the issue, but their trade-offs should also be aware.**

***Index Terms*—model interpretation, neural machine translation, grammatical structure, content and function words, translation evaluation**

I. Introduction

Machine Translation (MT) is a process of translating a text from a source language into a target language via logical/statistical computation. In data-driven approaches, translation can be learned solely from a set of parallel texts, or *bitext*, with a wide array of statistical approaches [1], e.g. phrase-based translation [2], induction of synchronous grammars [3], and tree-to-string transformation [4]. Neural networks can also be employed to learn a translation task by diving it into two subtasks: (1) language modeling [5], and (2) sequence-to-sequence generation via recurrent neural networks [6] and the state-of-the-art Transformer models [7].

Neural-network models yield quite promising translation results and satisfactory speed efficiency, thanks to multiple layers of the attention mechanism. They are relatively easy to interpret their behaviors, as attention on words can be visualized as a heatmap. This allows us to investigate which input words receive more focus, or *attention score*, when generating a particular output word. The attention scores are discovered to imply pseudo syntactic dependency during translation [8], offering the possibility to biligual parsing.

However, behavioral interpretation of the neural network models is still under-studied. These models are very delicate and easy to break when synthetic and natural noises are added to the data [9]. One reason is the attention mechanism tends to ignore past alignment information, resulting in over- and under-translation issues [10]. Particularly, systemic analysis of how the attention mechanism tackles a variety of complex grammatical structures has to be conducted in order to expand our epistomic horizon.

This paper presents a behavioral analysis of Transformer models in translating complex grammatical structures in the source language, where we particularly focus on multiple-word expressions and long-distance dependency. We introduce a novel and economical method for selecting translation pairs for the testing dataset based on the numbers of content and function words. Finally, we indirectly observed how well they capture complex grammatical structures via the BLEU score.

This remainder of this paper is organized as follows. Section II provides some background knowledge on Transformer-based machine translation. Section III describes our method for constructing such dataset for the behavioral study of Transformer. Section IV explains the training and testing datasets used herein and how we set up the experiments. Experimental results are elaborated in section V and discussed in section VI. Finally, section VII concludes this paper.

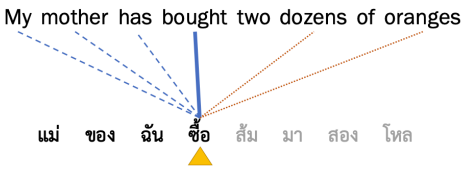
II. Transformer-based Machine Translation

Transformer [7] is a popular encoder-decoder framework for sequence-to-sequence modeling tasks. It can be employed in many applications, e.g. machine translation, automatic summarization, and chatbots.

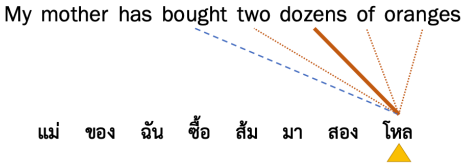
In a nutshell, Transformer consists of two steps: encoding and decoding. Each of them consists of a set of attention mechanisms (or *heads*) that learn to focus on different parts of the input and output sequences. Figure 1 shows how separate attention heads may learn to focus on verb phrases and noun phrases, respectively. The more attention heads, the more focus patterns Transformer can learn. Note that the learned focus patterns may sometimes be erratically sporadic when the training data is not large enough.

An overview of Transformer is illustrated in figure 2. First, the input sentence is encoded into a set of *input embedding vectors*, each of which learned from a separate attention head. These embedding vectors are then used to learn to focus on different parts of the output sentence, resulting in a set of *output embedding vectors*. Finally, these output vectors

978-0-7381-3068-2/20/$31.00 ©2020 IEEE



(a) Attention heads while translating 'bought'



(b) Attention heads while translating 'dozens'

Figure 1. Contextualized translation with the Transformer. Separate attention heads (blue and orange) may learn to focus and translate verb phrases and noun phrases, respectively. Bold lines denote the foci of translation, while dashed lines represent translation contexts.

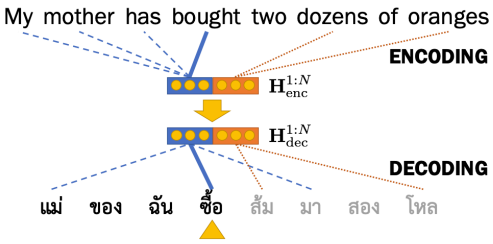


Figure 2. A conceptual overview of Transformer. The number of attention heads N is 2.

dictate the generation (or *decoding*) of each word of the output sentence. Transformer is also said to be *autoregressive*, because the generation of each output word is based on the previously generated outputs; i.e.

$$y_i = \arg\max_{y'} p(y'|x, y_{<i}) \quad (1)$$

where $y_{<i} = y_1 \dots y_{i-1}$ is the previously generated part of the output sequence at index i.

Encoder: The encoder encodes the input tokens into a memory matrix ($\mathbf{H}_{\text{enc}}^{1:N}$) using L layers of scaled dot-product network. Here we use the scaled dot-product attention as a self-attention mechnism for the sake of convenience. Suppose $\mathbf{H}^{(0)}$ is the input matrix consisting of word and positional embedding. The self-attention network is formulated in eq 2:

$$\begin{aligned} \mathbf{C}^{(l+1)} &= \text{LN}(\text{Attn}_{\text{enc}}(\mathbf{H}^{(l)}) + \mathbf{H}^{(l)}), \quad (2) \\ \mathbf{H}^{(l+1)} &= \text{LN}(\text{FNN}(\mathbf{C}^{(l+1)}) + \mathbf{C}^{(l+1)}), \end{aligned}$$

where LN is the layer normalization and FNN is the feed-forward neural network. For each head h, we define the self-attention mechanism $\text{Attn}_{\text{enc}}^{(h)}$ in eq 3:

$$\text{Attn}_{\text{enc}}^{(h)}(\tilde{\mathbf{H}}_h) = \text{softmax}\left(\frac{\tilde{\mathbf{H}}_h \tilde{\mathbf{H}}_h^\top}{\sqrt{d}}\right)\tilde{\mathbf{H}}_h, \quad (3)$$

where $\tilde{\mathbf{H}}_h$ is the submatrix of $\mathbf{H}$ corresponding to the attention head *(h)* and d is a scaling factor. Note that $\mathbf{H}^{(L)}$ is the hidden states in the last layer. The representation of the source sentence is $\mathbf{H}_{\text{enc}}^{1:N}$, which is the concatenation of N self-attentions.

$$\begin{aligned} \text{Attn}_{\text{enc}}(\mathbf{H}) &= \text{MultiheadAttn}_{\text{enc}}(\mathbf{H}) \quad (4) \\ &= \text{Attn}_{\text{enc}}^{(1)}(\tilde{\mathbf{H}}_1) \circ \dots \circ \text{Attn}_{\text{enc}}^{(N)}(\tilde{\mathbf{H}}_N) \end{aligned}$$

Decoder: The decoder also consists of L identical layers of Transformer. Suppose $\mathbf{D}^{(0)}$ is the output matrix consisting of word and positional embedding. The self-attention network for the decoder is formulated in eq 5:

$$\begin{aligned} \mathbf{R}^{(l+1)} &= \text{LN}(\text{Attn}_{\text{enc}}(\mathbf{D}^{(l)}) + \mathbf{D}^{(l)}) \quad (5) \\ \mathbf{C}^{(l+1)} &= \text{LN}(\text{Attn}_{\text{dec}}(\mathbf{R}^{(l+1)}, \mathbf{H}_{\text{enc}}^{1:N})) + \mathbf{R}^{(l+1)}) \\ \mathbf{D}^{(l+1)} &= \text{LN}(\text{FNN}(\mathbf{C}^{(l+1)}) + \mathbf{C}^{(l+1)}) \end{aligned}$$

For each head h, we define the decoder-encoder attention sub-layer $\text{Attn}_{\text{dec}}^{(h)}$ in eq 6.

$$\text{Attn}_{\text{dec}}^{(h)}(\tilde{\mathbf{R}}_\mathbf{h}, \tilde{\mathbf{H}}) = \text{softmax}\left(\frac{\tilde{\mathbf{R}}_\mathbf{h} \tilde{\mathbf{H}}^\top}{\sqrt{d}}\right)\tilde{\mathbf{H}}, \quad (6)$$

where $\tilde{\mathbf{R}}_h$ is the submatrix of the decoder intermediate states corresponding to the attention head (h). In practice, the decoder-encoder attention is calculated by multihead attentions.

$$\begin{aligned} \text{Attn}_{\text{dec}}(\mathbf{R}, \mathbf{H}_{\text{enc}}^{1:N})) &= \text{MultiheadAttn}_{\text{dec}}(\mathbf{R}, \mathbf{H}) \quad (7) \\ &= \text{Attn}_{\text{dec}}^{(1)}(\tilde{\mathbf{R}}_1, \mathbf{H}_{\text{enc}}^{1:N}) \\ &\quad \circ \dots \circ \text{Attn}_{\text{dec}}^{(N)}(\tilde{\mathbf{R}}_N, \mathbf{H}_{\text{enc}}^{1:N}) \end{aligned}$$

In the decoding step, the decoder queries the information from the encoder using the decoder hidden states $\mathbf{D}^{(L)}$ as keys. Finally, the decoder predicts the next target word (y_i) from the probability distribution $p(y'|x, y_{<i})$.

$$p(y'|x, y_{<i}) = \text{softmax}(W_o D_i), \quad (8)$$

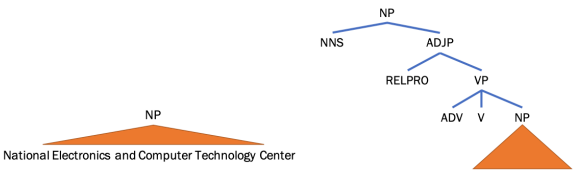
where W_o is the output weight matrix.

Objective function: Given a training set consisting of S sentence pairs $\{\langle x^{(s)}, y^{(s)}\rangle | 1 \le s \le S\}$, the training objective is to maximize the log likelihood $\mathcal{L}$ w.r.t. model parameters θ as shown below.

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{s=1}^{S} \log p(y^{(s)}|x^{(s)}; \theta) \quad (9) \\ &= \sum_{s=1}^{S} \sum_{j=1}^{J(s)} \log p(y^{(s)}|x^{(s)}, y_{<j}^{(s)}; \theta) \end{aligned}$$

where j is the index of each word in $y^{(s)}$, and $J(s)$ is the length of sentence $y^{(s)}$. As aforementioned, Transformer is an autoregressive model, because it also takes the previously generated sequence $y_{<j}$ as an input then generates the output symbol y_j.

Area to explore: Since Transformer is autoregressive, the decoding is by and large based on locality and the learned focus patterns. If it learns short locality and sporadic focus



(a) Flat and long phrase structure (b) Deep phrase structure

Figure 3. Two criteria for data selection

patterns, some linguistic phenomena may not be captured, e.g. multiple-word expressions and long-distance dependency.

Transformer-based machine translation has not been assessed in the linguistic perspective. Although its mechanism can be visually interpreted via heat maps, the syntactic learning capability is still under-studied. In this paper, we look forward to demonstrating how Transformer models cope with various grammatical structures in natural language, especially multiple-word expressions and long-distance dependency. The next section will describe how we construct a testset that covers both linguistic phenomena in an economical fashion.

III. Our Method for Testset Construction

A. Our Hypothesis

We desire to demonstrate how Transformer models cope with complex grammatical structures. We particularly focus on multiple-word expressions and long-distance dependency, which adversely affect the translation accuracy.

In regards to syntactic complexity, we classify the grammatical structures in natural language into two types: (1) flat and long phrase structure and (2) deep structure, as illustrated in figure 3. The first type is characterized by a relatively high ratio of content words to function words and non-recursive embedding structures. The latter, however, is characterized by the lower ratio and recursive embedding structures, e.g. subordinate clauses, relative clauses, and prepositional phrases. However, directly determining complex structures in plain text is computationally expensive owing to syntactic parsing being required.

To facilitate this step, we introduce a novel and economical method for selecting sentence pairs with complex structures. By the fact that any deep and wide parse tree always produces a lengthy sentence, it must contain more words than a shallow and narrow one. We then imply that the first contains more content words (i.e. noun, verb, adjective, and adverb) and function words (the remainders) than the latter.

In our method, we assume that we can rank the sentence pairs based on the numbers of content and function words in lieu to dependency structures. Based on our observation, we found that sentences with more content words tend to have flat and long phrases, e.g. named entities and noun phrases. Meanwhile, sentences with more function words tend to have deeper structures, e.g. preposition phrases and relative clauses.

B. Data Selection

We constructed our testset in the following fashion. We first selected the 680K EN-TH sentence pairs from COLING-2014 Workshop on Statistical Machine Translation [11], whose English side is automatically tokenized with NLTK's Punkt Tokenizer [12], because it was used as a universal testset for evaluating machine translation systems. On the Thai side, word segmentation was done automatically with SWATH [13].

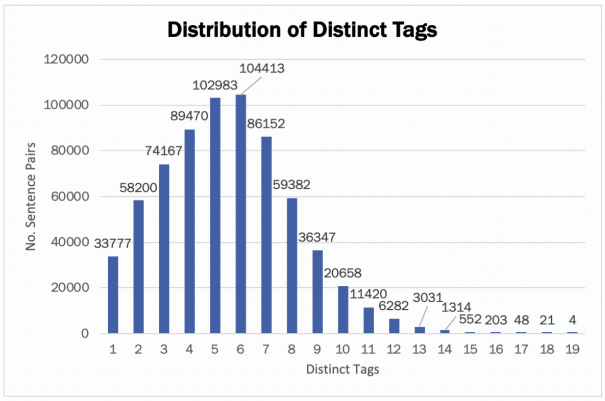


Figure 4. Distribution of distinct POS tags

In order to discern content words from function ones, we annotated the POS tags on the English part using NLTK's maximum entropy POS tagger [14] trained on Penn Treebank tagset [15]. We distinguish the POS tagset into two general groups:

- **Content words:** `JJ`, `JJR`, `JJS`, `NN`, `NNS`, `NNP`, `NNPS`, `RB`, `RBR`, `RBS`, `VB`, `VBD`, `VBG`, `VBN`, `VBP`, and `VBZ`
- **Function words:** `CC`, `CD`, `DT`, `EX`, `FW`, `IN`, `LS`, `MD`, `PDT`, `POS`, `PRP`, `PRP$`, `RP`, `SYM`, `TO`, `UH`, `WDT`, `WP`, `WP$`, and `WRB`

By discerning the POS tags with the content/function types, we are now able to rank all sentence pairs with their occurrence frequencies.

C. Ranking of Sentence Pairs

In this step, we ranked all sentences pairs w.r.t. the structural complexity on the English side. First, we descendingly sorted the sentence pairs by the number of distinct POS tags, because we can take into account flat and long phrase structures as well as deep structures in parallel. The distribution of distinct POS tags is illustrated in figure 4. Note that it merely resembles a Poisson distribution with the expectation of 6 distinct POS tags per sentence.

Next, we descendingly sorted them by the number of content words to rank them by flat and long structures. We again descendingly sorted them by the number of function words to rank them by deep structures. As a result, all sentences are ranked by flat and long structures and by deep structures, respectively.

Finally, we selected top-0.3% most complex sentences of each group of distinct tags, except that we selected the entire group of 19 distinct tags. The corpus statistics of the selected sentence pairs are shown in table I.

Table I
CORPUS STATISTICS OF OUR TESTING DATASET

Sentence pairs	2,359
No. content words	4–27
No. function words	0–20
No. distinct words	7,690
Sentence lengths	4–40
Average sentence lengths	14.72
Standard deviation	9.05

Table II
DISTRIBUTION OF SENTENCE PAIRS IN OUR TRAINING DATASET

Sources	**Sentence Pairs**
IWSLT 2015	83,528
HIT London Olympic	62,733
BTEC	100,012
`ThaiSubtitle.com`	1,164,040
ASEAN MT	20,543
TED talks	231,961
`OpenSubtitle.com`	3,216,815
Total	**4,879,632**

IV. EXPERIMENTAL SETUP

A. Training Datasets

We employed two training datasets of different sizes to compare the capability of Transformer in capturing complex grammatical structures when trained on different datasets.

OpenSubtitle+SCB: We used a combination of `OpenSubtitle.com` and Siam Commercial Bank's open bitext [16]. This latter dataset consists of 1M sentence pairs and was created from various sources: professional translators, crowdsourced translators, and web-crawled and automatically aligned data. Once combined, we eliminated any sentence pairs whose either side is shorter than 3 words or whose source-to-target length ratio is less than 0.4. As a result, this integral dataset consists of 4.18M sentence pairs, 302,398 distinct words, and 44,165,904 words in total.

Our training dataset: We created our own training dataset from various sources: IWSLT 2015, HIT London Olympic multilingual dialogs, Basic Travel Expression Corpus (BTEC), `ThaiSubtitle.com`, ASEAN Machine Translation Project, TED talks, and `OpenSubtitle.com`. As a result, this integral dataset consists of 4.88M sentence pairs, 200,096 distinct words, and 41,368,376 words in total. The distribution of sentence pairs is shown in table II.

Both training datasets are preprocessed in the following fashion. English sentences are tokenized into words with `tokenizer.perl` in Moses SMT framework [2] and pre-processed using byte-pair-encoding (BPE) [17] with BPE operation set to 50K steps. Thai source sentences are tokenized into words with Thai word segmentation DeepCut [18]. We are aware that our testing dataset is tokenized with a different Thai word tokenizer (i.e. SWATH). Despite slight discrepancies in their word segmentation guidelines, the results are quite comparable when used with colloquial Thai.

B. Testing Datasets

We will also compare the translation accuracy against two datasets to avoid any biases. The first one is IWSLT 2015 testset, which is available online and consists of 4,242 sentence pairs. The other one is the dataset we created in the method aforementioned in section III. Note that the construction of the latter was driven by wide coverage of complex grammatical structures.

Table III
OVERALL RESULTS.

Training datasets	**Distinct words**	**Sentence pairs**	**Testing datasets**	
			IWSLT	**Ours**
OpenSub+SCB	302,398	4.18M	20.52	12.55
Ours	200,096	4.88M	20.25	12.15

C. System Configuration

We employed the `FairSeq` Framework [19] as our Transformer implementation. The models were trained on 4 NVIDIA RTX2080Ti GPUs with update frequency set to 2, to simulate 8 GPUs. We set the number of attention heads for the encoder and decoder to be eight, and the scaling factors d to be 64. All models are trained for 100K steps and checkpoints are saved at every 500 updates. Then, we selected the best check point that perform best on development set to for assessment. We used Bilingual Evaluation Understudy (BLEU score) [20] for evaluating the translation performance.

V. RESULTS

In the following experiments, we performed English-to-Thai translation based on Transformer models and indirectly observed how well they capture complex grammatical structures via the BLEU score (i.e. translation accuracy).

A. Overall Results

We trained Transformer models on both training datasets and evaluated them against each testing dataset. The results are shown in table III.

Surprisingly, the general translation performance of Transformer trained on our training dataset is on par with that trained on OpenSubtitle+SCB. Although our training dataset has four times more distinct words while having almost equal number of sentence pairs, Transformer managed to circumvent the curse of dimensionality issue quite well.

However, the translation accuracies on our testing datasets are significantly lower than those of IWSLT 2015 across the board. We imply that our testing dataset has a different distribution of linguistic phenomena than IWSLT 2015 testset. This is because our testing dataset is designed to be experimentally more controllable in terms of grammatical complexity.

B. Results on Capturing Grammatical Structures

Next, we varied the complexity of grammatical structures by adjusting the number of content and function words, and observed the translation performance. This was solely done on our testing dataset.

First, we studied the effects of flat and long phrase structures towards the Transformer models in table IV. We classified our

Table IV
EFFECTS OF FLAT AND LONG PHRASE STRUCTURES

	No. content words		
	low (< 10)	medium (10-19)	high (> 19)
BLEU	16.2	12.6	10.3
Support	1,273	909	177

Table V
EFFECTS OF DEEP PHRASE STRUCTURES

	No. function words		
	low (< 10)	medium (10-19)	high (> 19)
BLEU	15.2	11.2	11.6
Support	1,917	431	11

testing dataset into three groups w.r.t. the number of content words in the sentences: *low* (< 10 words), *medium* (10-19 words), and *high* (> 19 words). As the number of content words increases, the translation accuracy steadily decreases.

Second, we examined the effects of deep phrase structures towards the Transformer models in table V. We again classified our testing dataset into three groups w.r.t. the number of function words in the sentences: *low* (< 10 words), *medium* (10-19 words), and *high* (> 19 words). As the number of function words increases, the translation accuracy decreases sharply.

Third and last, we emphasized the effects of deep structures in complex sentences as shown in table VI, by considering only sentences with less than 10 content words. We classified the remaining testing dataset into three groups w.r.t. the number of function words in the sentences: *low* (< 10 words), *medium* (10-19 words), and *high* (> 19 words). The translation accuracy initially saturates, then it sharply drops when the sentences get very complex (more than 19 function words).

VI. DISCUSSION

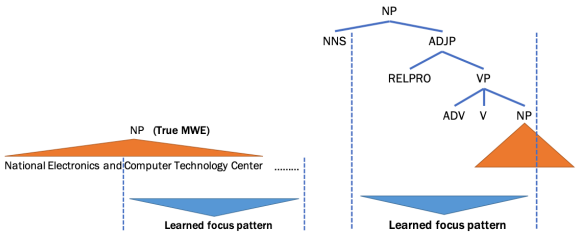
A. Summary of Findings

In table IV, the results suggest that flat and long phrase structures, as reflected by the number of content words, directly affect the translation accuracy. We suspect that as flat phrase structures get lengthened, the learned focus patterns may get erratically sporadic outside the phrase span as seen in figure 5(a), resulting in incorrect translation.

In table V, the results suggest that deep phrase structures, as reflected by the number of function words, adversely affect the translation accuracy. We postulate that as phrase structures get deepened, the ability to learn to focus locally deteriorates

Table VI
EFFECTS OF DEEP PHRASE STRUCTURES IN COMPLEX SENTENCES (LESS THAN 10 CONTENT WORDS)

	No. function words		
	low (< 10)	medium (10-19)	high (> 19)
BLEU	16.4	15.9	8.4
Support	1,145	113	15



(a) Sporadic attention on multiple-word expressions (b) Sporadic attention on deep phrase structures

Figure 5. Erratically sporadic attention

due to the data sparseness issue, resulting in context-less, or even context-free, decoding.

Quite surprisingly, in table VI, the results confirm the previous findings that the ability to learn to focus locally starts to deteriorate when evaluating the models on deeper phrase structures. We suspect that this is due to the scarceness of attention heads that cope with various phrase types inside a deep structure as seen in figure 5(b).

B. Error Analysis and Suggestions

Our error analysis reveals that translation of both types of grammatical structures are impacted by long-distance dependency. The flatter, longer, and deeper phrase structures are, the more distant the heads and corresponding modifiers become. In a language pair of different word orders (e.g. head-final English and head-initial Thai), this poses the issues of over-generation, undergeneration, mistranslation. and reordering of modifiers.

The following example demonstrates these issues in translating a flat and long phrase structure. The source text, the ground-truth translation, and a machine-generated translation are provided in (1a), (1b), and (1c), respectively. Note that the source text is a name of a CNC machine, which is a multiple-word expression.

(1) a. 'automatic bending machine for manufacturing of steel-rule cutting dies'

b. kʰruâŋ.muː sămràp dàt.kʰóːŋ ʔàttànoːmát
pʰuâː pʰàlìt mæ̀ː.pʰim lèk.pʰæ̀n
machine for bend automatic
for manufacture cutting-dies steel-rule

c. *kʰruâŋ dàt ʔàttànoːmát sămràp kaːn
pʰàlìt kʰɔ̌ːŋ lèk$_1$ tàt$_2$ kòt$_1$ taː$_{12}$
engine flex automatic for NOM
manufacture of metal cut rule decease

The nominalizing prefix kaːn is overgenerated in Thai (in blue). This is because the model learns to translate *for* into the pattern sămràp 'for' followed by a noun phrase. However, it can also be translated as the pattern conjunction pʰuâː 'for' followed by a verb phrase when long-distance dependency between the core noun and this preposition is taken into account. For undergeneration (in red), the words kʰruâŋ.muː 'machine, tool', dàt.kʰóːŋ 'bend', and lèk.pʰæ̀n are reduced to their cores kʰruâŋ 'engine', dàt 'flex', and lèk 'metal', respectively,

because their contexts are discarded in translation. Mistranslation (in violet) takes place when translating (*cutting*) *dies* into taːɪ 'pass away', because the context *cutting* is separated. Finally, the reordering issue (underlined) is a consequence of the mistranslation. As marked by numbers, two phrases *steel-rule* (1) and *cutting dies* (2) overlap each other when translated into Thai.

Translating a deep structure can be as problematic as seen in the following example. The source text, the ground-truth translation, and a machine-generated translation are provided in (2a), (2b), and (2c), respectively.

(2) a. 'I went shopping for a coat and met a friend for dinner.'

b. ɕʰăn paɪ sɯ́ː sɯâː kʰóːt lǽ paɪ kin ʔaːhăːn.kʰâm kàp pʰɯân
I go buy shirt coat and go eat dinner with friend

c. *ɕʰăn paɪ sɯ́ː sɯâː kʰóːt lǽ dâːɪ (paɪ) pʰóp pʰɯân pen ʔaːhăːn.kʰâm
I go buy shirt coat and PAST go meet friend as dinner

The past-tense marker dâːɪ is overgenerated (in blue) because the long-distance dependency between the two main verbs sɯ́ː 'buy' and pʰóp 'meet' is not detected. Owing to the same reason, the directional verb paɪ 'go' is discarded from the translation, resulting in undergeneration (in red). Last, the preposition *for* (*dinner*) is mistranslated into pen 'as' (in violet), because the dependency between the verb *met* and the preposition *for* is not detected.

In regards to the above error analysis, we suggest two possible ways to mitigate this issue. First, we may introduce the locality penalty to each attention head as a means to impose the focus beam locally, which copes with flat and long phrase structures. Second, we may increase the number of attention heads to handle more types of deep phrase structures. However, there are trade-offs in both solutions. By introducing locality penalty, the learning of long-distance dependency will become locally limited. By increasing the number of attention heads, more training data will be required as it introduces the data sparseness issue. An extra care might have to be taken when solving this issue.

VII. Conclusion

This paper has presented a behavioral study of Transformer models in coping with complex grammatical structures, especially multiple-word expressions and long-distance dependency. To reflect this issue, we introduced a novel and economical method for selecting translation pairs for the testing dataset based on the numbers of content and function words. Experiment results showed that the ability to learn to focus in translation deteriorates when flat, long, and deep structures are incorporated in the sentences, resulting in erratically sporadic focus patterns.

Our future work remains as follows. First, we will investigate the effects of attention heads in learning to focus and how to avoid the data sparseness issue as we increase its number. Second, we will examine the possibility to incorporate the locality penalty without losing the ability to capture long-distance dependency.

Acknowledgements

This work is supported by a close collaboration between Language and Semantic Technology (LST) laboratory, National Electronics and Computer Technology Center (NECTEC), and Institute of Computing Technology (ICT), Chinese Academy of Sciences (UCAS).

References

[1] P. Brown, S. della Pietra, V. della Pietra, and R. Mercer, "The mathematics of statistical machine translation: parameter estimation," *Computational Linguistics*, vol. 19, no. 2, pp. 263–311, 1993.

[2] P. Koehn, O. F. Josef, and D. Marcu, "Statistical phrase-based translation," in *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL 2003)*, pp. 48–54, 2003.

[3] D. Chiang, "A hierarchical phrase-based model for statistical machine translation," in *Proceedings of the Association for Computational Linguistics (ACL 2005)*, pp. 263–270, 2005.

[4] Y. Liu, Q. Liu, , and A. Lin, "Tree-to-string alignment template for statistical machine translation," in *Proceedings of the Association for Computational Linguistics (ACL 2006)*, pp. 609–616, 2006.

[5] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research (JMLR)*, vol. 3, pp. 1137–1155, 2003.

[6] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proceedings of ICLR 2015*, 2015.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Conference on Neural Information Processing Systems (NIPS 2017)*, 06 2017.

[8] K. Tran and Y. Bisk, "Inducing grammars with and for neural machine translation," in *Proceedings of the Workshop on Neural Machine Translation (WNMT 2018)*, 2018.

[9] Y. Belinkov and Y. Bisk, "Synthetic and natural noise both break neural machine translation," in *Proceedings of ICLR 2018*, 2018.

[10] Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, "Modeling coverage for neural machine translation," in *Proceedings of the Association for Computational Linguistics (ACL 2016)*, pp. 76–85, 2016.

[11] O. Bojar, C. Buck, C. Federmann, B. Haddow, P. Koehn, C. Monz, M. Post, and L. Specia, eds., *Proceedings of the Ninth Workshop on Statistical Machine Translation*, (Baltimore, Maryland, USA), Association for Computational Linguistics, June 2014.

[12] T. Kiss and J. Strunk, "Unsupervised multilingual sentence boundary detection," *Computational Linguistics*, vol. 32, pp. 485–525, 2006.

[13] S. Meknavin, P. Charoenpornsawat, and B. Kijsirikul, "Feature-based thai word segmentation," in *Proceedings of the Natural Language Processing Pacific Rim Symposium*, pp. 41–46, 1997.

[14] A. Ratnaparkhi, "A maximum entropy model for part-of-speech tagging," in *Conference on Empirical Methods in Natural Language Processing*, 1996.

[15] A. Taylor, M. Marcus, and B. Santorini, *Treebanks: Building and Using Parsed Corpora*, vol. 20 of *Text, Speech, and Language Technology*, ch. 1: The Penn Treebank: an Overview, pp. 5–22. Springer: Dordrecht, 2003.

[16] L. Lowphansirikul, C. Polpanumas, A. T. Rutherford, and S. Nutanong, "scb-mt-en-th-2020: A large english-thai parallel corpus," 2020.

[17] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proceedings of the Association for Computational Linguistics (ACL 2015)*, 2015.

[18] R. Kittinaradorn, T. Achakulvisut, K. Chaovavanich, K. Srithaworn, P. Chormai, C. Kaewkasi, T. Ruangrong, and K. Oparad, "DeepCut: A Thai word tokenization library using deep neural network," Sept. 2019.

[19] M. Ott, S. Edunov, A. Baevski, A. Fan, S. Gross, N. Ng, D. Grangier, and M. Auli, "fairseq: A fast, extensible toolkit for sequence modeling," in *Proceedings of NAACL-HLT 2019: Demonstrations*, 2019.

[20] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the Association for Computational Linguistics (ACL 2002)*, pp. 311–318, 2002.

Cryptocurrencies Asset Pricing Analysis: evidence from Thailand markets

Kanyawut Ariya
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
kanyawut_ariya@cmu.ac.th

Nathee Naktnasukanjn
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
nathee@cmuic.net

Tanarat Rattanadamrongaksorn
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
tanarat.r@cmuic.net

Piyachat Udomwong
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
piyachat.u@cmu.ac.th

Saronsad Sokantika
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
saronsad.s@cmu.ac.th

Nopasit Chakpitak
International college of digital innovation
Chiang Mai University
Chiang Mai, Thailand
nopasit@cmuic.net

***Abstract*—Can cryptocurrencies price variations be explained by exogenous classical market prices? We evaluate this issue by using daily data on some of the most important asset prices and indexes in Thailand i.e. Gold, Oil, SET50 index, Tourism index, Mutual fund, and THB/USD exchange rate in comparison with digital asset prices i.e. Bitcoin, Ethereum, Litecoin, Ripple, DASH, and Stellar. By performing both direct and inverse relationships using correlation matrix to find distance relationship and using minimum spanning tree to find the closest path between assets, we found strong direct relationship among cryptocurrencies in digital market with SET50 index and oil price in classical markets. We also found that THB-USD exchange rate has inverse relationship with Bitcoin price, SET50 index and oil price. There is a link between cryptocurrencies asset price and some classical assets' market price.**

Keywords: cryptocurrencies, digital asset pricing, capital market, correlation matrix, minimum spanning tree

I. INTRODUCTION

Since the technical white paper of Satoshi Nakamoto 2008 [1] has been released, Bitcoin is known as the first decentralized cryptocurrency. Cryptocurrency is a digital asset used as a medium of exchange [2]. Decentralized cryptocurrency means that there is no central financial control, it does not rely on any financial institution to issue, manage, destroy and re-issue for public usage. These cryptocurrencies therefore encourage the decentralization of the economy [3].

Bitcoin and/or other digital currencies may disrupt or have impact to the traditional financial systems. They can provide financial services independently and reduce, if not eliminate, the requirement of money transfer across states' borders and hence increase the business opportunities. However, the digital currency price is volatile compared to other investment assets. Price volatility is causing serious concern to investors [4]. These concerns lead to appeals for regulation of cryptocurrencies.

In Thailand, cryptocurrencies have started trading since 2013 with a trading website name www.bx.in.th. The platform allows traders to exchange baht to cryptocurrencies. People in the market claimed that they can make a lot of money. The profit has received many criticisms from financial analysts and economists, including the investment regulatory authority in Thailand. There are warnings for those who invest or trade in the market to be aware of risks due to the high fluctuation of the market and high uncertainty compare to the classical capital market [5].

The classical capital markets are financial markets that involve long-term debt or equity trading that provide financing for more than one year. Commonly, financial markets consist of money market and capital markets. Generally, money markets are financial markets where financial instruments with short-term maturity and high liquidity are traded, such as bills of exchange or promissory notes. These securities are issued by governments, financial institutions, and large companies. The capital market consists of two sub-markets which are the primary market, and secondary market. The primary market is concerned with newly issued securities and is responsible for creating new long-term capital. The secondary market manages the trading of previously released securities in the market, such as stocks and bonds, to maintain liquidity for investors. Capital markets with high liquidity and high transparency are of interest to investors as well [6].

Can we use classical financial asset pricing model like Fama and French (1993)'s [7] three-factor model or Naktnasukanjn et al. (2018) [8] six-factor model to evaluate the price and return of the cryptocurrencies? Let us take some examination whether there is any relationship between cryptocurrencies and asset price in classical financial markets. Conrad et al. (2018) [9], and Ciaian and Rajcaniova (2018) [10] show strong interconnected relationships between cryptocurrencies, but Giudici et al. (2019) [11] shows that Bitcoin prices are unrelated with classical market asset prices. By comparing between Bitcoin exchanges and traditional financial market, Mantegna (1999) [12] proposed the minimum spanning tree (MST) to identify the hierarchical structure of capital asset prices in financial market. Using the correlation matrix as a network structure between node group shows different periods due to the large amount of data to be analyzed and scattered. Giudici et al. (2019) [11] has adopted the random matrix theory (RMT) to filter the data by removing the noise contained in it, and that improves its interpretation

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

effectively following the methodology of Tola et al. (2008) [13]

In this research, we are interested in studying on investing in the digital asset markets compared to the classical asset in capital market in Thailand following the Giudici et al. (2019) [11] model. While some other papers examined this issue from statistical methods focusing on Bitcoin prices and capital asset price to find the relationship of both markets from many factors, we, however, are interested in studying the relationship between capital market and digital asset market in Thailand. Because the advancement of the law and regulation from the regulator (SEC) in Thailand creates the confident of the trading platform, we believe the cryptocurrencies trading are liquid in Thailand market and hence will allow the economic mechanism to perform properly. The research will focus on comparing asset price data in each period to find the direct relationship and inverse relationship of the returns from both markets.

II. Literature review

This section is a collection of literature studies relating to this research. To be a reference for efficient operations which will be the study of literature as following.

A. Digital asset market

The technology behind cryptocurrency is what makes cryptocurrency reliable. Related work is the paper of Alen Hrga [14] who researched blockchain technology which is the technology of cryptocurrency. The paper explains the concept of blockchain, which is a technology that decentralizes information. It does not store in any local system that can verify the accuracy of the data, also that cannot be updated and changed to encourage a cryptocurrency is that has the transparency of information. The research paper of Hassani et al. [15] shows that the combination of cryptocurrencies and blockchain technology can reduce the risk of bubbling, which is an advantage for digital market operators and investors. The blockchain technology is a security network that can analyze and share data in real-time, enabling big data analysis to improve operational efficiency and profitability for investors, while also protecting against the loss caused by price fluctuations.

Ethereum, one of the assets using blockchain technology. The Ethereum ecosystem, despite its high diversity in composition, popularity in use, service life, and asset trading volumes, continues to follow a global structure. Also, Shahar Somin et al. [16] show that while this economy is highly volatile from various traditional perspectives, it shows a stable and balanced form.

A study of the advantages and disadvantages of digital assets was studied in the document's Giancarlo Giudici (2019) [17] focuses on the emerging phenomenon of cryptocurrencies. Study the factors of digital currency price increases in the market and socio-economic issues and the behavior and sustainability of cryptocurrencies. It is hypothesized that cryptocurrencies may perform some useful functions and increase economic value. The results of the study show the advantages of currency which will increase the efficiency of the overall investment market due to the speed in the operation of each step, including no strict rules. On the other hand, there are still disadvantages of cryptocurrencies that are not properly controlled, causing the advantages of the cryptocurrencies to be misused such as money bubble financial crime or disruption to the traditional economy.

In Thailand, digital assets have been invested and have studied the investment in digital assets in the paper's Natnicha Tangwattanarat [5] is a study of the perception of Thai cryptocurrency investors towards the digital currency market by interviewing with 25 Thai cryptocurrency investors. The objective of this study is to understand the overall market situation and the perception of investors in Thailand towards investing in the digital currency and the digital currency market. The results from the study show that the interviewees. They are aware and have a good attitude towards investing in cryptocurrencies on the market situation. Most of them trust and continually increase their investment in cryptocurrencies regardless of the market situation. Also, investors do not seem to be extremely cautious when choosing the digital currency. They invest in technical reasons, most of which are about perspectives on technology and innovation and positive news about social media.

B. Capital market in Thailand

Thailand is a developing country that needs capital to invest in various industries and businesses, while the amount of savings in the country is insufficient due to many factors, making capital markets as a source of long-term funding to play a role in raising funds for investment and savings for the public, with the commercial banking group as the main business group of the national economy [18].

In 1997, Thailand was faced with the Thai baht currency crisis, resulting in the financial institutions of Thailand revised regulations and laws to meet international standards, and has given opportunities for foreigners to allow the Thai commercial banking business [18].

In paper's Apinya Wanaset [19] is to study the relationship between the capital market and economic growth in Thailand after the financial crisis in 1997. By using three methods which are unit root test, cointegration test, and causality test to find the relationship from the data between 1997 to 2016. The results show that economic growth has a positive impact on the stock market and the bond market. At the same time, the stock market also supports economic growth by allocating funds to private investment, which leads to the expansion of production and employment. However, the bond market does not affect economic growth. But it helps to push the stock market expansion by acting as investment management and diversification option for investors.

C. Comparative theory

The theory that people are generally risk-seeking on losses and risk-averse on profits, known as "reflection effect". In study's Yang-Yu Liu [20] aims to explore the empirical aspects from the reflection effect theory by analyzing and comparing the behavior of winning and losing trades from more than 28.5 million trades made by 81.3 thousand traders of an online financial trading community over 2 years and 4 months and by using the results to analyze the increase of online social trading opportunities. The results of the study show that the availability of changes in financial trading behavior is subject to individual human decisions, which means that trading efficiency will be significantly improved when humans understand investments.

Minimum Spanning Tree is an independent tree that connects every vertex in the graph, with the shortest path for

creating Prim's Algorithm. It is applied to find the relationship of assets by using the distance between assets in the research's Paolo Giudici (2019) [11] to study and understand the cryptocurrency property price changes, especially the price information during the exchange in the Bitcoin market among themselves and between markets Bitcoin and traditional markets. They have adopted the minimum spanning tree (MST) method which is used as a tool for studying the relationship between the digital asset market and the traditional market. When analyzing direct relationships, the results show that the relationship between the digital asset market does not have a direct relationship with the traditional market. In this research, there is no analysis of inverse relationships which may be one of the variables that will show the relationship between the two types of market.

The comparison between digital assets and original assets to determine their correlation was studied in paper's Jiaqi Liang et al. [21] aims to offer analytical insights to help understanding cryptocurrency by treating it as a financial asset. They position cryptocurrency by comparing their dynamic characteristics with two traditional and massively adopted financial assets: foreign exchange and stock. Based on the daily close prices for four years, they first construct the correlation matrix and asset trees of all the three markets, then conduct comparisons on five properties: volatility, centrality, clustering structure, robustness, and risk. From the analysis of the dynamic characteristics of the market cryptocurrency and comparison between the two traditional financial markets, foreign exchange markets, and stocks through experiments, they found that the correlation matrix and asset chart are powerful tools for market analysis. Cryptocurrency as well as used in analyzing the foreign exchange market and stock market

III. DATA

This research will focus on the relationship between the capital markets in Thailand and the digital asset market, by using daily asset prices for 3 years between 1 January 2016 and 31 December 2019. The datasets that used in this research are the capital market in Thailand including Tourism index, mutual fund SET50, Gold, Oil, THB/USD exchange rate, obtained from the Tradingeconomics's website and Bank of Thailand database. For the digital asset market including Bitcoin, Ethereum, Litecoin, Ripple, DASH, and Stellar, there are daily digital asset prices that collected from finance yahoo website (https://finance.yahoo.com/)

IV. METHODOLOGY

The data analyzed in this research are varied in data and sources. This data may be contaminated by noise, therefore, the random matrix theory (RMT) was used to filter noise from the correlation matrix to make the data more accurate. We use *PG + Filtering* methods [22] in this process. After that we display the relationship of assets in two ways, direct and inverse relationship, through networks by using a minimum spanning tree (MST).

Let $P_{i(t)}$ and $R_{i(t)}$ be the price and the return of an asset "*i*" at the time "*t*", respectively, where $R_{i(t)}$ defined as follows:

$$R_{i(t)} = logP_{i(t)} - logP_{i(t-1)} \quad (1)$$

Let a sequence $(R_{i(t)})$ be denoted by R_i. The elements of correlation matrix C can be obtained from the following formula

$$C_{i,j} = \frac{E(R_i R_j) - E(R_i)E(R_j)}{\sigma(R_i)\sigma(R_j)} \quad (2)$$

Where $C_{i,j}$ is Pearson correlation coefficient between asset *i* and asset *j*, *E* is the mean value, σ is the standard deviation.

Let $Q = L/N$ where *N* is the number of assets and *L* is the number of days. We call λ a noisy eigenvalue if $\lambda_- \leq \lambda \leq \lambda_+$ where λ_+ is the maximum eigenvalue and λ_- is the minimum eigenvalue, is given by.

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (3)$$

Let $D_{filtered}$ be a matrix obtained by replacing all noisy eigenvalues of eigenvalue matrix of *C* by zero, and *E* be an eigenvectors matrix of *C*. The correlation matrix after filtered $C_{filtered}$ can be obtained by replacing the diagonal of $ED_{filtered}E^{-1}$ by 1 (see [22, A.2] for more details).

To represent the direct relationship and inverse relationship between each asset, we create networks with each node as an asset and drawing a distance line between the nodes. The distance between node for direct relationship will be defined as follows:

$$D^+_{i,j} = \sqrt{2 - 2C_{i,j}} \quad (4)$$

Note that the result of this formula has a range between 0 and 2 and if $D^+_{i,j}$ close to 0, it means that data set *i* and data set *j* have a strong direct relationship.

In the same fashion, we define the distance between node for inverse relationship by

$$D^-_{i,j} = \sqrt{2C_{i,j} + 2} \quad (5)$$

Note that the result of this formula has a range between 0 and 2 and if $D^-_{i,j}$ close to 0, it means that data set *i* and data set *j* have a strong inverse relationship.

After that, the minimum spanning tree (MST) method is applied to these networks.

V. RESULT

From the previous section, the original correlation matrix *C* between return of each asset can be found in Figure 1. The closer of the value to 1 the stronger of direct relationship between financial assets. On the contrary the closer of the value to -1 the stronger of their inverse relationship between assets.

	SET50	Tourism	Vayupak	GOLD	Oil	THB-USD	BTC	ETH	LTC	XLM	XRP	DASH
SET50	1	-0.4936444	0.9573160	0.4142216	0.9049003	-0.8332192	0.8141923	0.7075056	0.7471151	0.7201127	0.6698209	0.5804724
Tourism	-0.4936444	1	-0.4554432	-0.6394518	-0.5163210	0.8068773	-0.5055604	-0.0472855	-0.2419805	-0.1353158	-0.1696156	0.0764173
Vayupak	0.9573160	-0.4554432	1	0.2977953	0.9150829	-0.7435983	0.6723214	0.5693232	0.6073305	0.6180514	0.5437102	0.4461080
GOLD	0.4142216	-0.6394518	0.2977953	1	0.2592968	-0.7240668	0.5055722	0.1816031	0.2817219	0.0743305	0.1746860	0.0687081
Oil	0.9049003	-0.5163210	0.9150829	0.2592968	1	-0.7533827	0.7118812	0.5880437	0.5947880	0.6512687	0.5627008	0.4491432
THB-USD	-0.8332192	0.8068773	-0.7435983	-0.7240668	-0.7533827	1	-0.8020897	-0.5022079	-0.6142514	-0.4883094	-0.5024438	-0.3484960
BTC	0.8141923	-0.5055604	0.6723214	0.5055722	0.7118812	-0.8020897	1	0.7582869	0.8768070	0.6877420	0.7255774	0.7175632
ETH	0.7075056	-0.0472855	0.5693232	0.1816031	0.5880437	-0.5022079	0.7582869	1	0.9031607	0.8707535	0.8671070	0.8957773
LTC	0.7471151	-0.2419805	0.6073305	0.2817219	0.5947880	-0.6142514	0.8768070	0.9031607	1	0.7983780	0.8480202	0.8814907
XLM	0.7201127	-0.1353158	0.6180514	0.0743305	0.6512687	-0.4883094	0.6877420	0.8707535	0.7983780	1	0.9010340	0.7295313
XRP	0.6698209	-0.1696156	0.5437102	0.1746860	0.5627008	-0.5024438	0.7255774	0.8671070	0.8480202	0.9010340	1	0.8243755
DASH	0.5804724	0.0764173	0.4461080	0.0687081	0.4491432	-0.3484960	0.7175632	0.8957773	0.8814907	0.7295313	0.8243755	1

Fig. 1. Correlation matrix between return

To use PG + Filtering methods, we compute Q by using *N* = 12 assets, and *L* = 1461 days. Thus *Q* = 132.818, and we found minimum eigenvalue (λ_-) = 0.826956 and maximum eigenvalue (λ_+) = 1.189471 from Equation (3). Finding eigenvalues matrix from correlation matrix *C* and replaced the noisy eigenvalues within the bound [λ_-, λ_+] with zero. Then the correlation matrix after filtered $C_{filtered}$ can be obtained by

using the method in section 4. We found that the filtered correlation matrix $C_{filtered}$, as shown in Figure 2, is slightly changed from original correlation matrix C.

	SET50	Tourism	Vayupak	GOLD	Oil	THB-USD	BTC	ETH	LTC	XLM	XRP	DASH
SET50	1	-0.51041	0.844716	0.544843	0.793827	-0.85344	0.864635	0.737687	0.798417	0.697259	0.700565	0.629992
Tourism	-0.510408	1	-0.48635	-0.60359	-0.54681	0.801326	-0.49171	-0.039	-0.2279	-0.14159	-0.16118	0.090012
Vayupak	0.8447165	-0.48635	1	0.538658	0.710266	-0.78089	0.765337	0.624977	0.70193	0.57591	0.600401	0.537421
GOLD	0.5448429	-0.60359	0.538658	1	0.496895	-0.68081	0.39767	0.117041	0.171981	0.123217	0.108922	-0.03722
Oil	0.7938267	-0.54681	0.710266	0.496895	1	-0.79017	0.803636	0.642944	0.688106	0.609698	0.618623	0.539218
THB-USD	-0.853442	0.801326	-0.78089	-0.68081	-0.79017	1	-0.78538	-0.49221	-0.59726	-0.49588	-0.49226	-0.3321
BTC	0.8646351	-0.49171	0.765337	0.39767	0.803636	-0.78538	1	0.733355	0.834428	0.706621	0.700181	0.676657
ETH	0.7376872	-0.039	0.624977	0.117041	0.642944	-0.49221	0.733355	1	0.877804	0.882049	0.851911	0.871302
LTC	0.7984171	-0.2279	0.70193	0.171981	0.688106	-0.59726	0.834428	0.877804	1	0.817578	0.822191	0.839887
XLM	0.6972593	-0.14159	0.57591	0.123217	0.609698	-0.49588	0.706621	0.882049	0.817578	1	0.91254	0.748064
XRP	0.7005646	-0.16118	0.600401	0.108922	0.618623	-0.49226	0.700181	0.851911	0.822191	0.91254	1	0.799444
DASH	0.6299918	0.090012	0.537421	-0.03722	0.539218	-0.3321	0.676657	0.871302	0.839887	0.748064	0.799444	1

Fig. 2. Filtered correlation between digital asset market and capital market in Thailand

From Figure 2, the direct distance relationship between assets can be determined from Equation (4) and the distance can be drawn as a network map and use minimum spanning tree to display these relations. The closer the distance between assets, the strong relationship shown in Figure 3.

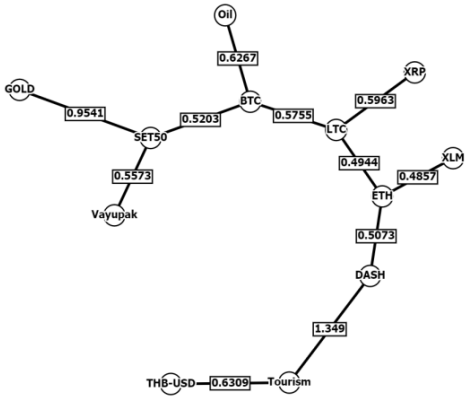


Fig. 3. Minimum spanning tree of direct relationship network

From Figure 3, we found that the distance between ETH and XLM is the shortest, showing the most direct relationship. On the other hand, we found that the distance between DASH and Tourism is the longest, showing that there is no relation. The shortest distance of each asset that show in Figure 3 is displayed in Figure 4.

			Distance
1	ETH	XLM	0.4857
2	LTC	ETH	0.4944
3	ETH	DASH	0.5073
4	BTC	SET50	0.5203
5	Vayupak	SET50	0.5573
6	BTC	LTC	0.5755
7	LTC	XRP	0.5963
8	BTC	Oil	0.6267
9	Tourism	THB-USD	0.6309
10	GOLD	SET50	0.9541
11	DASH	Tourism	1.3490

Fig. 4. The shortest distance between assets that show in Figure 3

From Figure 4, we can estimate the direct relationship between assets, each asset has an equal level correlation, except Tourism that has no relation to DASH since the distance is close to $\sqrt{2}$.

From Figure 2, the inverse distance relationship between assets can be determined from Equation (5). The distance can be drawn as a network map and use a minimum spanning tree to find the shortest distance between assets shown in Figure 5.

From minimum spanning tree, the shortest distance of each asset that shows inverse relationship when the value close to 0 in Figure 6.

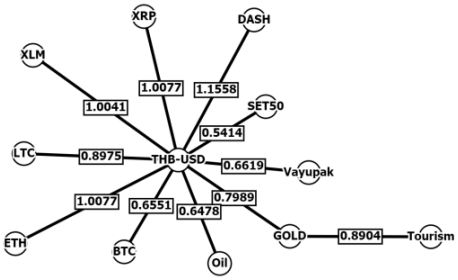


Fig. 5. Minimum spanning tree of inverse relationship network

From Figure 5, we found that the distance between THB-USD exchange and SET50 is the shortest, showing the most inverse relationship. As well as the distance between THB-USD and Oil, BTC, Vayupak fund showing an inverse relationship. The shortest distance of each asset that show in Figure 5 is displayed in Figure 6.

			Distance
1	THB-USD	SET50	0.5414
2	THB-USD	Oil	0.6478
3	THB-USD	BTC	0.6551
4	THB-USD	Vayupak	0.6619
5	THB-USD	Gold	0.7989
6	Gold	Tourism	0.8904
7	THB-USD	LTC	0.8975
8	THB-USD	XLM	1.0041
9	THB-USD	XRP	1.0077
10	THB-USD	ETH	1.0077
11	THB-USD	DASH	1.1558

Fig. 6. The shortest distance between assets that show in Figure 5

From Figure 6, we found that the inverse relationship between assets is weakly correlated, and we found that THB-USD is weakly related to most of the assets in this research.

VI. CONCLUSION

This research paper contributes evidence to academy that there is a link between cryptocurrencies asset price and some classical assets' market price. From three years daily data of Thailand asset prices in digital assets and conventional assets in classical capital market, we found strong direct relationship not only among cryptocurrencies in digital market but also with SET50 index and oil price in classical markets. Interestingly, the relationship among Ethereum, Stellar, Litecoin and DASH are even stronger than between mutual fund and SET50 index. This could imply that relationships between digital assets are stronger than between conventional assets, for example between mutual fund and stock market index, which are highly positively correlated. This phenomenon that digital assets show stronger relationship than conventional assets could be from the differences in trading platform or other factors, which is interesting for further investigation.

The strong direct relationships between Bitcoin and SET50 index, and between Bitcoin and oil price reveal the connection between digital asset market and classical markets. Digital assets could be additional options for investment to diversify the risk of investment portfolio. The inverse relationship between THB-USD exchange rate and Bitcoin and SET50 index could imply the relationships between foreign fund flow and investments in stock market and digital asset markets, especially Bitcoin. But for other digital assets, the weak signal might reveal that the trading is mostly among local investors and not much from foreign investment. The further investigation to prove these suspicious hypothesis could be interesting research papers.

ACKNOWLEDGMENT

The authors are very grateful acknowledge the support from International College of Digital Innovation, Chiang Mai University for all support for improving the paper.

REFERENCES

[1] Nakamoto, S., "Bitcoin: A peer-to-peer electronic cash system," www.bitcoin.org, 2008

[2] Andy Greenberg, "Crypto Currency," Forbes, 2019.

[3] Zixuan Zhang, Michael Zargham and Victor M. Preciado, "On modeling blockchain-enabled economic networks as stochastic dynamical systems," Springer, 2020.

[4] A.S. Hayes., "Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin" Telematics and Informatics. 2017.

[5] Natnicha tangwattanarat. "a study of the perception of Thai cryptocurrency investors towards digital currency market," 2017.

[6] Atiphat Rojanawutthitukun. "The determinants on performance of banking-sector stocks," 2011.

[7] Fama, E. F. and French, K.R. (1993) 'Common risk factors in the return on stocks and bonds', Journal of Financial Economics, Vol. 33, No. 1, pp.3–56.

[8] Naktnasukanjn, N., Chunhachinda, P. and Padungsaksawasdi, C. (2018) 'Multifactor asset pricing model incorporating coskewness and cokurtisis: the evidence from Asian mutual funds', Int. J. Economics and Business Research, Vol. 16, No. 3, pp.308-325.

[9] Conrad, C., Custovic A., and Ghysels, E., "Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis," Journal of Risk and Financial Management. 2018.

[10] Ciaian, P., and Rajcaniova, M., "Virtual relationships: Short-and long-run evidence from bitcoin and altcoin markets," Institutions and Money. 2018.

[11] Giudici, Giancarlo, Alistair Milne, and Dmitri Vinogradov, "Cryptocurrencies: Market analysis and perspectives.," Journal of Industrial and Business Economics, 2019.

[12] Mantegna., "Hierarchical structure in financial markets," The European Physical Journal BCondensed Matter and Complex Systems. 1999.

[13] Tola,V., Lillo, F., Gallegati, M.,&Mantegna, R.N., "Cluster analysis for portfolio optimization," Journal of Economic Dynamics and Control. 2008.

[14] Alen Hrga, Federico-Matteo Ben˘ci´c and Ivana Podnar˘Zarko. "Technical Analysis of an Initial Coin Offering," Institute of Electrical and Electronics Engineers Inc. 2019.

[15] Hassani, Hossein, Xu Huang, and Emmanuel Sirimal Silva. "Blockchain and Cryptocurrency. In Fusing Big Data, Blockchain and Cryptocurrency," Cham: Palgrave Pivot. 2019.

[16] Somin, S. "Network Dynamics of a Financial Ecosystem," 2020.

[17] Giudici, Giancarlo, Alistair Milne, and Dmitri Vinogradov. "Cryptocurrencies: Market analysis and perspectives," Journal of Industrial and Business Economics. 2019.

[18] Atiphat Rojanawutthitukun. "The determinants on performance of banking-sector stocks," 2011.

[19] Apinya Wanaset., "The Relationalship between Capital Market and Economic Growth in Thailand," Journal of Economic and Management Strategy. 2018.

[20] Yang-Yu Liu, Jose C. Nacher, Tomoshiro Ochiai, Mauro Martino and Yaniv Altshuler., "Prospect Theory for Online Financial Trading," Public Library of Science. 2014.

[21] Jiaqi Liang et al. "Towards an Understanding of Cryptocurrency: A Comparative Analysis of Cryptocurrency, Foreign Exchange, and Stock," 2019

[22] Vasiliki Plerou et al. "A Random Matrix Approach to Cross-Correlations in Financial Data," American Physical Society. 2002.

[23] Quoc Khanh Nguyen. "Blockchain – A Financial Technology For Future Sustainable Development," Institute of Electrical and Electronics Engineers Inc. 2016.

[24] Otabek Sattarov. "Recommending Cryptocurrency Trading Points with Deep Reinforcement Learning Approach," 2020.

[25] Gandal, Neil. "Competition in the cryptocurrency market. Bank of Canada Working Paper," 2014.

[26] Arianna Agosto. "Financial Bubbles: A Study of Co-Explosivity in the Cryptocurrency Market," 2020.

Simulation of Autonomous Mobile Robot System for Food Delivery in In-patient Ward with Unity

Supachai Vongbunyong
Institute of Field Robotics
King Mongkut's University of Technology Thoburi
Bangkok, Thailand
supachai.von@kmutt.ac.th

Salil Parth Tripathi
School of Engineering
IMT Atlantique Bretagne-Pays de la Loire
Nantes, France
Parth.tripathi@imt-atlantique.net

Kitti Thamrongaphichartkul
Institute of Field Robotics
King Mongkut's University of Technology Thoburi
Bangkok, Thailand
kitti.th@mail.kmutt.ac.th

Nitisak Worrasittichai
Institute of Field Robotics
King Mongkut's University of Technology Thoburi
Bangkok, Thailand
nitisak.w@mail.kmutt.ac.th

Aphisit Takutruea
Institute of Field Robotics
King Mongkut's University of Technology Thoburi
Bangkok, Thailand
aphisit.ta@mail.kmutt.ac.th

Teeraya Prayongrak
Institute of Field Robotics
King Mongkut's University of Technology Thoburi
Bangkok, Thailand
teeraya.ningg@gmail.com

***Abstract*—Logistic management is crucial for effective and efficient transportation of various items in hospitals. During pandemic situations, especially COVID-19, special in-patient cohort ward is established to treat patients who require special treatment due to the quarantine protocol. Autonomous Mobile Robot (AMR) is used for delivering food and medical supplies to individual patients in order to keep the physical distance between patients and health workers. In this research, delivery by using multiple AMRs working in the in-patient ward is simulated. The simulation software is developed in Unity platform to study the operations of AMRs in various scenarios.**

Keywords—Autonomous Mobile Robot, Simulation, Hospital Logistics, Automated Delivery, COVID-19

I. INTRODUCTION

A. Robot systems in COVID-19 situation

Logistics in hospital is a process designed for efficient and effective management of transportation of various types of items in hospitals. [1] classified the items into 7 groups, including, food and dietary, medical devices and supplies, pharmaceutical items, clothes, general housekeeping items, regulated medical waste, and general waste. Nowadays, the transportation is generally labor intensive which is carried out by skilled and unskilled health workers. These can be concerned a waste in the process and can be improved by using robotics and automation technology.

Autonomous Mobile Robot (AMR) is one of the feasible solutions to improve logistics capability. AMRs are generally used in industry to automatically transport items between specific locations. Fleet management software is generally used to control the operations of multiple robots to work cooperatively on shop floors in a factory in order to achieve the transportation tasks. Similar functions can be adapted to be used in hospital environment. A number of research works and commercial systems are available (see Section 2-A). However, the systems need to be customized according to the specific requirements of each hospitals.

During pandemic situation, especially COVID-19, implementation of mobile robots in hospital is not only for improving work efficiency but also for the infection control purpose. One of the infection control protocols to lower the risk of the health workers from the infectious patients is to reduce physical contacts between them. Mobile robots are one of the best solutions. They are equipped with particular extension for specific functions to serve various tasks in pandemic situation [2], e.g. tele-medicine, items delivery, UVC disinfection, general communication, and etc. The robots will operate in the special controlled in-patient ward while the doctors and nurses control the system from outside.

B. CARVER AMR system overview

In this research, CARVER-Cab (see Fig. 1), an AMR designed for bulk delivering items (i.e. food, medicine, and medical supplies) to individual patients are developed [3]–[5]. The AMR can automatically navigate to specific locations - load-in/out station and patient beds – in the cohort ward. The operation area is predetermined with CAD model of the floor plan. Currently, CARVER system has been installed and operated in at least four public hospitals in Bangkok, Thailand.

Fig. 1 Functions and features of CARVER-Cab

In this article, one cohort ward is selected as a case study. The objective of this project is to evaluate the performance and utilization of the system that consists of multiple robots

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

working together. The process simulation is carried out by using Unity, which is a real-time 2D-3D development platform [6]. The operational environment is dynamic as the state of items can change according to the assigned conditions. From the simulation, the performance and limitation of the system is tested in 2 scenarios with variations of the number of robots and service schedule. The result from the simulation will be discussed in this article.

This article is organized as follows: Process simulation, multiple mobile robot system, and development tools are reviewed in *Section II*. Methodology for system developing based on hospital environment is described in *Section III*. The development of process simulation is focused on *Section IV*. Results are in *Section V* and discussion in *Section VI*.

II. Literature Review

A. Fleet management of AMR/AGV and simulation

Fleet Management (FM) is a platform for managing multiple robots to optimize the operations of the robots, generally AMRs, to perform desired tasks. The performance is evaluated in several terms, e.g. time, efficiency, cost-effectiveness, and etc. FM is developed in many approaches. [7] worked on fleet schedules in material handling based on timetable execution with the Multimodal Transportation Network. [8] proposed local coordination strategy with Quadratic Programming for optimization the utilization of robots and time. [9] proposed centralized method to guarantee the safety when the communication is lost which is a one of the critical issues of FM. [10] developed algorithm for coordinating multiple robots by partitioning the working area to sectors and manage multiple robots. [11] proposed using Congestion Prevention Rule-based Bi-level Genetic Algorithm for optimizing the scheduling in automated container terminals to simultaneously operate components.

Actual environment can be complex and needs algorithms mentioned above to operate the system effectively. However, simulation in virtual environment is also required and conduct a priori in order to roughly evaluate the performance of the system. Software for simulation specifically configured for AMR in hospitals is developed in this research.

B. Software development with Unity?

Unity is a real-time 2D-3D development platform that is originally designed for game programming. One of the key advantages of Unity platform is available libraries that support a number of applications, not only for game development but also for engineering applications, e.g.[12]. In regard to the process simulation, a number of research and development works were carried out by using this platform, e.g. [13].

In this research, the simulation of AMR operations can be concerned a part of fleet management that controls and monitors multiple AMRs working together. An individual AMR can be treated as an *agent,* in which the behavior can be characterized with script assigned. Therefore, other types of robots and health works can be assigned for the further study.

In addition, Unity is equipped with powerful rendering engine. The simulation can be visualized for the end-users, e.g. doctors and nurses, so that the operation can be planned in advance. The platform also supports Augment Reality and Virtual Reality (AR/VR) which can be used for enhancing user experience on the simulation in the future. All of these features make Unity flexible for future modification improvement required in this project. It is eventually selected for developing simulation in this research.

III. Methodology

A. CARVER AMR spec & parameters

CARVER-Cab is an AMR specifically designed for bulk delivery. It is developed based on CARVER-AMR Platform. It is equipped with Lidar and other sensors for localization. It can automatically navigate point-to-point with ±30 mm precision in areas with known map. For the safety protocol, obstacle avoidance with Elastic Bands algorithm is used [14]. Battery can be recharged automatically at the station while the battery level is lower than the designed level.

TABLE I. Summary of specification of CARVER-AMR for Simulation

Item	Detail
Dimension (mm)	0.45 m x 1.0 m x 1.5 m
Mass	200 kg
k1 = Acceleration constant (m/s^2)	0.1 m/s^2
k2 = Vmax (m/s)	0.5 m/s
k3 = Average serving time at one bed (sec)	60 sec
k4 = Allowable minimum battery level (%)	4%
k5 = Max battery level (%)	100 %
k6 = Energy consumption rate (%/position)	-2% /position
k7 = Battery charging rate (%/hr)	+13% /hr
k8 = Obstacle avoidance radius (m)	0.1m

B. Simulation with Unity – NavMesh

The simulation is developed based on the actual robot's specifications (see TABLE I.) to reflect robot's behavior in reality. *NavMesh* module is used for navigation of the robot in Unity environment. *NavMesh* is a surface where the robot movement is allowed (see blue surface in Fig. 2). Therefore, the path calculation can be simplified as geometry space is limited by predetermined *NavMesh*. The *NavMesh* is calculated on the basis of (a) *robot size* and (b) *minimum distance from obstacles. NavMesh module* is used to generate the *NavMesh* with an obstacle avoidance radius of 0.1m (*k8*).

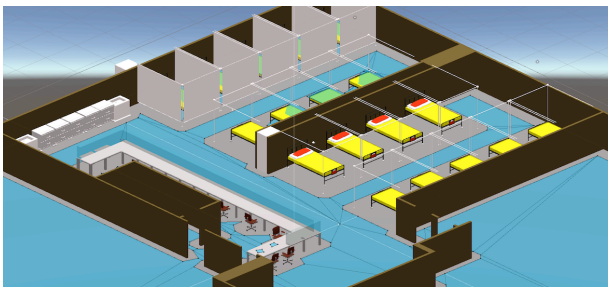


Fig. 2. *NavMesh* setup on hospital floor

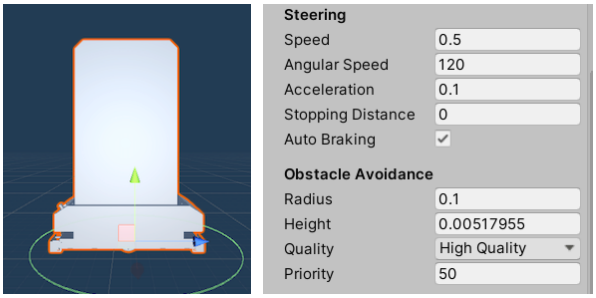


Fig. 3. 3D Model of CARVER-AMR and NavMesh Agent

IV. Implementation

A. Unity Platform and Hardware Setup

The simulation is developed on Unity 2019.3.11f1 x64 as a 3D project with *NavMesh* module added-on. Scripts are in C#. The simulation is eventually built as an standalone app for Windows 10. The computer used for developing the software is equipped with Core I7 8750M processor, 16 GB DDR4 RAM, and NVIDIA GTX 1650M graphic card.

B. CAD Modeling vs Actual Objects

The delivery operation takes place in an in-patient cohort ward in a hospital in Thailand. This 300 m^2 cohort has been modified to be a large negative pressure room for supporting COVID-19 patients (see layout, Fig. 4). The access of health workers are limited according to infection control protocol. The operating area in this room consisting of 18 beds divided in 2 rows separated by a wall, each row has 4-5 beds on each side of the main alley. The beds are separated by a removable curtain. The model of the room is shown in Fig. 5.

For the virtual environment, *beds* and all physical objects are included in the simulations. The curtain can be removed in case that larger workspace is required for robot movement. The *battery charging ports* are not presented as 3D objects but located virtually in the room in the battery charging area (see Fig. 6). To be noted that human staffs are excluded from this simulation due to the access restriction.

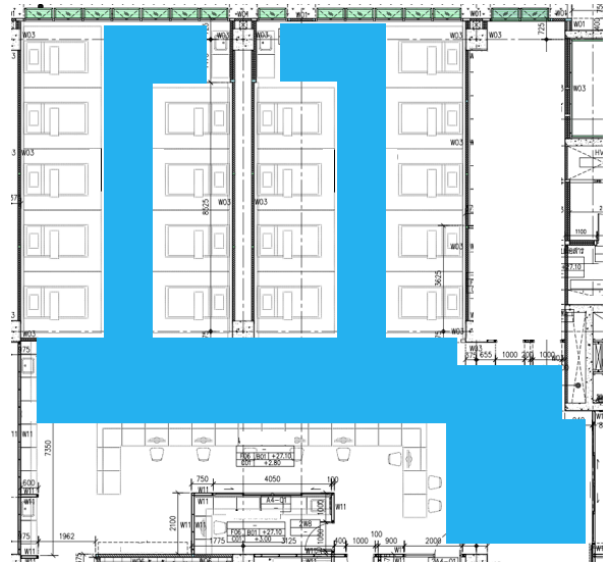


Fig. 4. Layout of the hospital in-patient ward with operational area

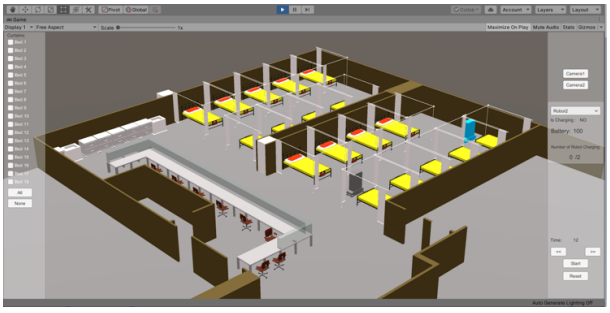


Fig. 5. 3D CAD Model of hospital ward from perspective view

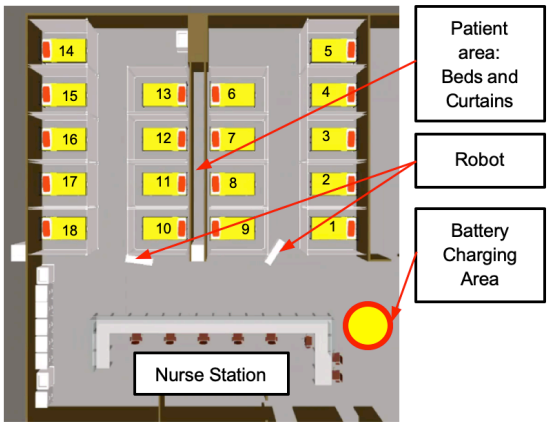


Fig. 6. 3D CAD Model of hospital ward from top view

C. Programming & Scripting

The scripts are attached to the objects to get affected. Therefore, the behavior of the robot can be controlled by using script. Multiples scripts are created for this simulation. They can be classified into 3 categories: (a) *Communication*, (b) *Graphical User Interface (GUI)*, and (c) *Core simulation*.

(a) **Communication scripts:** The goal is to import the user inputs and export the results. The *import* is done through a CSV file that is created by a Microsoft Excel/VBA program that we will detail in the next section The *export* is done in a simple text file. Different types of data are exported for each robot every second (in the simulation environment time, not real-time), including *Instant speed* and *Battery level*. This data is then exploited to get the results (see Section V).

(b) **GUI scripts:** The users can monitor and control the simulation from GUI. The configuration of the simulation can be set and run-time variables can be monitored via this GUI. The key variables are: the number of robots, specification parameters of the robot, battery level, battery charging status, states of the robot, and etc.

(c) **Core scripts** are the ones running the simulation. A multitude of them are needed to make the simulation work.

Robot Navigation scripts: are used to assign the target position for each robot. One script is used to obtain the user's defined parameters from CSV files, e.g. bed access schedule. One script is used to calculate the path that the robot needs to follow to reach the different positions. The path is calculated automatically by *NavMesh* module path calculator. The parameters are set for *NavMesh Agent* (see TABLE I). The path is calculated to be the shortest path possible. The robot moves 0.5 m/s at max. speed (*k2*) with 0.1 m/s^2 acceleration constant (*k1*). The robot stops for 60s (*k3*) whenever the robot reach the boundary within 0.1 m radius (*k8*) around target position. The navigation procedure is in a flowchart, Fig. 7.

Battery Level script is used to manage the battery level of each robot. It updates the battery level of every robot every time it reaches a position and every hour if it is charging. The recharging rate assumed to be linear with +13%/hr (*k7*) and the battery loss is averaged -2%/position reached (*k6*).

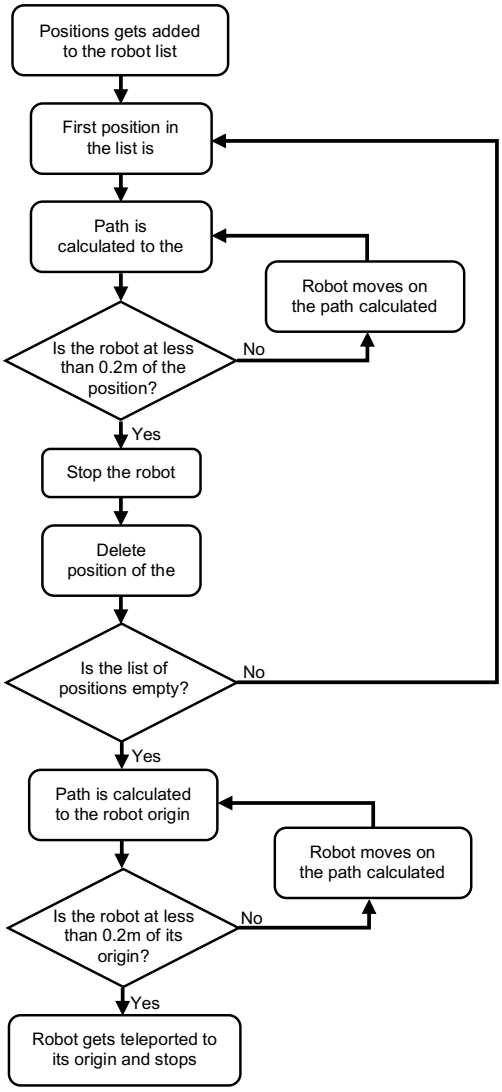


Fig. 7. Flowchart for navigation

D. Robot Operation with NavMesh Agent

The robots' operations are controlled by *NavMesh Agent* which will execute according to the following process.

First, the list of the target positions in order is given to the robot at the given hour. Then, the path calculation is done at every instant with the *NavMesh* path calculation algorithm. It constantly calculates the shortest path between the robot and the target position. It does so at every instant because new obstacles, including other robots, can appear along the robot path and it should make sure to avoid them.

Another important aspect of the simulation is the *schedule management* with respect to the user input data (which will be explained in Section IV-E). When the robots are on duty, *Robot Navigation script* is executed in order to move the robots between target positions and *Battery Level script* is executed for managing the battery usage. On the other hand, when the robots are not on duty, only *Battery Level script* is executed in order to manage the battery recharging. The robot will move to the charging ports as required.

E. Input CSV – Parameters for Scheduling Management

The simulation is required to run multiple times with different initial conditions. Directly modify the scripts or enter the input via GUI is too complicated. Therefore, the conditions are input via CSV file and VBA script is used to transfer the user inputs to the Unity program.

CSV file is designed similar to the actual CARVER-AMR is used to input the data by health workers. The file consists of two parts: (a) schedule table and (b) parameters settings of robot and simulation. This data will be read and set in the Unity program by the *Communication script.*

(a) *Schedule table:* the vertical axis is divided into slots ranging from 00:00 to 23:00 with 1-hour steps. For the horizontal axis, the beds number is presented. Therefore, the user can indicate which bed the meal will be served at which time slot (see Fig. 8).

(b*) Parameters settings*: These parameters need to be modifiable as the user will change in order to explore the variations of the fleet operation.

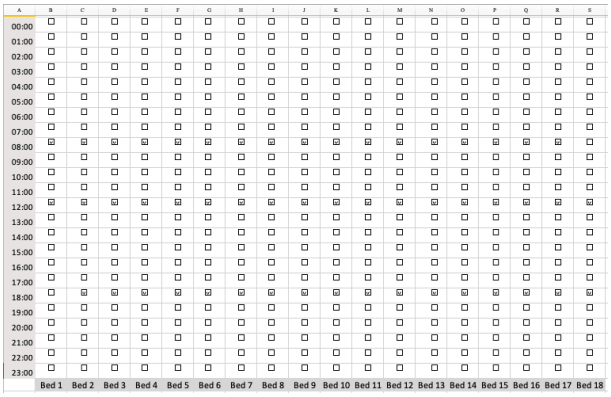


Fig. 8. Schedule setting in CSV file

Parameter 1: Number of robots: It goes from 1 to 5, it is the main parameter since it allows to know how the tasks are realized with the different number of robots. This parameter has been limited to 5 due to the size of the actual room. To be noted that, from the simulation results shown in Section V, no significant difference when more than 3 robots are used.

Parameter 2: Number of charging ports: This parameter is capped by the number of robots as it represents the number of robots that can charge simultaneously. It is an interesting parameter since the bottleneck in charging process will be occurred when the number of ports is less than the demand. The optimal number of ports can be found out

Parameter 3: Task sharing strategy: two strategies have been implemented. The strategy for fleet management can be studied from the variation of this parameters.

(i) To share the load between robots by giving the priority to the robots with the most battery. However, to minimize useless movements, the robot is assigned a minimum of 4 target positions (if there are at least 4 positions remained) before assigning positions to another robot (see Fig. 9).

(ii) Each robot is given all the positions until its battery level gets below 50%. Then the next robot in line takes this responsibility so on and so forth.

```
For r in G:
    L(r) = [ ]
    G = Sortexp(G)
EndFor
For p in H:
    If Batexp(G[0]) > Batexp(G[1]) OR  length(L(G[0])) < 4 :
        L(G[0]).append(p)
    Else:
        G = Sortexp(G)
        L(G[0]).append(p)
    EndIf
EndFor
```

Where:

L(r) is list containing the target positions of robot *r* that need to reach in the next run.

Bat(r) is the level of battery of the robot *r*.

Batexp(r) = *Bat*(*r*) – *length*(*L*(*r*)) *2

G is the list containing all the robots.

H is the list containing the positions to be assigned to robots.

Sortexp(G) sorts *G* in function of *Batexp* level of robots in *G*.

Fig. 9. Algorithm for load sharing between robots

V. Simulation Result

Two schedule scenarios were used in this simulation to study how the fleet behaved in different situations. The purpose is to study the time required to complete the delivery tasks and the needs on battery management strategy. Load sharing strategy is implemented. The graph between robots' speed versus time are plotted to observe the movement of the robot at each state (see Fig. 10 - Fig. 14). The robot moves between points with 0.5 m/s maximum speed. When the robot reaches the target position, the speed is 0 m/s.

A. Scenario 1

In the first scenario the robots were asked to reach the beds for 3 meals: 8:00, 12:00, and 17:00. This is considered the simplest and easiest scenario as the robots were able to achieve the tasks without having a battery level issue. A single robot was able to finish the task by itself without running out of battery as it has plenty of charging time to regain the battery lost between each run. However, it took almost half an hour to complete an entire round (see Fig. 10).

The simulations were repeated for 1-5 robots and 3 meals schedule. The graphs for 8:00 slots for 1-5 robots are shown in Fig. 10 - Fig. 14, respectively. The time usage to complete the task is summarized in TABLE *II*. To be noted that the time for other two time slots express in the same way; hence, only the 8:00 slot is selected to demonstrate the trend.

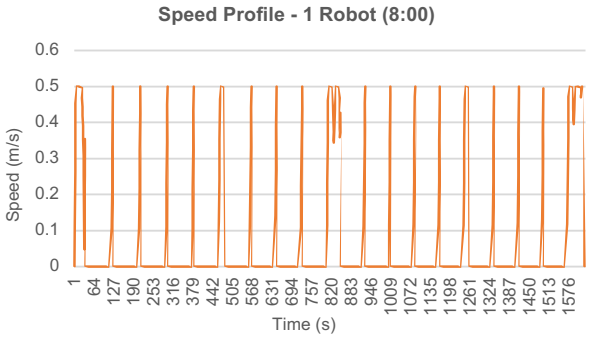


Fig. 10. Speed profile of 1 robot – Scenario 1 at 8:00

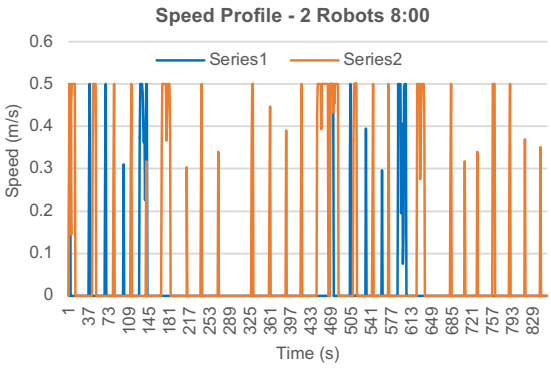


Fig. 11. Speed profile of 2 robot – Scenario 1 at 8:00

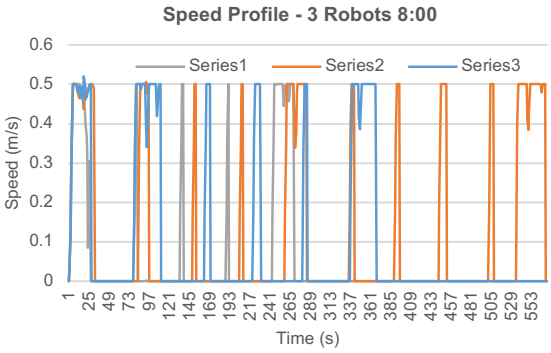


Fig. 12. Speed profile of 3 robot – Scenario 1 at 8:00

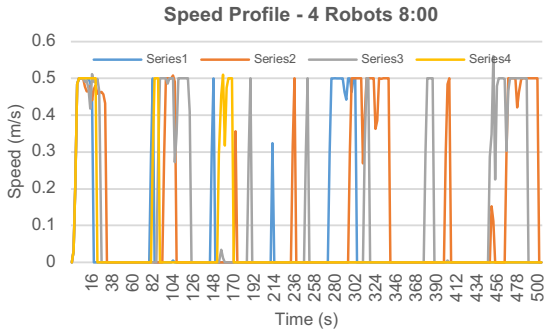


Fig. 13. Speed profile of 4 robot – Scenario 1 at 8:00

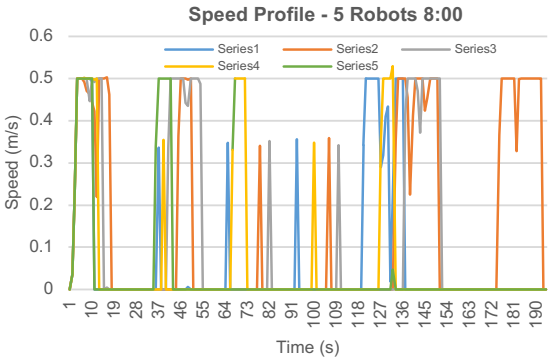


Fig. 14. Speed profile of 5 robot – Scenario 1 at 8:00

TABLE II. Simulation result of Scenario 1

No. of Robot	Time (s)						
	8:00	10:00	12:00	15:00	17:00	20:00	Avg.
1	1548	-	1429	-	1429	-	1468.7
2	817	-	780	-	723	-	773.3
3	545	-	731	-	732	-	669.3
4	504	-	239	-	205	-	316.0
5	188	-	188	-	177	-	184.3

B. Scenario 2

The second scenario refers to a busy day where the robots needs to reach every bed every few hours: 8:00,10:00,12:00, 15:00, 17:00, and 20:00. The results hour by hour are close to the first scenario. However, from the simulation result, a single robot was not able to achieve the task as the battery died out at 17:00. From the time shown in TABLE III, using 2 robots was sufficient to complete the task. The time usage reduced marginally when the number of robot increased. However, there is no difference between using 4 and 5 robots.

TABLE III. Simulation result of Scenario 2

No. of Robot	Time (s)						
	8:00	10:00	12:00	15:00	17:00	20:00	Avg.
1	1630	1522	1340	1390	-	-	1470.5
2	901	780	493	528	510	493	617.5
3	545	528	443	647	630	630	570.5
4	341	307	392	545	443	460	414.7
5	341	324	392	477	460	409	400.5

C. Discussion

From the simulation results, we have found out that only 1 robot was capable to complete the scenario-1 task in around 1468s (24 mins) and 2 robots were capable to complete the scenario-2 task in around 617s (10 mins). The time reduced when the number of robots increased. However, there was no more time reduction when the number of the robot exceeded the operations required, for example 4 robots in scenario-2 took around 7 mins to complete the task.

In regard to the Unity software, the limitation due to hardware dependent has been noticed. Indeed, if multiple simulations need to be launched, the need will arise to speed up the process (in this case, x50 speed up). However, without a superior hardware configuration, there will be problems to record data. Indeed, it does not take only 180 seconds for the 5 robots to complete a run, but the refreshing rate of the simulation on this hardware is too low to cope with the quick changes in the simulation. Consequently, the results can demonstrate whether the fleet is able to complete its task or not. However, the real-time data cannot be obtained correctly under the hardware with this configuration.

VI. Conclusion

Process simulation is a tool helping users to manage the process effectively and efficiently. In this case, AMRs are actually implemented in hospital logistics applications where food and medicines are delivered to specific patients in a COVID-19 cohort ward. In this research, Unity software is used to develop a problem-specific simulation program for study the operations of an AMRs fleet working in that room.

NavMesh module is used for developing this program. The *NavMesh Agent* is used to control the robots to imitate the actual AMRs. The movement paths are automatically generated with obstacle avoidance capability. The battery level and recharging process is managed by the a Unity script.

The simulation was run for two scenarios according to the food serving time for patients. The level of operation duty is different for two scenarios. Only 1 and 2 robots are sufficient for the operations in scenario-1 and scenario-2, respectively. The critical condition that makes the operation fail is the depletion of battery. The robots cannot complete the tasks if the recharging rate is not able to cope with the depletion. A more realistic model of battery recharging and consumption rate can improve the simulation result.

Acknowledgment

This project is partially funded and supported by (a) Project “ModBoriraksa” FACO: FIBO Against COVID-19, and (b) NXPO Office of Higher Education Science Research and Innovation Policy Council, Thailand, as part of project “Promotion of Robotics for All & Development of Robotics Innovators, Researchers, Engineers, and Startups”.

References

[1] M. Gunjan and L. Pankaj, “Service Robotics Market - Global Opportunity Analysis and Industry Forecast, 2014 - 2022,” 2016.

[2] R. Murphy, V. Gandudi, and J. Adams, “Reported Use of Robots (Ground, Aerial) Worldwide for COVID-19 as of 12 April 2020,” 2020.

[3] “Carver Robotic System,” 2020. [Online]. Available: www.facebook.com/CarverAMR.

[4] S. Vongbunyong, “Self-Identification System for Security in Personal Food Delivery for Inpatients with Autonomous Mobile Robot System,” 2003001083, 2020.

[5] S. Vongbunyong, “Autonomous Mobile Robot System Reducing Cross-Contamination between Persons in Food Delivery in Hospital Ward,” 2020.

[6] “Unity,” Unity Technologies, 2020. [Online]. Available: https://unity.com/.

[7] G. Bocewicz, I. Nielsen, and Z. Banaszak, “Automated guided vehicles fleet match-up scheduling with production flow constraints,” Eng. Appl. Artif. Intell., vol. 30, pp. 49–62, 2014.

[8] V. Digani, M. A. Hsieh, L. Sabattini, and C. Secchi, “A Quadratic Programming approach for coordinating multi-AGV systems,” IEEE Int. Conf. Autom. Sci. Eng., vol. 2015-October, pp. 600–605, 2015.

[9] A. Mannucci, L. Pallottino, and F. Pecora, “Provably safe multi-robot coordination with unreliable communication,” IEEE Robot. Autom. Lett., vol. 4, no. 4, pp. 3232–3239, 2019.

[10] L. Sabattini, V. Digani, C. Secchi, and C. Fantuzzi, “Hierarchical coordination strategy for multi-AGV systems based on dynamic geodesic environment partitioning,” IEEE Int. Conf. Intell. Robot. Syst., vol. 2016-November, pp. 4418–4423, 2016.

[11] Y. Yang, M. Zhong, Y. Dessouky, and O. Postolache, “An integrated scheduling method for AGV routing in automated container terminals,” Comput. Ind. Eng., vol. 126, no. July, pp. 482–493, 2018.

[12] T. Tothong, S. Vongbunyong, W. Chinthammit, and A. Piyasinchart, “Augmented Reality for Dies Alignment in Machine Setup,” IOP Conf. Ser. Mater. Sci. Eng., vol. 886, p. 012051, 2020.

[13] G. Corke, “Unity for manufacturing,” X3DMedia, 2020. [Online]. Available: https://develop3d.com/visualisation/unity-visualisation-vr-manufacturing-industrial-design-game-on-simulation/.

[14] S. Quinlan and O. Khatib, “Elastic bands. Connecting path planning and control,” Proc. - IEEE Int. Conf. Robot. Autom., vol. 2, pp. 802–807, 1993.

Myanmar POS Resource Extension Effects on Automatic Tagging Methods

Zar Zar Hlaing
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
zarzarhlaing.it@gmail.com

Ye Kyaw Thu
National Electronics and Computer Technology Center, NECTEC
Pathum Thani, Thailand
yktnlp@gmail.com

Myat Myo Nwe Wai
Faculty of Computer Science
Myanmar Institute of Information Technology, Mandalay, Myanmar
myatmyonwewai@gmail.com

Thepchai Supnithi
National Electronics and Computer Technology Center, NECTEC
Pathum Thani, Thailand
thepchai.supnithi@nectec.or.th

Ponrudee Netisopakul
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
ponrudee@it.kmitl.ac.th

***Abstract*—Part-of-speech (POS) tagging is the process of assigning the part-of-speech tag or other lexical class marker to each word in a sentence. It is also one of the most important steps in Natural Language Processing (NLP) task pipeline. There are several research works in Myanmar POS tagging implemented with different approaches. However, there is only one publicly available tagged corpus named myPOS corpus. The size of this corpus is only 11 thousand sentences. It is not enough to train downstream NLP tasks, such as machine learning. For this reason, we manually extended the original myPOS corpus as myPOS version 2.0 and the size of the extended corpus becomes approximately triple size of the original myPOS corpus. To evaluate the effects of the extended corpus versus the original corpus, the accuracies of four supervised tagging algorithms, namely, Conditional Random Fields (CRFs), Hidden Markov Model (HMM), Ripple Down Rules-based (RDR), and neural sequence labeling approach of Conditional Random Fields (NCRF^{++}) are compared. The results showed that the extended myPOS version 2.0 improved the accuracies of automatic POS tagging methods compared with the original myPOS.**

***Index Terms*—POS Tagging, Myanmar Language, CRFs, NCRF^{++}, HMM, RDR**

I. Introduction

POS tagging is the essential basis of Natural language processing (NLP). It is the process of tagging each word in a sentence with a corresponding POS tag such as noun, verb, adjective, adverb and so on. The POS information is very useful in many areas of NLP applications such as machine translation, information retrieval, sentiment analysis and word sense disambiguation systems, etc. In many Asian languages, such as Khmer, Lao, Thai, and Myanmar, POS tagging mainly depends on word segmentation. This is because these languages have no explicit word boundary. White space can be used for many purposes but not a symbol to separate subsequent words. Hence, the availability of resources such as POS-tagged corpus is an important issue. Although there are many available large POS tagging resources for English and other languages, there is only one available small POS tagged corpus [1] for Myanmar Language, namely myPOS with about 11 thousand sentences. Hence, many advanced machine learning models cannot be applied to NLP tasks in Myanmar. The objective of this study is to manually extend the original myPOS corpus to myPOS corpus version 2.0, which contains over 30 thousand tagged sentences and to study the effects of this larger size POS tagging corpus toward improving the accuracies of automatic tagging methods. In this paper, we compare the tagging accuracies of four methods, namely, Conditional Random Fields (CRFs), Hidden Markov Model (HMM), Ripple Down Rules-based (RDR), and neural sequence labeling approach of Conditional Random Fields ($NCRF^{++}$).

II. Related Work

For the performance of POS tagging, Htike et al. [1] evaluated and compared the accuracy of six POS tagging approaches (CRFs, HMM, MaxEnt, SVM, RDR and Two Hours of Annotation Approach) on a manually annotated 10K sentences POS tagged corpus. They defined 16 Myanmar POS tags for their work and showed the best result of RDR is 97.05 % on open-test data.

The author [2] compared HMM with post-editing rules and a purely rule-based approach for the POS tagging performance. They defined 36 POS tags and showed the results on increasing the number of training corpus size from 5,000 to 1 million words and gained the best result in term of accuracy in HMM with rules compared with the pure rule-based approach.

Myanmar Language Part-of-Speech (POS) tagging using deep learning models is also developed in [3]. In this work, Recurrent Neural Network (RNN) with Bi-directional Long Short-Term Memory (Bi-LSTM RNN) is applied in Myanmar Word segmentation and POS tagging process. GloVe was also used for performing the syllable vector representation and word vector representation. The author used 14 POS tags of myPOS [1] such as Abbreviation, Adjective, Adverb, Conjunction, Foreign Word, Interjection, Noun, Number, Particle, Post-positional Marker, Pronoun, Punctuation, Text number, Verb for the POS tagging process. The proposed system achieved better accuracy on 4K open test-set compared

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

with the POS tagging approaches CRFs, HMM and Max-Ent.

Myint et al. [4] applied bigram in POS tagging based on a very small training corpus that contains 1,000 sentences with the basic 14 POS tags (54 tags in detail). Although the test corpora contained close and open test data with many ambiguous words, their work gained 95.77% F-score on open test data.

Backpropagation Neural Network-based Myanmar POS tagged was developed by Hnin et al. [5]. They used manually tagged 5K sentences POS tagged corpus for training (82,892 words) defining 5 more tags over the traditional POS tag sets for Myanmar language. The training was done with 3-grams, 4-grams and 5-grams BPNNs on 3 different test sets and resulted in 80% F-score on open test data with 4-grams.

POS tags information was applied in Machine translation with an unsupervised novel bilingual infinite HMM approach [6]. The authors showed that this approach gained two points in BLEU for Myanmar to English translation.

III. Word Segmentation

In Myanmar texts, words are composed of single or multiple syllables that are usually not separated by white spaces. Spaces are used for easy reading and generally put between phrases, but there are no clear rules for using spaces in the Myanmar language. Therefore, word segmentation is needed before doing the POS tagging processes. A Burmese word can usually be identified by the combination of a root word, prefix and suffix. For example, a Myanmar word သွားသည် (go) can be segmented into two units: one is root verb, "သွား" and the other unit is a postpositional marker "သည်" , both combine to be one complete verb. Usually, a method such as Conditional Random Fields (CRF) is used for Myanmar word segmentation [7]. The example of segmented Burmese sentence, "You can't get nothing without risk." is shown as follow:

Unsegmented sentence: မင်းစွန့်စားမှုမရှိပဲဘာမှမရဘူး။
Segmented sentence: မင်း_စွန့်စား_မှု_မ_ရှိ_ပဲ_ဘာ_မှ_မ_ရ_ဘူး_။

Most of the Myanmar words are formed by one to three syllables and in the above example sentence, most words are formed by one syllable and then one word is formed by two syllables. Myanmar negative statements are constructed with the prefix မ (ma) as in negative imperatives and prohibitions and generally followed by suffix as we have shown in the above sentence မ + ရှိ + ပဲ (don't have) and မ + ရ + ဘူး (not get). It is similar to **ne + conjugated verb + pas** in the French language.

IV. POS Tag-set

In Myanmar language, there are 10 POS tags are defined by Myanmar Language Commission [8], [9]. These POS tags are Noun, Pronoun, Adjective, Adverb, Verb, Post-positional-market, Particles, Conjunction, Interjection and Punctuation. 16 Myanmar POS are applied in our tag set as used in the previous work [1]. The definitions and descriptions of POS tags are described in detail in Table I.

V. POS Tagging Methodologies

In this section, we will describe the methodologies of POS tagging processes used in the experiments in this paper.

A. Conditional Random Fields (CRFs)

Linear-chain conditional random fields (CRFs) [10] are models that consider dependencies among the predicted segmentation labels that are inherent in the state transitions of finite state sequence models and can incorporate domain knowledge effectively into segmentation. Unlike heuristic methods, they are principled probabilistic finite state models on which exact inference over sequences can be efficiently performed. The model computes the following probability of a label sequence $Y = \{y_1, \ldots, y_T\}$ of a particular character string $W = \{w_1, \ldots, w_T\}$.

$$P_\lambda(Y|W) = \frac{1}{Z(W)} exp(\sum_{i=1}^{T}\sum_{k=1}^{|\lambda|}\lambda_k f_k(y_{t-1}, W, t)) \quad (1)$$

where Z(W) is a normalization term, f_k is a feature function, and λ is a feature weight vector.

B. Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) is a probabilistic sequence model: given a sequence of units (words, letters, morphemes, sentences, whatever), it computes a probability distribution over possible sequences of labels and chooses the best label sequence [11], [12]. In an HMM for POS tagging, the observation is a sequence of words $\mathbf{o} = x_1, \ldots, x_n$ and is associated with a state sequence of POS tags that we cannot observe $\mathbf{s} = y_1, \ldots, y_n$. The model describes the joint state and observation sequence:

$$\begin{aligned} &p(y_1, \ldots, y_n, x_1, \ldots, x_n) \\ &= p(y_1)p(x_1|y_1)\prod_{i=2}^{n} p(y_i|y_{i-1})p(x_i|y_i) \end{aligned} \quad (2)$$

and the probability of the observation sequence can be obtained by marginalizing:

$$p(x_1, \ldots, x_n) = \sum_y p(x_1, \ldots, x_n, y_1, \ldots, y_n) \quad (3)$$

Here, the key assumptions of HMM are:

- The state sequence $p(y_i|y_1; \ldots; y_{i-1}) = p(y_i|y_{i-1})$ is Markovian
- The observations are conditionally independent of next and previous states and observations given the current state:

$$p(x_i|x_1, \ldots, x_n, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = p(x_i|y_i) \quad (4)$$

A graphical representation of a HMM can be seen in Figure 1.

TABLE I: Part-of-Speech Tag-set for Myanmar.

POS Tab	Brief Definition	Examples
abb	Abbreviation	အိုင်တီ (Information Technology), အ.လ.က (Basic Education Middle School)
adj	Adjective	လိမ္မာ (clever), ကြင်နာ (kind), ကျော်ကြား (famous)
adv	Adverb	လျင်မြန် (quick), တည်တည်ငြိမ်ငြိမ် (quietly)
conj	Conjunction	နှင့် (and), ထို့ကြောင့် (therefore), သို့မဟုတ် (or)
fw	Foreign word	Facebook, VOA, 1, 2, 3, Myanmar, ミャンマー (Myanmar in Japanese)
int	Interjection	အမလေး (Oh My God!)
n	Noun	သစ်ပင် (tree), ခဲတံ (pencil), ခဲဖျက် (eraser), မိန်းကလေး (girl)
num	Number	၁ (1), ၂ (2), ၃ (3), ၁၀ (10), ၁၀၀ (100), ၁၀၀၀ (1000)
part	Particle	များ (used to form the plural nouns as "-s , -es"), ခဲ့ (the past tense "-ed") သင့် (modal verb "shall"), လိမ့် (modal verb "will"), နိုင် (modal verb "can")
part_neg	Negative Particle: Particle that is used to form negative meaning of adjective and verb	မကောင်းပါဘူး (not well), မလုပ်နိုင်ဘူး (cannot do)
ppm	Post-positional Marker	သည်, က, ကို, အား, သို့, မှာ, တွင် (at, on, in, to)
pron	Pronoun	ကျွန်တော် (I), ကျွန်မ (I), သင် (you) , သူ (he), သူမ (she)
punc	Punctuation	။, ၊, (,), \, _, " "
sb	Symbol	?, #, &, %, $, €, ₹, π, λ, ÷, +, ×,
tn	Text Number	တစ် (one), နှစ် (two), သုံး (three), တစ်ရာ (one hundred), တစ်ထောင် (one thousand)
v	Verb	စား (eat), လေ့လာ (learn), နားထောင် (listen)

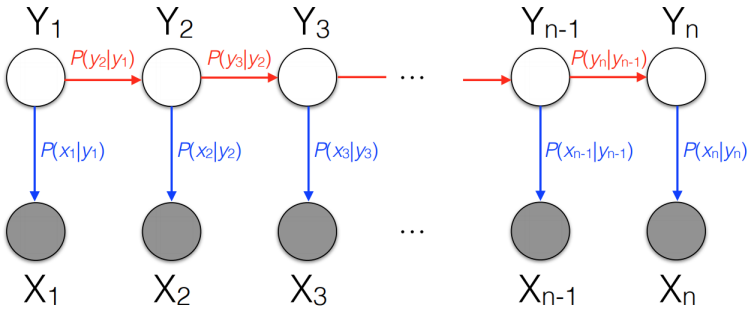


Fig. 1: Graphical representation of a Hidden Markov Model. Here, state variables $Y_1; Y_2; \ldots; Y_n$ form a Markov chain, but this sequence of variables is not observed (i.e. hidden). The $X1; X_2; \ldots; X_n$ are observable variables (i.e. output) of the Markov chain. Horizontal and vertical arrows indicate conditional dependence relations of variables.

C. Ripple Down Rules-based (RDR)

Ripple Down Rules (RDR) [13], [14] is a approach of building knowledge-based systems. RDRPOSTagger provides a failure-driven approach to automatically reconstruct transformation rules in the form of a Single Classification Ripple Down Rules (SCRDR) tree [15]. It can be written as *if X then Y* where X is condition and Y is the conclusion. Cases in SCRDR are defined by passing a case to the root (Rule 0 in Figure 2). At any node in the SCRDR tree (i.e. Rule 1 to Rule 6), if the condition of a node n met, the case is passed on to the exceptional child of n by using the *except* link if it exists. Otherwise, the case is passed on to the *if-not* child of n. A binary tree of Single Classification Ripple Down Rules (SCRDR) is shown in Figure 2.

D. Open-source Neural Sequence Labeling Toolkit ($NCRF^{++}$)

$NCRF^{++}$ [16] is a toolkit for neural sequence labeling. This toolkit can provide the users flexible running time and state-of-the-art results because it is a PyTorch[1] based framework with the customizable configuration file. $NCRF^{++}$ is designed for the rapid implementation of different neural sequence labeling models with a CRF inference layer. $NCRF^{++}$ can be regarded as a neural version of CRF^{++}, which is a famous statistical CRF framework. To build this toolkit, users will need the basic requirements of Python version 2 or 3 and PyTorch version 1.0. $NCRF^{++}$ supports different structure combinations on three levels: character sequence representation, word sequence representation and inference layer.

– Character sequence representation: character LSTM, character GRU, character CNN and handcrafted word

[1]http://pytorch.org/

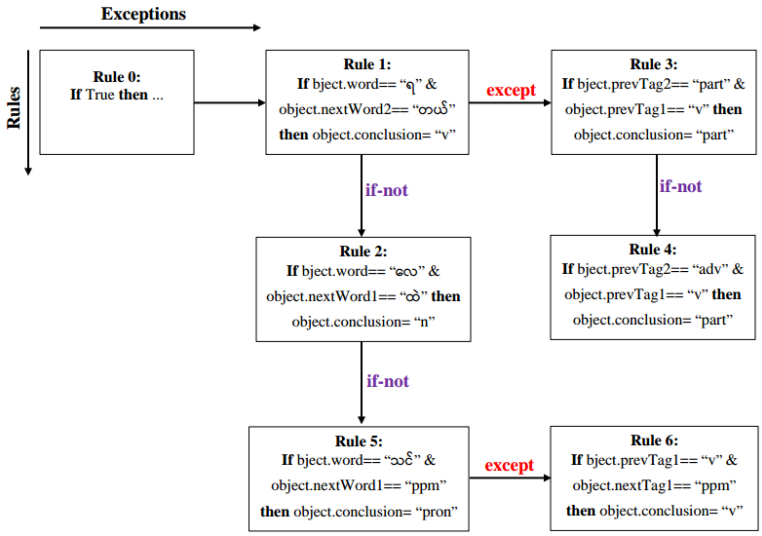


Fig. 2: A binary tree of Single Classification Ripple Down Rules.

features.
- Word sequence representation: word LSTM, word GRU, word CNN.
- Inference layer: Softmax, CRF.

Users can compare twelve neural sequence labeling models (charLSTM, charCNN, None ×wordLSTM, wordCNN × softmax, CRF). In the experiment of $NCRF^{++}$ in this paper, we compared two models : (1) CharLSTM+WordCNN+CRF (2) CharCNN+WordLSTM.

$NCRF^{++}$ framework is shown in Figure 3. It is designed with three layers: a character sequence layer; a word sequence layer and an inference layer. For each input word sequence, words are represented with word embeddings. The character sequence layer can be used to automatically extract word-level features by encoding the character sequence within the word. Word representations are the concatenation of word embeddings (red circles), character sequence encoding hidden vector (yellow circles) and handcrafted neural features (grey circles). Then the word sequence layer takes the word representations as input and extracts the sentence level features, which are fed into the inference layer to assign a label to each word. When building the network, users need only to edit the configuration file to configure the model structure, training settings and hyperparameters. Users can extend $NCRF^{++}$ by defining their own structure in any layer and integrate it into $NCRF^{++}$ easily.

VI. Experimental Setup

A. Corpus Extension

We extended the myPOS corpus of 11,000 sentences to 31,052 sentences from the Myanmar-Chinese parallel corpus (Extension-1) and Myanmar-Korea parallel corpus (Extension-2). Word segmentation, tagging with defined POS tags for each word and error checking was done manually. The shortest sentence in the corpus contained 2 words (for example: သူခိုး ။ , "thief"in English). The longest sentence of the current corpus contained 423 words. Sample sentences of these two corpora are as follows:

Myanmar-Chinese parallel corpus:
- "Happy" ကို တရုတ် လို ဘယ်လို ပြော လဲ ။ (How to say "Happy" in Chinese?)
- ပညာသင် ကာလ ၁ ခု ကို ကျောင်း လခ ဘယ်လောက် လဲ ။ (How much does it cost the tuition fee per semester?)
- ခဏ နေ မှ ပြန် တွေ့ မယ် ။ (See you soon.)

Myanmar-Korean parallel corpus:
- အလုပ် ပြီး ရင် မီး ပိတ် ပါ ။ (Turn off the light after doing work.)
- သတင်းစာ က နေ တဆင့် ကိုရီးယား ရဲ့ သတင်း ကို နားစွင့် နေ တယ် ။ (I am listening to Korean news through newspapers.)
- သူ့ ကို ကျေးဇူး ပြု ၍ ဒီ သတင်း ပို့ ပေး ပါ ။ (Please send this message to him.)

Firstly, we checked and corrected manually the segmented Myanmar sentences from our developing Myanmar-Chinese and Myanmar-Korean parallel corpora. These two-segmented corpora are then manually tagged using the POS tag-sets mentioned above and then checked repeatedly. In this paper, we defined the 10K POS tagged Myanmar sentences of *Myanmar-Chinese Parallel Corpus* as *Extension-1* and another 11K POS tagged Myanmar sentences of *Myanmar-Korean Parallel Corpus* as *Extension-2*. These two different domains POS tagged sentences are combined with the original myPOS corpus into the extended myPOS corpus version 2.0. The extended corpus has 31,052 sentences, which approximately triple size of the original myPOS corpus. The comparison of the original corpus and the extended corpus is shown in Table II.

B. Training Data and Test Sets

For the comparison of the original myPOS model and the extended myPOS version 2.0 model, we used two training data sets. One for the baseline myPOS model which is taken from the whole myPOS corpus (11,000 sentences) and another is also taken from the whole extended myPOS

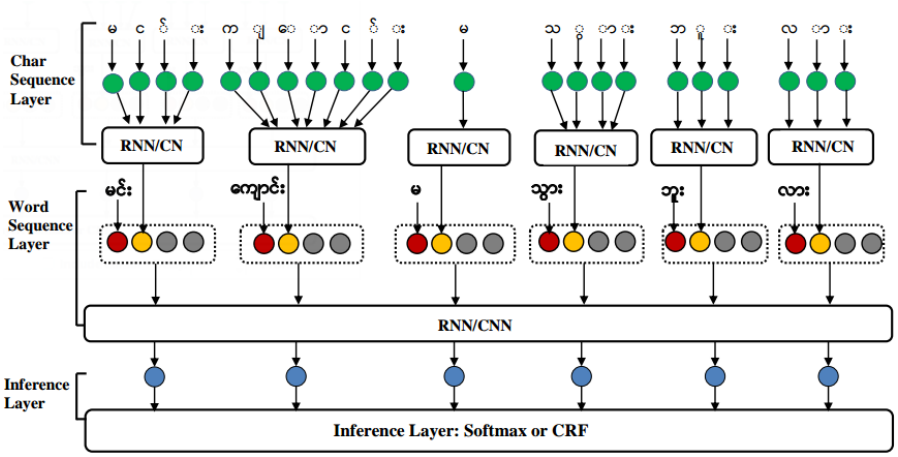


Fig. 3: $NCRF^{++}$ for Burmese sentence "မင်းကျောင်းမသွားဘူးလား" (Don't you go to school?) . Green, red, yellow and blue circles represent character embeddings, word embeddings, character sequence representations and word sequence representations, respectively. The grey circles represent the embeddings of sparse features

TABLE II: Data Comparison between Original Baseline Corpus and Extended Corpus (myPOS version 2.0)

Count	Original Corpus	Extended Corpus		
	myPOS	Extension-1	Extension-2	myPOS (version 2.0)
Total no. of Sentences	11,000	10,000	10,052	31,052
Total no. of Words	239,598	103,909	106,864	450,371
Average Words per Sentence	21.78	10.39	10.63	14.50

TABLE III: Comparison of accuracies between original myPOS and the extended version 2.0 among various automatic tagging methods

Methods	ctest:Closed Test-set		otest-1		otest-2	
	Original myPOS	myPOS (version 2.0)	Original myPOS	myPOS (Version 2.0)	Original myPOS	myPOS (version 2.0)
CRFs	**98.40 %**	98.19 %	91.69 %	**94.61 %**	91.79 %	**94.75 %**
$NCRF^{++}$(CharLSTM+WordCNN+CRF)	**96.50 %**	95.10 %	91.07 %	**91.65 %**	90.15 %	**94.64 %**
$NCRF^{++}$(CharCNN+WordLSTM)	**97.61 %**	97.40 %	93.14 %	**95.79 %**	93.23 %	**95.98 %**
HMM	**97.14 %**	96.44 %	93.66 %	**94.97 %**	93.78 %	**95.06 %**
RDR	**98.43 %**	98.35 %	95.10 %	**97.54 %**	95.15 %	**97.57 %**

version 2.0 (31,052 sentences) for the extended model. There are two types of test sets: closed test-set and open test-set. 50% of training data of original myPOS (5,500) is taken for closed test-set (ctest). Two open test sets defined as otest-1, otest-2. For the otest-1 and 2, we used sentences from the travel and tour domain ASEAN-MT corpus [17], which is not from original myPOS and our extended POS corpus. We randomly selected 6,072 sentences (50 % of ASEAN-MT corpus) for otest-1. The whole 12,144 POS tagged sentences of the ASEAN-MT domain data are used for otest-2 (i.e. big size open test data). Myanmar sentences from the Thai-Myanmar parallel corpus of the ASEAN-MT corpus are manually tagged. We thoroughly checked the manual POS tagged data to get the optimal open test data-sets.

C. Evaluation

To evaluate the effect of our extended POS corpus on automatic tagging task using various methods, we conduct the experiments on several test sets such as ctest, otest-1 and otest-2. The POS tagging performance was measured using the accuracy defined as follows:

$$Accuracy = \frac{number\ of\ correct\ POS-tags}{number\ of\ tokens\ in\ test\ corpus} \quad (5)$$

D. Software

The following open source POS Taggers were used for the experiments:

– CRFSuite: We used the CRFsuite tool (version 0.12) [18], (https://github.com/chokkan/crfsuite) for training and testing CRFs models. The main reason for this tool was its speed compared to other CRFs toolkits.

- $NCRF^{++}$ is a a toolkit for neural sequence labeling [16]. (https://github.com/jiesutd/NCRFpp) for quick implementation of different neural sequence labeling models with a CRF inference layer.
- Jitar (version 0.3.3): is a simple part-of-speech tagger, based on a trigram Hidden Markov Model (HMM). It (partly) implements the ideas set forth in [19]. Jitar is written in Java [20] and thus easy to use in other Java programs, or languages that run on the JVM.
- RDRPOSTagger (Version 1.2.3): is a rule-based Part-of-Speech and morphological tagging toolkit [14]. It is a robust, easy-to-use and language-independent toolkit.

VII. Results and Conclusion

The comparison of four automatic POS tagging accuracies based on different test-sets using the training model from original myPOS and extended myPOS version 2.0 is shown in Table III. When the two models are applied to the closed test-set from the original myPOS (ctest), the model trained from the original myPOS yields slightly higher accuracies for all methods. This should have no surprises since the test-set and the training set for building the model are from the same corpus; while the model trained from the extended version 2.0 yield comparable but slightly lower accuracies due to its wider scope of tagging ability. However, when the two models are applied to the open test-set outside the training corpus, the benefit of the extended myPOS version 2.0 becomes apparent. For all automatic tagging methods, the models trained from extended myPOS version 2.0 yield about 1 % to 3 % higher accuracies. This signifies that the tagging abilities of automatic POS tagging models trained from the larger training set perform better in general. Note that, when the test-set is larger (otest-2 compare to otest-1), the accuracies do not drop but actually yield slightly higher accuracies. In conclusion, the larger the training dataset for POS tagging, the better the model for any automatic tagging methods. Hence, the benefit of this huge effort to manually extend myPOS corpus is apparent from the evaluation experiments.

Acknowledgment

This work was supported by King Mongkut's Institute of Technology Ladkrabang and in collaboration with Artificial Intelligence Research Group, National Electronics and Computer Technology Center, Thailand.

References

[1] K. W. W. Htike, Y. K. Thu, Z. Zhang, W. P. Pa, Y. Sagisaka and N. Iwahashi, "Comparison of Six POS Tagging Methods on 10K Sentences Myanmar Language (Burmese) POS Tagged Corpus" , at 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017), Budapest, Hungary, April 17–23, 2017.

[2] K. Zin, "Hidden Markov Model with Rule Based Approach for Part-of-Speech Tagging of Myanmar Language", In Proceedings of the 3rd International Conference on Communications and Information Technology, pp. 123–128, 2009.

[3] T. T. Wai, "Myanmar Language Part-of-Speech Tagging Using Deep Learning Models" , International Journal of Scientific & Engineering Research ISSN:2229-5518, vol. 10, 2019.

[4] P. H. Myint, M. H. Tin and T. N. Lar, "Bigram Part-of-Speech Tagger for Myanmar Language", In Proceedings of 2011 International Conference on Information Communication and Management, pp. 147–152, 2011.

[5] H. M. Hnin, W. P. Pa and Y. K. Thu, "BackPropagation Neural Network Approach to Myanmar Part-of-Speech Tagging" , In Genetic and Evolutionary Computing - Proceedings of the Tenth International Conference on Genetic and Evolutionary Computing, ICGEC 2016, Fuzhou City, Fujian Province, China, pp. 212–220, November 7–9, 2016.

[6] Y. K. Thu, T. Akihiro, F. Andrew, S. Eiichiro and S. Yoshinori, "Unsupervised POS Tagging of Low Resource Language for Machine Translation" , In Proceedings NLP 2014, pp. 590–593, 2014.

[7] W. P. Pa, Y. K. Thu, A. Finch and E. Sumita, "Word Boundary Identification for Myanmar Text Using Conditional Random Fields" , Springer International Publishing, Cham, pp. 447–456, 2016.

[8] S. Lwin, "Myanmar - English Dictionary ", Department of the Myanmar Language Commission, Ministry of Education, Union of Myanmar, 1993.

[9] Myanmar Language Commission, Department 2005, Myanmar Thdda (2005), Department of the Myanmar Language Commission, Ministry of Education, Union of Myanmar.

[10] J. Lafferty, A. McCallum and F. C. N. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data" , In Proceedings of the Eighteenth International Conference on Machine Learning, ICML ' 01, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc, pp. 282–289, 2001.

[11] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition" , Proceedings of the IEEE, vol. 77, pp. 257–286, 1989.

[12] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" , Prentice Hall PTR, Upper Saddle River, NJ, USA, 1^{st} edition, 2000.

[13] Scheffer, Tobias, "Algebraic Foundation and Improved Methods of Induction of Ripple Down Rules" , pp. 23–25, 1996.

[14] Dat Q. Nguyen, Dai Q. Nguyen, D. D. Pham and S. B. Pham, "RDRPOSTagger: A Ripple Down Rules-based Part-Of-Speech Tagger" , In Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, Association for Computational Linguistics, pp. 17–20, 2014.

[15] D. Richards, "Two decades of Ripple Down Rules research" , Knowledge Eng. Review.24, pp. 159–184, 2009.

[16] J. Yang and Y. Zhang, "$NCRF^{++}$: An Open-source Neural Sequence Labeling Toolkit" , Proceedings of ACL 2018, System Demonstrations, pp. 74–79, July 2018, http://aclweb.org/anthology/P18-4013.

[17] P. Boonkwan and T. Supnithi, "Technical Report for The Network-based ASEAN Language Translation Public Service Project," Online Materials of Network-based ASEAN Languages Translation Public Service for Members, NECTEC, 2013.

[18] Okazaki, Naoaki, "CRFsuite: a fast implementation of Conditional Random Fields (CRFs)" , 2007.

[19] T. Brants, "TnT: A Statistical Part-of-speech Tagger" , In Proceedings of the Sixth Conference on Applied Natural Language Processing, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 224–231, April 2000.

[20] de Kok, Daniël, "Jitar: A simple Trigram HMM part-of-speech tagger" , 2014, [accessed 2016].

Mushroom Classification by Physical Characteristics by Technique of *k*-Nearest Neighbor

Narumol Chumuang[1], Kittisak Sukkanchana [2], Mahasak Ketcham[3], Worawut Yimyam[4],
Jiragorn Chalermdit[5], Nawarat Wittayakhom[6], Patiyuth Pramkeaw[7]

[1,2,5,6]Dep. of Digital Media Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand.
[3]Dep. of Management Information System, King Mongkut's University of Technology North Bangkok, Thailand.
[4]Dep. of Business Information Management, Phetchaburi Rajabhat University, Thailand.
[7]Media Technology Program, King Mongkut's University of Technology Thonburi, Thailand.

lecho20@hotmail.com[1], kittisak.oomsin@gmail.com[2], mahasak.k@it.kmutnb.ac.th[3], worawut_yimyam@hotmail.com[4],
jiragorncha@mcru.ac.th [5], nawaratwit@mcru.ac.th [6], patiyuth.pra@kmutt.ac.th [7]

***Abstract*— This paper proposed the principles of data analysis in order to present the prototype of mushroom classification based on physical characteristics. We created a model of mushroom classification by using Machine Learning (ML) with the mushroom dataset, comprising a total of 800 samples from the physical data of 22 attributes and it divide into two class as a toxic and non-toxic. The investigators designed the experiment in which 200 samples were randomly assigned to the mushroom population, consisting of 200 equally toxic and non-toxic mushrooms. For the quality, many ML were comparison such as Naïve Bayes Updateable, Naïve Bayes, SGD Text, LWL and K-Nearest Neighbor (k-NN). The results showed that K-NN gave the highest classification accuracy rate of 100%.**

Keywords*— *Mushroom, Classification, k-Nearest Neighbor, k-NN,

I. INTRODUCTION

According to the data of the Department of Disease Control [1],[2], it was found that at present. In the past 2017, there were 1,093 patients who had eaten poisonous mushrooms, 5 deaths [3]-[5]. Dr. Suwanchai. Wattana Yingcharoenchai Director-General of the Department of Disease Control said that during this period, many areas continued to have rain. Causing a large number of wild mushrooms in nature. Especially in the same area that used to collect wild mushrooms. Therefore people prefer to collect or buy wild mushrooms to cook for eating [3]. Especially people in the northeast and the north which each year will find patients and die from eating poisonous mushrooms that grow naturally on a regular basis. Misunderstood because wild mushrooms contain both edible and poisonous mushrooms which are very similar [6] .

Mushrooms are living organisms, classified in the Fungi Kingdom classified in Phylum Basidiomycota and Phylum Agaricomycota [7]. The mushroom is the reproductive structure of the gus, consisting of a stipe and a hat (pileus) under the hat, it can be a lamella or a tube in which the spore formed is very small, called microscopic, requires a microscope to help them see [8]. Currently, there are more than 30,000 types of mushrooms, including edible mushrooms, inedible mushrooms, and some poisonous mushrooms. Some have eaten and are fatal [9].

This paper focus to create a model for the classification of mushrooms which will be divided into toxin and mushrooms that are non- toxic to the body that will be examined from the characteristics of each mushroom species. That's or what kind of elements are and analyzed that it is a mushroom that can be eaten or is it a toxic mushroom.

The organize paper consists of the following topics: part 2 discusses related literature, part 3 research methodology, part 4 results of experiment and the last one is a conclusion.

II. LITERATURE REVIEW AND RELATED WORK

A. Mushroom

It is a mold that has evolved higher than other types. There are eukaryotic cells in the kingdom fungi, and most of them are in the phylum basidiomycota and phylum ascomycota [4]. Mushrooms cannot photosynthesize by themselves because they do not contain chlorophyll [8]. They grow into branched fibers called hypha, which clusters of these fibers are called mycelium. Most of them are in the phylum basidio mycota. sexual reproduction get a spore called besidiospor. The mushroom spores are formed and are attached to the spores called basidium, which line up on the gills or in pores. The am asco mycota produces a spore called ascospore in the sac called ascus.

The life cycle of all mushroom species is similar starting with the spores falling into various places and if blown into the right environment, it will grow into hypha [7]-[9]. The mycelium group will then gather into a group to form a mushroom. If the mushroom grows, it will be able to regenerate spores and be blown into hypha again. Mushrooms have different shapes. In general, the structure of the mushroom consists of a cap, a gill or lamella, a stalk, an annulus or ring, a volva, and a scale.

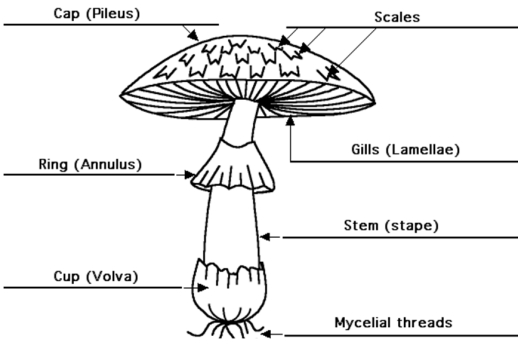


Fig. 1. Mushroom components.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

Mushrooms are both edible mushrooms and poisonous mushrooms [8]. There are many types of edible mushrooms and the well-known ones are straw mushroom, sajor-caju mushroom, golden needle mushroom, shiitake mushroom, abalone mushroom, the king oyster mushroom etc. These edible mushrooms are nutritious because they are good sources of fiber and vitamins, such as vitamin B1 (thiamin), vitamin B2 (riboflavin), vitamin B3 (niacin), etc. They also contain minerals such as potassium, phosphorus, magnesium and zinc. Mushrooms also contain a number of compounds that can be used as a medicine. There are many types of inedible mushrooms or poisonous mushrooms. Some are edible mushrooms in shape, but have a more showy color, such as those belonging to the Amanita genus. The toxin of this group of mushrooms is Amanitin or Amatoxin, which if eaten will affect the intestines and digestive system. It can cause dizziness, nausea, vomiting, and severe stomach pain or diarrhea.

Fig. 2.Example of the mushroom poisons.

B. k-Nearest Neighbor

Technique of the nearest neighbor algorithm to find the distance between each variable (attribute) in the data and then calculate the value [10]. This method is suitable for numeric data but a variable that is a non-continuous value can be done. Moreover ,it need more special handling as if it is a matter of color. What can we use to measure the difference between blue and green?, see in Fig.3. Next, we need a way to combine the distance of every attribute that can be measured when the distance between conditions or cases can be calculated, then a set of conditions to classify [11]. It become a base for classifying in new conditions yes, we can judge that the extent of the potential side is how big should be and it may be decided how to count the number of adjacent points. By the algorithm, the nearest neighbor has steps as follows [12]-[14]:

Nearest neighbor *k* algorithm] is a classification method that uses a method to find the distance between the characteristics of each data, which is suitable for numerical data by the algorithm. The steps are summarized as follows.

1) Set the number of neighbors *k* (preferably set to odd)

2) Calculate the distance of the data to be considered with the teaching data set which can be calculated from Euclidean distance equation is shown in eq. (1).

$$dist(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

where *dist (p, q)* represents the distance between *p* and *q* data, p_i refers to the value of the *i* property of the p and q_i data refers to the value of the *i* property of the *q* data.

3) Arrange the order of the ascending distance and select the least number of *k* sets the least data set.

4) Determine that the answer to the data to be predicted is the group that has the largest number of the first *k* of the data set.

5) Assign a *class* to the class nearest to the consideration point.

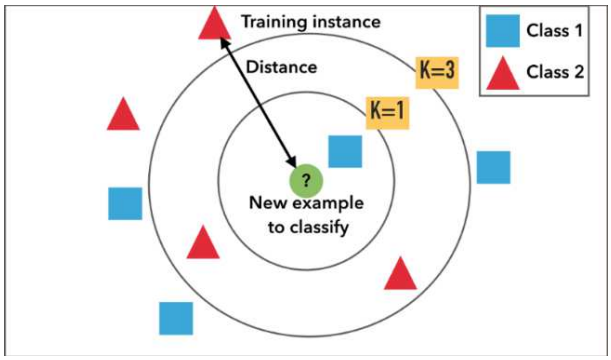


Fig. 3. Classification example with k-NN.

The main operation is to calculate the distance between two records. In order to measure the similarity of the data and a combination of the results obtained from the calculation of the distance by sorting the distance in ascending order, then look at the value of *k* that is set to what and take the sorted order compared with the sorted data class to answer. Distance Function Properties use to calculated distance. It must not be negative. If the same position, the function must be zero (the same value) and the round trip distance measurement must be the same. To calculate the distance function, add the absolute value to the distance value | AB | square the distance $(AB)^2$ and normalize it | (A-mean) / (SD) -. (B-mean) / (SD) |. To include the distance (the record) by measuring the distance Manhattan distance) is to combine the calculated values in one record. And Euclidean distances distance) is to find the square root in each variable (attribute) and put them together and bring the calculated values in one record together [15].

Nearest neighbor algorithm is an algorithm that is used to group information by arranging information that is close to each other as a class, this technique can make decisions that Which classes can replace new conditions or instances? By examining the number of *k*, if the conditions of the decision are complex [16]. This method can create an efficient model but the nearest neighbor algorithm takes a long time to calculate. If there are many variables (attribute), there will be problems in computation. And it consumes very high workloads on the computer because the time it takes for calculations will the factorial increases by the total number of points. Therefore, in order to increase the speed for the nearest neighbor algorithm techniques to be more all frequently used information must be stored in memory by the Memory-Based Reasoning method, which is frequently referred to [17]. To store the nearest neighbor algorithm class group in memory,

and if the data to be sought is few independent variables already will allow us to understand easier to model nearest neighbor algorithms, these variables are also useful for building models [18],]19]. Regarding non-standard data types, such as text, there may be a measurement standard for that type of data appropriate. In addition, the efficiency of this nearest neighbor algorithm will depend on the number of spacing explanation between the two that can effectively divide between normal data and unusual information describing the distance between data is a huge challenge when data is complex, such as graph data and sequential data [20].

C. Related work

K. Thirunavukkarasu, et al.,[21] proposed a method to classify plant species using appropriate agent selection method using the Genetic Algorithm and using the k-nearest neighbor technique to classify the type of plants. The samples used for this experiment consisted of 340 leaves of 30 plant species. The results were performed using the nearest neighbor technique alone, yielding 79.12% accuracy, 79.50% recall and 79.50% accuracy. Precision 73.17%, and when appropriate agent selection by genetic algorithm in conjunction with nearest neighbor technique results in plant classification accuracy 86.27%, recall 83.06% and the precision 84.06%.

T. Pinthong, et al., [22] presented techniques for recognizing letters on car license plates. And classify vehicle license plates in Thailand. The work process consists of three main processes: First one is detecting the location of the letters on the car license plate using the division of characters with a threshold method. Next is recognition of characters on the car license plate by using the median filter. Using the connected component label technique and *k*-NN for character recognition, and the last, car license plate classification in Thailand by using the color of the letters and license plate background for assortment of cars by means of the HSV color structure color sensing method from the results of 110 vehicle registration images, it was found that the proposed method was able to classify and recognize vehicle license plates at an average accuracy of 96.94% and 93.88% respectively.

X. Zheng, J. Liu and J. Liu, [23] presents the results of the comparison of the data classification of internet network services of higher education institutions by comparing the data classification efficiency of 3 techniques: Decision Tree technique, *k*-NN technique and Deep Learning technique. There are 8,530 sets of data that can be used in research studies by tertiary institutions. Cross-validation test by randomizing the data to divide the data was divided into 5 parts, each of which consisted of 1,706 data sets. The model was modeled from 4 data parts and tested the performance of the system with 1 part of the data. Techniques used to classify network service usage data internet of higher education institutions. The most effective is Deep Learning, with an accuracy of 88.79%, a memory of 88.55%, an accuracy of 89.10%, and a balance of 0.289, an acceptable assessment level.

K. Theresia Diah, A. Faqih and B. Kusumoputro, [24] Currently, communication via electronic mail (e-mail) is one of the most popular channels. Because it is easy to use, convenient, fast and free of cost. For this reason, it is a channel for those who do not wish well Used to send spam e-mail for other purposes, such as advertising or deception. This article presents a comparative study on the effectiveness of classified e-mail spam using data mining techniques. This consisted of a Decision tree, Naive Bayes, and K-Nearest Neighbor. Evna Eve Bay Provide the highest accuracy as 92.48%

C. Liou et al., [25] intended to compare performance how to choose suitable characteristics for classification Micro array data types. In this experiment, two stages of testing were performed: 1) characterization by Cosine and SNR by selecting appropriate characteristics, divided into subgroups from 100 to 1000 characteristics, and 2) Data classification performance testing. It consists of three data classification techniques: decision tree, *k*-NN and simple bay. The results was performed with 47 samples of lymphoma, 4,026. Characteristics the results of the trial were as follows. The SNR characterization method can provide effective data classification. The best average is 94.33% with only 100 attributes, while all data classification provides. The efficiency of 86.52% indicates that the selection of the appropriate characteristics before classification of the data is a guideline. This is especially suitable for enhancing the efficiency of the classification of information as well. This is because the micro-array data can be highly fragmented. Which some of the information may not be consistent It is therefore imperative that filtering is performed first.

III. RESEARCH METHODOLOGY

In this research, the research team has set the steps into 5 main steps, as shown in Fig. 4, the process of conducting research and describe follows step.

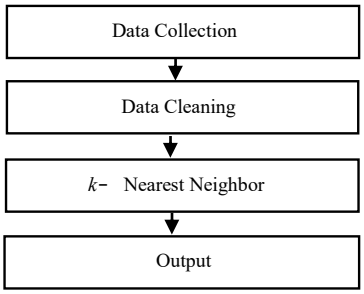


Fig. 4. Shows an overview of the work system.

A. Data Collection

The dataset were used in our work must be numerical. Because in the process of dividing the data into two parts is the data that is overlaid and the non-overlapping data has a process to find the distance between the data. And the data used must be data without data loss or missing-value because such data will not be able to find the distance between the data, need to Pre-Processing before using the data. In table I shows examples of data available. By this dataset used taken from the UCI Dataset Machine Learning Repository database (https://archive.ics.uci.edu/ml/datasets/Mushroom).

It consists of 800 records of data and has 23 attributes as: Cap shape, Cap surface , Cap color , Bruises , Odor , Gill attachment , Gill spacing , Gill size , Gill color , Stalk shape ,

Stalk root , Stalk surface above ring , Stalk surface below ring , Stalk color above ring, Stalk color below ring , Veil type , Veil color , Ring number , Ring type , Spore print color , Population, Habitat and Class. It has 2 classes, which are in the class attribute, a toxic and non-toxic classes. From the data in table II, it can be shown as a feature of the dataset.

TABLE I. EXAMPLE OF MUSHROOM DATASET

Cap shape	Cap surface	Cap color	Bruises	Odor	...	Cass
Convex	Smooth	Brown	Yes	Musty	.	Toxic
Convex	Smooth	Brown	Yes	Musty	.	Toxic
Convex	Scaly	Pink	Yes	Musty	.	Toxic
Convex	Scaly	Brown	Yes	Almond	.	Toxic
Flat	Scaly	White	No	Anise	.	Non-toxic
Flat	Scaly	Green	No	Almond	.	Toxic
Convex	Smooth	Pink	Yes	Spicy	.	Non-toxic
Sunken	Grooves	Green	Yes	Musty	.	Toxic
Bell	Fibrous	Pink	No	Almond	.	Non-toxic
Bell	Fibrous	Purple	No	Anise	.	Toxic
Sunken	Grooves	Yellow	Yes	Almond	.	Non-toxic
Bell.	Fibrous	Purple	No	Spicy	.	Toxic
Sunken	Grooves	Green	Yes	Musty	.	Toxic
.	.	.	.	.	.	.
Convex	Smooth	White	Yes	Fishy	.	Non-toxic

TABLE II. EXPLANATION OF THE MEANING OF ATTRIBUTES

Attribute	Feature meaning
Cap shape	Bell (1) , Conical (2) , Convex (3) , Flat (4) , Knobbed (5) , Sunken (6)
Cap surface	Fibrous (1) , Grooves (2) , Scaly (3) , Smooth (4)
Cap color	Brown (1) , Buff (2) , Cinnamon (3) , Gray (4) , Green (5) , Pink (6) , Purple (7) , Red (8) , White (9) , Yellow (10)
Bruises	Yes (1) , No (2)
Odor	Almond (1) , Anise (2) , Creosote (3) , Fishy (4) , Foul (5) , Musty (6) , None (7) , Pungent (8) , Spicy (9)
Gill attachment	Attached (1) , Descending (2) , Free (3) , Notched (4)
Gill spacing	Close (1) , Crowded (2) , Distant (3)
Gill size	Broad (1) , Narrow (2)
Gill color	Black (1) , Brown (2) , Buff (3) , Chocolate (4) , Gray (5) , Green (6) , Orange (7) , Pink (8) , Purple (9) , Red (10) , White (11) , Yellow (12)
Stalk shape	Enlarging (1) , Tapering (2)
Stalk root	Bulbous (1) , Club (2) , Cup (3) , Equal (4) , Rhizomorphs (5) , Rooted (6) , Missing (7)
Stalk surface above ring	Fibrous (1) , Scaly (2) , Silky (3) , Smooth (4)
Stalk surface below ring	Fibrous (1) , Scaly (2) , Silky (3) ,Smooth (4)
Stalk color above ring	Brown (1) , Buff (2) , Cinnamon (3) , Gray (4) , Orange (5) , Pink (6) , Red (7) , White (8) , Yellow (9)
Stalk color below ring	Brown (1) , Buff (2) , Cinnamon (3) , Gray (4) , Orange (5) , Pink (6) , Red (7) , White (8) , Yellow (9)
Veil type	Partial (1) , Universal (2)
Veil color	Brown (1) , Orange (2) , White (3) ,Yellow (4)
Ring number	None (5) , One (6) , Two (7)
Ring type	Cobwebby (1) , Evanescent (2) , Flaring (3) , Large (4) , None (5) , Pendant (6) , Sheathing (7) , Zone (8)
Spore print color	Black (1) , Brown (2) , Buff (3) , Chocolate (4) , Green (5) , Orange (6) , Purple (7) , White (8) , Yellow (9)
Population	Abundant (1) , Clustered (2) , Numerous (3) , Scattered (4) , Several (5) , Solitary (6)
Habitat	Grasses (1) , Leaves (l) , Meadows (2) , Paths (3p) , Urban (4) , Waste (5) , Woods (6)
Class	Non-toxic, Toxic

B. Data Cleaning

The vast majority of real-world data scientists spend 80% of their time cleaning data and the remaining 20% building a clean model. This data is like cooking when we have raw materials. In addition to carefully selecting raw materials. We still have to take the raw material to clean the bark and trim the rotten parts. Cut into shapes that are ready to cook and many more steps in order for the dish to be cooked to the best. Moreover, the dataset in table I. must transform to Table III.

TABLE III. MUSHROOM DATASET AFTER DATA CLEANING

Cap shape	Cap surface	Cap color	Bruises	Odor	...	Cass
3	4	1	1	6	.	Toxic
3	4	1	1	6	.	Toxic
3	3	6	1	6	.	Toxic
3	3	1	1	1	.	Toxic
4	3	9	2	2	.	Non-toxic
4	3	5	2	1	.	Toxic
3	4	6	1	9	.	Non-toxic
6	2	5	1	6	.	Toxic
1	1	6	2	1	.	Non-toxic
1	1	7	2	2	.	Toxic
6	2	10	1	1	.	Non-toxic
1.	1	7	2	9	.	Toxic
6	2	5	1	6	.	Toxic
.	.	.	.	.	.	.
3	4	9	1	4	.	Non-toxic

C. Classification

From the Fig. 3, we have two types of data, Toxic (triangle data) and Non-toxic (rectangular data), with 1 point representing 1 data, while the green circle is the data that we will predict that the circle should be in a group of squares or a triangle using the principle of *k*-NN, which first what we need to do is assign a number to *k*. For example, from the figure we select *k* equal to 3. Then by the principle of *k*-NN will run for the nearest data point. It is the most number of 3 data points, from which the data in the solid circle line is the three data that we have selected as close to the green circle.

Then, what *k*-NN will do next is figure out what type of data is the most in the group of the three captured data. Here it can be seen that the 3 data is divided into 2 triangles and 1 square, which means that the triangular data has a larger number, then *k*-NN will turn the green circle into triangle type data or if we were to change *k* equal to 5, then it would be displayed according to the information in the dotted circle, see in Fig 5. This time, there are only 2 triangles, but with three squares of data, which means that in this round *k*-NN will assign our green circle data to be square data instead. And

that's it he preliminary work of *k*-NN or the nearest neighbor's algorithm.

```
k−Nearest Neighbor
Determine (k, distance)

Classify (□, □, □) // □ is training data, □ is labels of
□, □ is unknown sample
for i = 1 to n do // n is all the training data
            Compute distance d(□□ ,□)
end for

Compute set I containing indices for the k smallest distance
d(□□ ,□).
return majority label for {□ where i ∈ I}
```

Fig. 5. pseudo-code of mushroom classification by *k*-NN.

IV. EXPERIMENTAL AND RESULTS

In the design of our experiment, we divided the Mushroom Data Set into two groups using a randomized method, training set and test set. where the sampling rate was 50%. We took the total number of samples of each data set out of 22 features through the classification efficiency process, it was seen that the classification by Naïve Bayes Multinomial Text, Naïve Bayes Updateable, Naïve Bayes, SGD Text, LWL, *k*-NN and Stacking method. The *k*-NN was high accurate than other methods show in table IV. The bar chart is show the comparison accuracy rate between method as show in Fig. 6.

TABLE IV. COMPARISONS OF MUSHROOM CLASSIFICATION EFFICIENCY

Classification Method	Accuracy rate
Naïve Bayes Multinomial Text	49.94%
Naïve Bayes Updateable	96.38%
Naïve Bayes	96.38%
SGD Text	50.06%
LWL	92.88%
k-NN	100.00%
Stacking	49.94%

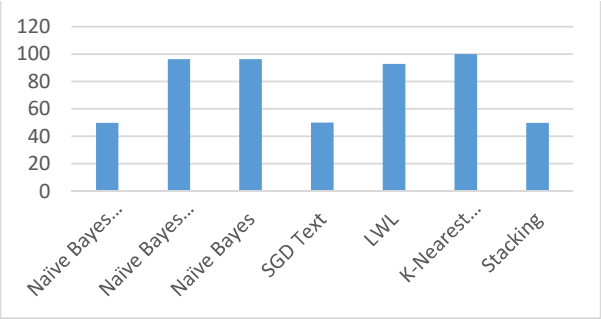


Fig. 6. Bar chart compare seven classification method.

From table IV. The result comparison of dataset classification efficiency highest is k-NN with accuracy rate 100%. Followed by Naïve Bayes Updateable and Naïve Bayes with 96.38%, LWL 92.88%, SGD Text 50.06%, Naïve Bayes Multinomial Text and Stacking 49.94%, respectively. The confusion matrix in table V., precision and recall 1.00.

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1.000	0.000	1.000	1.000	1.000	Toxic
1.000	0.000	1.000	1.000	1.000	Non-Toxic
1.000	0.000	1.000	1.000	1.000	

TABLE V. CONFUSION MATRIX OF CLASSIFICATION

CONCLUSION

Our work focus to classify by physical characteristics Mushroom between toxic and non-toxic. In addition, we comparison the efficiency with seven classification method with Mushroom dataset. The mushroom dataset consists of 22 feature attributes and 800 records. All of them divided into two class as toxic and non-toxic. For academic standards, we separate dataset for training and testing set by randomly at 50/50, then in each set have 400 records. For efficiency classification, we compared seven methods for get the highest accuracy rate. The results k-NN show 100% accuracy rate. The reason is this dataset are numeric data with discrete values then suitable for *k*-NN, thus making the classification highly accurate.

For a line of extension from the research results in this paper. We see the main points in two trends: the first issue is developed to be an application such as a phone application for general users to use easily or developed as a tool for tourists who enjoy trekking in the forest. The second issue is to improve the experimental process to find results with other algorithms into new knowledge.

REFERENCES

[1] H. Chen, L. Fu, P. Dong, X. Zhang and X. Lu, "Acute Toxicity Evaluation and Compositional Analysis of a Polysaccharide from the Medicinal Mushroom Inonotus obliquus," 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, Beijing, 2009, pp. 1-4, doi: 10.1109/ICBBE.2009.5162275.

[2] A. Masoudian and K. A. Mcisaac, "Application of Support Vector Machine to Detect Microbial Spoilage of Mushrooms," 2013 International Conference on Computer and Robot Vision, Regina, SK, 2013, pp. 281-287, doi: 10.1109/CRV.2013.10.

[3] A. Wibowo, Y. Rahayu, A. Riyanto and T. Hidayatulloh, "Classification algorithm for edible mushroom identification," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, 2018, pp. 250-253, doi: 10.1109/ICOIACT.2018.8350746.

[4] T. Kaewwiset and P. Yodkhad, "Automatic temperature and humidity control system by using Fuzzy Logic algorithm for mushroom nursery," 2017 International Conference on Digital Arts, Media and Technology (ICDAMT), Chiang Mai, 2017, pp. 396-399, doi: 10.1109/ICDAMT.2017.7905000.

[5] J. B. Frenila, O. V. Silvela and A. C. Paglinawan, "Novel approach to urban farming: A case study of a solar-powered automated mushroom cultivation in a plastic box," 2015 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Cebu City, 2015, pp. 1-6, doi: 10.1109/R10-HTC.2015.7391865.

[6] K. Mallikarjuns, N. John Sushma, G. Narasimha, K. Venkateswara Rao, L. Manoj and B. Deva Prasad Raju, "Synthesis and spectroscopic characterization of palladium nanoparticles by using broth of edible mushroom extract," International Conference on Nanoscience, Engineering and Technology (ICONSET 2011), Chennai, 2011, pp. 612-615, doi: 10.1109/ICONSET.2011.6168045.

[7] Fangkun Zhu, Li Qu, Wang Xuejing and Fan Wenxiu, "Heavy metals in some edible mushrooms consumed in Xinxiang, China," 2011 International Conference on New Technology of Agricultural, Zibo, 2011, pp. 1034-1036, doi: 10.1109/ICAE.2011.5943965.

[8] G. Lv, Z. Zhang, H. Pan and L. Fan, "Antioxidant Properties of Different Solvents Extracts from Three Edible Mushrooms," 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, Beijing, 2009, pp. 1-4, doi: 10.1109/ICBBE.2009.5162375.

[9] S. Ismail, A. R. Zainal and A. Mustapha, "Behavioural features for mushroom classification," 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, 2018, pp. 412-415, doi: 10.1109/ISCAIE.2018.8405508.

[10] N. Chumuang, P. Pramkeaw and A. Farooq, "Electrical Impedance Of Breast's Tissue Classification By Using Bootstrap Aggregating," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy, 2019, pp. 551-556, doi: 10.1109/SITIS.2019.00093.

[11] W. On-num, N. Chumuang and C. Siladech, "The Development of Intelligent Models for Health Classification," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045374.

[12] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 5, pp. 1774-1785, May 2018, doi: 10.1109/TNNLS.2017.2673241.

[13] Y. Zhang, J. Wu, J. Wang and C. Xing, "A Transformation-Based Framework for KNN Set Similarity Search," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 3, pp. 409-423, 1 March 2020, doi: 10.1109/TKDE.2018.2886189.

[14] F. Zhao and Q. Tang, "A KNN Learning Algorithm for Collusion-Resistant Spectrum Auction in Small Cell Networks," in IEEE Access, vol. 6, pp. 45796-45803, 2018, doi: 10.1109/ACCESS.2018.2861840.

[15] M. S. Sarma, Y. Srinivas, M. Abhiram, L. Ullala, M. S. Prasanthi and J. R. Rao, "Insider Threat Detection with Face Recognition and KNN User Classification," 2017 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, 2017, pp. 39-44, doi: 10.1109/CCEM.2017.16.

[16] J. Vieira, R. P. Duarte and H. C. Neto, "kNN-STUFF: kNN STreaming Unit for Fpgas," in IEEE Access, vol. 7, pp. 170864-170877, 2019, doi: 10.1109/ACCESS.2019.2955864.

[17] M. P. Vaishnnave, K. S. Devi, P. Srinivasan and G. A. P. Jothi, "Detection and Classification of Groundnut Leaf Diseases using KNN classifier," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878733.

[18] M. Türkoğlu and D. Hanbay, "Combination of Deep Features and KNN Algorithm for Classification of Leaf-Based Plant Species," 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 2019, pp. 1-5, doi: 10.1109/IDAP.2019.8875911.

[19] P. Tamrakar, S. S. Roy, B. Satapathy and S. P. S. Ibrahim, "Integration of lazy learning associative classification with kNN algorithm," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-4, doi: 10.1109/ViTECoN.2019.8899415.

[20] A. A. Bharate and M. S. Shirdhonkar, "Classification of Grape Leaves using KNN and SVM Classifiers," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 745-749, doi: 10.1109/ICCMC48092.2020.ICCMC-000139.

[21] K. Thirunavukkarasu, A. S. Singh, P. Rai and S. Gupta, "Classification of IRIS Dataset using Classification Based KNN Algorithm in Supervised Learning," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777643.

[22] T. Pinthong, W. Yimyam, N. Chumuang and M. Ketcham, "License Plate Tracking Based on Template Matching Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 299-303, doi: 10.1109/ISCIT.2018.8587992.

[23] X. Zheng, J. Liu and J. Liu, "An Equalization Method based on W-KNN for PON with PAM4," 2019 18th International Conference on Optical Communications and Networks (ICOCN), Huangshan, China, 2019, pp. 1-3, doi: 10.1109/ICOCN.2019.8933884.

[24] K. Theresia Diah, A. Faqih and B. Kusumoputro, "Exploring the Feature Selection of the EEG Signal Time and Frequency Domain Features for k - NN and Weighted k-NN," 2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129), Depok, West Java, Indonesia, 2019, pp. 196-199, doi: 10.1109/R10-HTC47129.2019.9042448.

[25] C. Liou et al., "Design and Fabrication of Micro-LED Display with Sapphire Micro-Reflector Array," 2020 IEEE 33rd International Conference on Micro Electro Mechanical Systems (MEMS), Vancouver, BC, Canada, 2020, pp. 1153-1156, doi: 10.1109/MEMS46641.2020.9056340.

An Efficiency Random Forest Algorithm for Classification of Patients with Kidney Dysfunction

Narumol Chumuang[1], Nuttawoot Meesang[2], Mahasak Ketcham[3], Worawut Yimyam[4],
Jiragorn Chalermdit[5], Nawarat Wittayakhom[6], Patiyuth Pramkeaw[7]

[1,2,5,6]Dep. of Digital Media Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand.
[3]Dep. of Management Information System, King Mongkut's University of Technology North Bangkok, Thailand.
[4]Dep. of Business Information Management, Phetchaburi Rajabhat University, Thailand.
[7]Media Technology Program, King Mongkut's University of Technology Thonburi, Thailand.

lecho20@hotmail.com[1], nuttawootzaaaaa@gmail.com[2], mahasak.k@it.kmutnb.ac.th[3], worawut_yimyam@hotmail.com[4],
jiragorncha@mcru.ac.th[5], nawaratwit@mcru.ac.th[6], patiyuth.pra@kmutt.ac.th[7]

***Abstract*— In this paper, we presented a separate separation and comparison of data of people with renal impairment. By collecting information on CKD. The data was collected for selection in data mining using the CKD data set from UCI Machine Learn Repository to compare the classification of 400 CKD patients, comprising 25 attributes and dividing into two class, which one is for patients with CKD and those who do not suffer from CKD. In the experimental designing with 5-folds cross validation test, the result is separation by technique as Random Forest shows an accuracy of 100 %, BayesNet 98.75 %, Stochastic Gradient Descent (SGD) 98.25%, Sequential Minimal Optimization (SMO) 95.75%, Multinomial Logistic Regression (MLR) 95.75% respectively.**

Keywords*— *Kidney Dysfunction, Classification, Random Forest, RF.

I. INTRODUCTION

CKD calls CKD is a disease that is currently a major global problem [1]-[3]. Because it adversely affects the health of the people in the country and resulted in death [1],[4]. According to the International Nephrology Society, 10% of the global population is suffering from CKD and each year more than 1 million people die of the disease [1], [2] which in the present in the world [4]-[6]. There are 17.6 % of the total population of CKD patients or approximately 8 million people who are terminally ill patients 2 hundred thousand and more than 7,800 people get sick each year causes of CKD [7]. It is caused by a condition where there is continuous deterioration of kidney function over a period of months or years. Including patients with other diseases such as diabetes, hypertension, gout, kidney stones, nephritis, etc. often result in permanent deterioration of the kidneys. This is due to side effects from drugs and chemicals, as well as other risk factors such as aging. Family history of CKD, overweight, and smokers. All of these factors result in the kidneys being unable to return to function normally.

In the present era, that is the age of technology. The technology is widely used and it can be applied in many areas, including communication, agriculture, business, education, or even medicine [8]. One of the technology that is very popular today is data mining technology which is the process of doing information to find patterns, guidelines, and hidden relationships in that dataset based on statistics, recognition, machine learning and mathematics [9]. In which data mining techniques can be applied can be applied in medicine [10]-[12] as well by classification of information find the relationship between the various physical examination results with disease using stored patient and doctor diagnosis information to help diagnose the patient's disease or medical research.

We see the importance of this reason. Therefore, we have implementation data mining technology or data mining techniques to be used with medical and assisted in the classification analysis using a total of 400 data samples for finding abnormalities in patients with renal dysfunction and compare the performance of the algorithms to find the most accurate analysis results.

The organize paper consists of the following topics: part II discusses related literature, part III research methodology, part IV results of experiment and the last one is a conclusion.

II. LITERATURE REVIEW AND RELATED WORK

As for the theory and related research. We studied algorithms for data classification by studying the fundamentals of nephrotic syndrome and algorithms with different fundamental characteristics as show in Fig. 1.

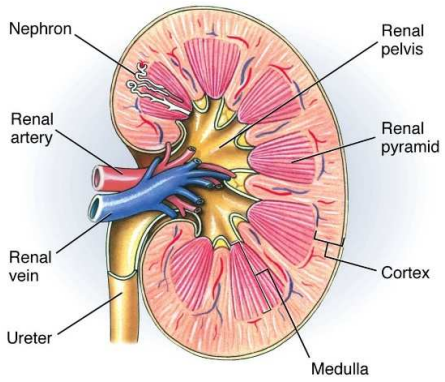


Fig.1. Kidney characteristics.

The kidneys are an important part of the urinary system. They include the kidneys, ureters, bladder, urethra. The kidneys are the body's vital organs, as are other vital organs such as the heart, liver, lungs, stomach, brain, which all of these organs perform specific functions [13]. But they are well coordinated thus keeping the body normal and happy. If one organ is damaged or destroyed, it can affect the function of the other [14]. The kidneys look like a red bean. It is located in the 2 lumbar areas under the ribs and on the two sides of the spine. It is red like fresh pig kidneys. It is long, measuring 11-

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

12 cm. in diameter and weighing 150 grams per side [15]. Each kidney receives blood through the aorta. Out of the heart as the blood flows through the kidneys, it is filtered through the tiny kidney units called the nephron, which are located 1 million on each side [16]. These tiny kidneys filter waste products from the blood through the ureter and the urine is excreted from the body through the urethra. The body can survive with only one normal kidney because it has a very good balance. Therefore, those who donate one kidney can lead a normal life with only one kidney left [15].

A. CKD

CKD [17] is a major public health problem, affecting people of all ages, races and economic conditions, increasing prevalence and incidence of the disease. Due to diabetes high blood pressure and obesity. The physical differences between normal and chronic kidneys are shown in Fig. 2. In the United States, there are more than 20 million people, or 1 in 9 people with CKD, and 20 million people at risk of CKD. No symptoms in the first stage, the symptoms of kidney failure appear when the function of kidney failure is greater than 70-80%. In order for the patient to be diagnosed with CKD quickly, this is to raise awareness among doctors, healthcare professionals and patients and initiate treatment or behavior modification live early in accordance with the disease situation. Instead of neglecting to the point of being CKD that is very severe and therefore becomes alert as it was in the past.

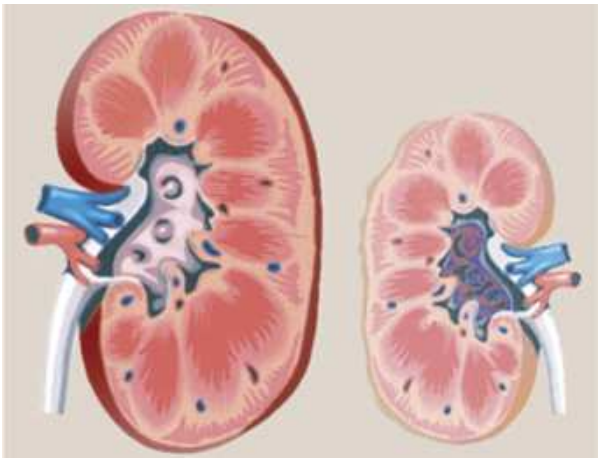


Fig. 2. The normal (left) kidney is large, red with smooth edges, and the chronic (right) kidney is smaller, pale, and jagged.

B. Random Forest

Random Forest [18] is a type of algorithmic decision tree algorithm that is either unpruned or regression trees, which is generated by randomly selecting data samples and data attributes. Build a decision tree where part of the unselected sample will be used to test the decision tree. This section is called Out-of-Bag (OOB), the method is called bagging. The results independently of each decision tree are taken into account for the largest number of votes. The Random Forest algorithm does not need to. There is test data to estimate errors because OOB data has been used to test decision trees.

The principle of Random Forest [19] is to create a model from multiple decision tree models (from 10 models to more than 1000 models) as in Fig. 3, each model receives a different data set, which is a subset of all data sets. When making prediction, each decision tree is assigned for their prediction and the prediction of the vote output chosen by the most decision tree classification case as in eq (1). When performing Random Forests based on classification data, you should know that you are often using the Gini index, or the formula used to decide how nodes on a decision tree branch.

$$Gini = 1 - \sum_{i=1}^{C}(p_i)^2 \quad (1)$$

This formula uses the class and probability to determine the Gini of each branch on a node, determining which of the branches is more likely to occur [20]. Here, p_i represents the relative frequency of the class you are observing in the dataset and c represents the number of classes. You can also use entropy to determine how nodes branch in a decision tree.

$$Entropy = \sum_{i=1}^{C} -p_i * log_2(p_i) \quad (2)$$

Entropy uses the probability of a certain outcome in order to make a decision on how the node should branch. Unlike the Gini index, it is more mathematical intensive due to the logarithmic function used in calculating it.

The mean value from the output of each decision tree in regression case in eq (3) the mean squared error (MSE) to how your data branches from each node. Each decision tree model in Random Forest is considered to be a weak learner - which is not a very good model [21]. But when taking each decision tree to make prediction together, it will have a total model that is skilled and more accurate than the decision tree that made a single prediction.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2 \quad (3)$$

Where N is the number of data points, f_i is the value returned by the model and y_i is the actual value of data point i.

The principles of analyzing data classification using Random Forest are:

1) Sample the data (bootstrapping) from all data sets to get n different sets according to the number of decision trees in RF. For example, the default data set has ten features (X_1, X_2, ..., X_{10}), each decision tree will have a feature go not the same and will not get all data rows from all data sets (X_1 -> X_1 ', X_2-> X_2', ...).

2) Create a model decision tree for each data set.

3) Aggregation of results from each model (bagging) e.g. voting in classification or mean in regression case.

Random Forest works for both problems [22]. classification and regression. Moreover, it can be used for both structured (column / table style data) and unstructured e.g. image, text. Do hyper-parameter tune for Random Forest not overfitting, not difficult. It does not assumption and features whether to distribute data as normal distribution, or relative to target linearly, nor to correlate additional features called interaction - e.g. feature X_1 * X_2 from X_1 and X_2.

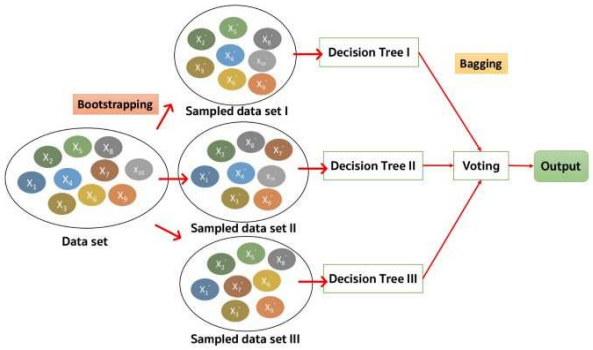


Fig. 3. The process of classification with Random Forest.

C. *Related work*

R. Gupta, et al., [23] present a comparision the efficiency of the classification method by choosing to use the method as close as possible how the tree decides neural network method, support vector machine, rule- based methods, log- logistic regression methods and naave bay to measure the efficiency of group classification. They used the data set on CKD from Apollo Hospital, India by dividing the data into model building blocks And the model test kits at a ratio of 70 and 30, respectively. By comparing the mean square validity and tolerance The classification method is effective. The classification is best. How the tree decides Which gives an accuracy of 100% and the mean square error is 0.0059.

M. P. N. M. Wickramasinghe, et al., [24] developed a medical support system for classifying patients with CKD using data mining technique. The comparison of the efficiancy of algorithms for predicting CKD and assess the effectiveness of the kidney disease prediction system. The main functions of the system include patient information management medical personnel physical examination, disease prediction, diagnosis of disease prediction results. The model management and various reports from this benchmark were found that Random Forest, which used a dataset with oversampling data, had the highest accuracy rate of all models at 97.29%, precision 95.76% and F-Measure 97. 44% , and used this data mining technique. This development as a model for predicting kidney disease. The results of the evaluation of the efficacy of systemic kidney disease were found that the unknown data was 95. 71% accurate in predicting kidney disease. The techniques and developed models could be further developed to develop an effective medical support system in the future.

M. Muñoz- Organero and V. Corcoba- Magaña, [25] present a suitable models for predicting stress with data mining techniques. The process was initiated with a group of 300 samples to conduct a stress assessment test in four cycles. The experimental was conducted for a period of two months, and then all the results were fed for creating a model with six algorithms as Bayesian Network, Naïve Bayesian, Decision Tree: 4. 5, Decision Table, Partial Rules (PART) , and Multilayer Perceptron (MLP). In order to create the model and the model test, the 10- fold cross- validation method was used. They are used to divide the data into two sets: the teaching dataset and the test dataset. From all the model tests, it was found that the most suitable model in used to predict stress is the model is based on the MLP algorithm applied to the historical data of six months. The accuracy values were 81%, 0.81, 0.81 and 0.81 respectively.

L. Li, [26] purpose of this paper was to analyze the career choice interests of learners from John Holland's theoretical test with data mining techniques. The data set from student, 3,000 whose characteristics include gender, age, subject, cumulative GPA. Educational goals, domicile, occupation, parents And median family income. The model was developed with 3 decision tree techniques with the J48 algorithm, LMT, Random Forest, data classification techniques using Bayesian learning and data classification techniques using the neural network using 10- fold cross validation to divide the data into sets. Which contains the data set training and test set. The accuracy, precision and F-measure forecasting performance of the model. The results of the model forecasting efficiency were found that the data classification technique by the decision tree with the Random Forest can predict better than other. The results were 91.37%, 91.30%, 91.40% and 91.30%, respectively.

V. Mittal and D. Singh, [27] presented of cassava inputs from the Office of Agriculture to create a forecast model by using the classification with decisionTree, 5 algorithms were used, namely J48, RandomTree, SimpleCart, NaïveBayes, and LADTree, and the forecasting model was tested. Cross-validation test was found to be accurate at 70.96%, which the researcher thought was not good enough Therefore, the testing method was improved by dividing the data set into 2 parts, namely Training Set and Test Set, 5 sets and then creating a forecasting model with algorithms J48 , RandomTree, SimpleCart. Again, NaïveBayes and LADTree were found to have higher accuracy than the method. Cross-validation test in all algorithms. Therefore, the researcher chooses a forecasting model that offers the highest accuracy. In terms of test data, the J48 algorithm provides the highest accuracy at 75.64%, the SimpleCart algorithm provides the highest accuracy at 80.12%, and the LADTree algorithm provides the highest accuracy at 89.55%. The RandomTree algorithm and Naïve Bayes algorithm provide an accuracy of less than 70%, so they are not taken into account.

III. RESEARCH METHODOLOGY

In this research team has set the steps into 5 main steps, as shown in Fig. 4, the process of conducting research and describe the details follows step by step.

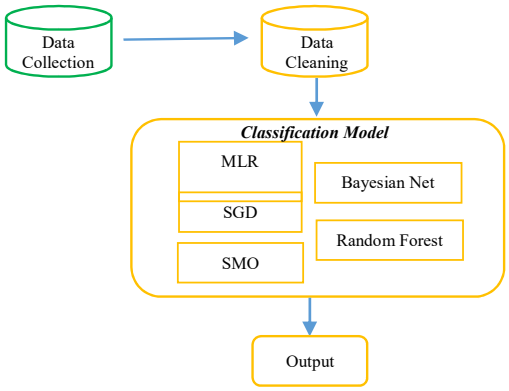


Fig. 4. Shows an overview of the work flow system.

A. Data Collection

The dataset were used in our work must be the numerical. Because in the process of dividing the data into two parts is the data that is overlaid and the non-overlapping data has a process to find the distance between the data and the data used must be data without data loss or missing value because such data will not be able to find the distance between the data, need to pre-processing before using the data. In table I shows examples of data available. This dataset used taken from CKD from a popular database to classify data as well. Machine Learning UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/ckd).

Twenty-five attributes have been widely misinterpreted in the past as a dependent variable representing presence or absence of a CKD. This is incorrect [1]. The twenty-five field was created by BUPA researchers as a train/test selector. The dataset does not contain any variable representing presence or absence of a CKD. we used this dataset as a classification benchmark should follow the method used in experiments by the donor and others.

From the data, 401 samples consisted of 25 attributes are used in our work. The description of each attributes are show in Table I.

TABLE I. DESCRIPTION OF ATTRIBUTES

No.	Feature	Description	No.	Feature	Description
1	age	Age of patient.	14	pot	Potassium
2	bp	Blood pressure.	15	hemo	Hemoglobin
3	sg	Specific gravity.	16	pcv	Packed cell volume
4	al	Albumin.	17	wc	White blood cell count.
5	su	Sugar level.	18	rc	Red blood cell count
6	rbc	Red blood cells.	19	htn	Hypertension
7	pc	Pus cell.	20	dm	Diabetes mellitus
8	pcc	Pus cell clumps	21	cad	Coronary artery disease
9	ba	Bacteria.	22	appet	Appetite
10	bgr	Blood glucose random.	23	pe	Pedal edema
11	bu	Blood urea.	24	ane	Anemia
12	sc	Serum creatinine	25	class	class

The Example format of the data set looks like an example of the information as follows:

TABLE II. EXAMPLE OF CHRONIC KIDNEY DATASET

age	bp	sg	al	su	…	class
48	80	1.02	1	0	…	n
7	50	1.02	4	0	…	a
62	80	1.01	2	3	…	a
48	70	1.005	4	0	…	a
51	80	1.01	2	0	…	a
60	90	1.015	3	0	…	a
48	80	1.02	1	0	…	n
7	50	1.02	4	0	…	n
61	79	1.001	2	3	…	n
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

The data in this study was divided into two classes of abnomally, including hepatic impairment and normally means no hepatic impairment.

TABLE III. DEFINED CLASS TYPE VARIABLE

Type of meaning	Class
People who do not have CKD.	“nockd”
People who have CKD.	“ckd”

B. Data Cleaning

Data cleansing or data cleaning or data scrubbing means data cleaning. It is the process of reviewing and correcting (or removing) invalid entries from the dataset. Table or database which is the cornerstone of the database, because it means imperfection Inaccuracy Inconsistent with other information, for example, these inaccurate information must be replaced, updated or deleted o provide quality information as shows in Table II. Data cleaning arise due to the data inconsistency. This may be due to errors in recording, transmission, or different interpretations of the stored data. The more it needs to be integrated with other databases such as data warehouses or multiple databases. Therefore has a high chance of being born "unclean data".

C. Classification Model

Modeling the classification of patients with hepatic dysfunction to help alleviate the burden of small number of healthcare workers and nursing homes with seven variables obtained from the UCI dataset, data mining and a number of techniques used in modeling five decision tree techniques are follows.

1) Bayesian Net is a learning method that minimizes the limitation of learning easily. In the premise of the incompatibility between characteristics in the simple rubay method in the hypothesis of inequality. In fact, some characteristics are found to be dependent on each other, and this dependency should be included in the model. BayesNet describe condition independent incompatibility between work network context variables. BayesNet that "variable" is used instead of "attribute" to make the learning process effective. Which can put knowledge before in the job network BayesNet are provided in the form of a conditional structure, network and probability tables which is considered an advantage of the Bay Area network. The Bay network is characterized by being able to describe the relationships between variables in a graphical model based on prior knowledge in describing and constructing the network

2) Multinomial logistic regression (MLR) is a technique used to predict the value of one dependent variable from the value of more than one independent variable, where the dependent variable is Is a nominal variable that has the There are three or more values, each representing a type / type / group of variables.

3) Stochastic Gradient Descent (SGD) is an algorithm that updates the parameters in every practice dataset. It is a fairly

fast algorithm only one update per training cycle. In which every time there is an update. The updated parameters have a high variance and result in loss function values varying in different intensities. These steps are good because they allow us to find the smallest values possible (to the center of the star itself), but this method has a problem, the lower we converge. The resulting value becomes more variable and complex.

4) Sequential Minimal Optimization (SMO) is a method used to replace the lost values and transform the data. The group of features (nominal) as binary data and all attribute data is in standard (Normalized) format SMO is also effective for non-linear use. It also helps to manage the structure of the model and reduce the risk of data being generated. The information is reliable in the practice of forecast.

5) Random Forest (RF) technique is a technique to randomly select data and feature decision tree, which is generated from sampling with replacement of the data and then using it to create a tree, where part of the sample is not selected. This information, called Out-of-Bag (OOB), is used in decision tree testing. This method is called bagging. The results independently of each decision tree are accounted for voting. The maximum number of votes indicates the status of the Random Forest technique class test data is not required to estimate the error, because the OOB data has already been used to test the decision tree.

We can describe the working of Random Forest algorithm for classification of patients with liver disorders with the help of following and this following pseudo-code will illustrate its working in Fig 5.

Step 1 − First, start with the selection of random samples from a liver disorders dataset.

Step 2 − Next, Random Forest will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 − In this step, voting will be performed for every classify liver disorders result.

Step 4 − At last, select the most voted prediction result as the final classification result.

To generate c bootstrap samples:
for $i = 1$ to c **do**
  Randomly sample the training data D with replacement to produce D_i
  Create a root node, N_i containing D_i
  Call BuildTree(N_i)
end for

BuildTree(N):
if N contains instances of only one class **then**
  return
else
  Randomly select x% of the possible splitting features in N
  Select the feature F with the highest information gain to split on
  Create f child nodes of N, N_1,..., N_f, where F has f possible values (F_1, … , F_f)
  for $i = 1$ to f **do**
    Set the contents of N_i to D_i, where D_i is all instances in N that match F_i
    Call BuildTree(N_i)
  end for
end if

Fig. 5. Pseudo-code of liver disorders classification by Random Forest.

IV. EXPERIMENTAL AND RESULTS

In our experiment we got the Chronic Kidney dataset was used to use 7 aerosols: mcv, alkphos, sgpt, sgot, gammagt, drinks and class which field created by the BUPA researchers to split the data into train / test sets. For 359 samples due to the small amount of data then performed a 5-fold cross-validation type experiment. For more reliable performance testing by dividing training data into 5 equal parts, then testing the performance of five models with the following illustrate in Fig. 6.

Round 1 uses data sections 2,3,4 and 5 to model and predict part 1 to perform the test.

Round 2 uses data sections 1,3,4 and 5 to model and predict part 2 for testing.

Round 3 uses the data sections 1,2,4 and 5 to model and predict part 3 for testing.

Round 4 uses the data sections 1,2,3 and 5 to model and predict part 4 to perform the test.

Round 5 uses the data sections 1,2,3 and 4. Model and predict Part 5 to perform the test.

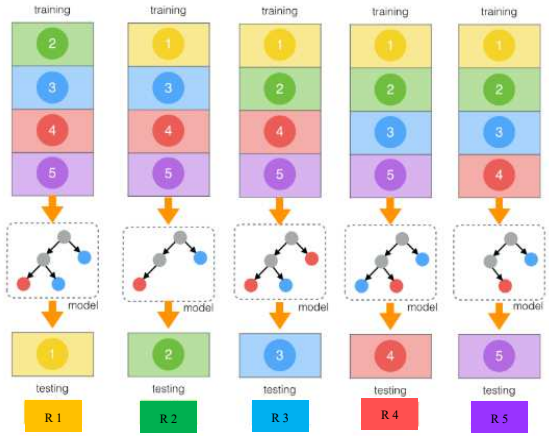


Fig. 6 An experiment with 5-fold cross validation.

From the mean difference test, the accuracy of the algorithm. The results are performed by using a 5-fold cross validation method with different algorithms. Resulting in different tolerances and precision of all 400 data samples, the Random Forest algorithm had 100% of the accuracy rate, which is the most accurate compared to the five methods tested in this work. show in table IV.

TABLE IV. COMPARISON OF LIVER DISORDERS DATA SET CLASSIFICATION EFFICIENCY

Classified	Accuracy Rate	Error Rate
Bayesian Net	98.75 %	1.25 %
MLR	95.75 %	4.25 %
SGD	98.25 %	1.75 %
SMO	97.75%	2.25%
Random Forest	100.00 %	0.00 %

In table IV. The result comparison of Chronic Kidney dataset classification efficiency highest is Random Forest with accuracy rate 100.00%. Followed with Bayesian Net 98.75%, Stochastic Gradient Descent (SGD) with 98.25%, Sequential Minimal Optimization (SMO) 97.75%, and Multinomial logistic regression (MLR) 95.75%.

CONCLUSION

For the purpose of this paper, the study was conducted on a comparative model to separate data from CKD patients. The CKD dataset from UCI Machine Learn Repository was used to compare the classification of 400 CKD patients, comprised of 25 attributes, and were divided into 2 classes: CKD patients and non-CKD patients. The experimental was performed using a 5-fold cross validation method, with the Random Forest algorithm having 100% accuracy rate which is the highest accuracy when comparison with others. It can be concluded that Random Forest is the most effective. Therefore, the forecasting model a comparative model to sort out patient data using data mining using Random Forest is therefore appropriate for further forecasting.

REFERENCES

[1] M. L. Braunstein, "Health Care in the Age of Interoperability: The Potential and Challenges," in IEEE Pulse, vol. 9, no. 5, pp. 34-36, Sept.-Oct. 2018, doi: 10.1109/MPUL.2018.2856941.

[2] G. Chen et al., "Prediction of CKD Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 100497-100508, 2020, doi: 10.1109/ACCESS.2020.2995310.

[3] E. Hodneland et al., "In Vivo Detection of CKD Using Tissue Deformation Fields From Dynamic MR Imaging," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 6, pp. 1779-1790, June 2019, doi: 10.1109/TBME.2018.2879362.

[4] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng and B. Chen, "A Machine Learning Methodology for Diagnosing CKD," in IEEE Access, vol. 8, pp. 20991-21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[5] L. Zhang, H. Fu, L. Bai, X. Liu, Y. Zeng and Y. Wu, "Good management of CKD craves for registration system: A case of a single center in China," 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Belfast, 2014, pp. 213-216, doi: 10.1109/BIBM.2014.6999361.

[6] Y. Amirgaliyev, S. Shamiuulu and A. Serek, "Analysis of CKD Dataset by Applying Machine Learning Methods," 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, 2018, pp. 1-4, doi: 10.1109/ICAICT.2018.8747140.

[7] N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha and C. Likhitha, "Prediction of CKD (CKD) using Data Science," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 642-647, doi: 10.1109/ICCS45141.2019.9065309.

[8] J. Edwards, "Medicine on the Move: Wearable devices supply health-care providers with the data and insights necessary to diagnose medical issues and create optimal treatment plans [Special Reports]," in IEEE Signal Processing Magazine, vol. 36, no. 6, pp. 8-11, Nov. 2019, doi: 10.1109/MSP.2019.2930767.

[9] T. Yingthawornsuk, N. Chumuang and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 418-423, doi: 10.1109/SITIS.2017.75.

[10] N. Chumuang, P. Pramkeaw and A. Farooq, "Electrical Impedance Of Breast's Tissue Classification By Using Bootstrap Aggregating," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy, 2019, pp. 551-556, doi: 10.1109/SITIS.2019.00093.

[11] K. Vongprechakorn, N. Chumuang and A. Farooq, "Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045560.

[12] N. Chumuang, "Comparative Algorithm for Predicting the Protein Localization Sites with Yeast Dataset," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 369-374, doi: 10.1109/SITIS.2018.00064.

[13] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav and R. Dakshayani, "CKD Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-4, doi: 10.1109/ICNTE44896.2019.8946029.

[14] S. R. Raghavan, V. Ladik and K. B. Meyer, "Developing decision support for dialysis treatment of chronic kidney failure," in IEEE Transactions on Information Technology in Biomedicine, vol. 9, no. 2, pp. 229-238, June 2005, doi: 10.1109/TITB.2005.847133.

[15] Wallace MA. Anatomy and physiology of the kidney. AORN J. 1998 Nov;68(5):800, 803-16, 819-20; quiz 821-4. doi: 10.1016/s0001-2092(06)62377-6. PMID: 9829131.

[16] Wilk A, Wiszniewska B, Rzuchowska A, Romanowski M, Różański J, Słojewski M, Ciechanowski K, Kalisińska E. Comparison of Copper Concentration Between Rejected Renal Grafts and Cancerous Kidneys. Biol Trace Elem Res. 2019 Oct;191(2):300-305. doi: 10.1007/s12011-018-1621-6. Epub 2019 Jan 15. PMID: 30645698; PMCID: PMC6706355.

[17] Paniagua-Sierra JR, Galván-Plata ME. Enfermedad renal crónica [CKD]. Rev Med Inst Mex Seguro Soc. 2017;55(Suppl 2):S116-7. Spanish. PMID: 29697216.

[18] Asoke K. Nandi; Hosameldin Ahmed, "Decision Trees and Random Forests," in Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines , IEEE, 2019, pp.199-224, doi: 10.1002/9781119544678.ch10.

[19] S. Thaiparnit, N. Chumuang and M. Ketcham, "A Comparitive Study of Clasification Liver Dysfunction with Machine Learning," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-4, doi: 10.1109/iSAI-NLP.2018.8692808.

[20] N. Chumuang, P. Pramkeaw and A. Farooq, "Electrical Impedance Of Breast's Tissue Classification By Using Bootstrap Aggregating," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy, 2019, pp. 551-556, doi: 10.1109/SITIS.2019.00093.

[21] K. Vongprechakorn, N. Chumuang and A. Farooq, "Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045560.

[22] S. F. Noshahi, A. Farooq, M. Irfan, T. Ansar and N. Chumuang, "Design and Fabrication of an Affordable SCARA 4-DOF Robotic Manipulator for Pick and Place Objects," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2019, pp. 1-5, doi: 10.1109/iSAI-NLP48611.2019.9045203.

[23] R. Gupta, N. Koli, N. Mahor and N. Tejashri, "Performance Analysis of Machine Learning Classifier for Predicting CKD," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154147.

[24] M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with CKD (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303, doi: 10.1109/LSC.2017.8268202.

[25] M. Muñoz-Organero and V. Corcoba-Magaña, "Predicting Upcoming Values of Stress While Driving," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1802-1811, July 2017, doi: 10.1109/TITS.2016.2618424.

[26] L. Li, "Evaluation Model of Education Service Quality Satisfaction in Colleges and Universities Dependent on Classification Attribute Big Data Feature Selection Algorithm," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 2019, pp. 645-649, doi: 10.1109/ICITBS.2019.00160.

[27] V. Mittal and D. Singh, "Critical Analysis of Fusion Algorithms for Digital Agriculture: An Efficient Application of PALSAR Data," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 5917-5920, doi: 10.1109/IGARSS.2019.8898580.

Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm

Pronab Ghosh
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
pronab1712@gmail.com

F. M. Javed Mehedi Shamrat
Department of Software Engineering
Daffodil International University
Dhaka, Bangladesh
javedmehedicom@gmail.com

Shahana Shultana
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
shahanashomi246@gmail.com

Saima Afrin
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
saima.cse@diu.edu.bd

Atqiya Abida Anjum
Department of Computer Science & Engineering
United International University
Dhaka, Bangladesh
aanjum151026@bscse.uiu.ac.bd

Aliza Ahmed Khan
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
alizaahmedkhancse22ju@gmail.com

***Abstract*—Chronic Kidney disease (CKD), a slow and late-diagnosed disease, is one of the most important problems of mortality rate in the medical sector nowadays. Based on this critical issue, a significant number of men and women are now suffering due to the lack of early screening systems and appropriate care each year. However, patients' lives can be saved with the fast detection of disease in the earliest stage. In addition, the evaluation process of machine learning algorithm can detect the stage of this deadly disease much quicker with a reliable dataset. In this paper, the overall study has been implemented based on four reliable approaches, such as Support Vector Machine (henceforth SVM), AdaBoost (henceforth AB), Linear Discriminant Analysis (henceforth LDA), and Gradient Boosting (henceforth GB) to get highly accurate results of prediction. These algorithms are implemented on an online dataset of UCI machine learning repository. The highest predictable accuracy is obtained from Gradient Boosting (GB) Classifiers which is about to 99.80% accuracy. Later, different performance evaluation metrics have also been displayed to show appropriate outcomes. To end with, the most efficient and optimized algorithms for the proposed job can be selected depending on these benchmarks.**

Keywords—Support Vector Machine, AdaBoost, Linear Discriminant Analysis, Gradient Boosting.

I. INTRODUCTION

Kidney disease develops very slowly without revealing any symptoms. There are various forms of kidney disease around the world. So most doctors usually waste their precious time to detect whether a patient is affected by kidney disease or not. In this paper, we are basically working on "Chronic Kidney Disease"[1] based on different performance indices to figure out which algorithm is best to use in this type of problem. The fast prediction of the disease can help save thousands of lives worldwide before severe damage has been done to the patients. Besides, machine learning algorithms can be used [2] to detect this disease. To detect this rising disease, several machine learning algorithms can be trained [3] depending on medical patients' data. But the challenge is to get the most accurate prediction in the shortest time.

The crucial aim of the proposed research is to build a kidney disease system focused entirely on machine-learning. The research aims to solve various algorithms, such as SVM, AB, LDA, and GB to classify the people affected by kidney disease. To make it more accurate, this study is performed using different performance assessment metrics such as False Negative Rate (FNR), Accuracy (ACC), Precision (PRE), Negative predictive value (NPV), F1 Score (F1), False Discovery Rate (FDR), Standard Deviation (SD), Specificity (SPE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Sensitivity (SEN), Root Mean Squared Error (RMSE), False Positive Rate (FPR), ROC, AUC, Error Rate, and Execution time to properly evaluate classifiers performance. The main aims of this research are:

- All missing value issues have been solved through the imputation method of K-Nearest Neighbors to obtain more reliable outcomes.
- With the aid of a standard scaler technique, all features are prepossessed to hold the values within the range of [0, 1].
- The assessment process of different models has been experimented with the 80:20 distinction.
- This study elucidates accuracy, error rate, execution time, AUC, and ROC figures to demonstrate the efficiency of different classifiers.

II. LITERATRE REVIEW

Over the past researchers have shown the use of machine learning algorithms to perform various calculations and evaluate data to come up with decisions to better human life. From the field of math and science to business, medicine to every day human life, experiments using machine learning algorithms bring fruitful results.

Using AB ensemble classifiers, the authors in [4] performed human activity recognition. The data for the experiment data was gather from human body sensor. The high performance shows the feasibility of the algorithm. Developmental Dysplasia of Hip (DDH) in infants can be deadly. In [5] the authors use SVM technique to detect DDH in acoustic non-invasive data. Using an acoustic noise of 10-2500Hz the data is collected and the performance of the model is calculated. Here, SVM gives an accuracy rate of 79%. ECG signal frequently contains noises and speckles. After filtering through various image processing techniques, machine learning algorithm is used in [6]. Here, LDA is used to features from the input ECG signals and SMV to recognize pattern. Specificity, Sensitivity and mean square root is calculated to measure to efficiency of the algorithms. In [7] the authors studied the data of socio-demographic, clinical and magnetic

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

resonance imaging to predict Alzheimer's disease using GB algorithm. The algorithm predicted the disease on the aspect of socio economical state, age, education, gender, and mini-mental state exam. The model achieved 91.3% accuracy in prediction.

From the research gap, SVM, AB, LDA, and GB classifiers have been addressed to detect CKD. The data are preprocessed before the algorithms are implemented on them. Here, the main object of the study is to find the best and most optimal algorithm for prediction of kidney disease.

III. Research Methodology

A. System Overview:

This deals with the theoretical ability of the research work. It will give a piece of clear information about the concept of work. For the study, the dataset is collected form an online data repository. This data is cleaned using several pre-processing techniques. The feature selection is done on the dataset obtained from the repository. Once the data is cleaned and processed, it is divided in training set and test set data. Four machine learning classification algorithms are trained using the training data. After the algorithms are train, they are implemented on the test data to obtain prediction. In this study, the accuracy and performance of the prediction among the four algorithms are compared to determine the most efficient algorithm to predict the chronic kidney disease among the patients. The following Fig. 1 is illustrated to show the proposed model of this research.

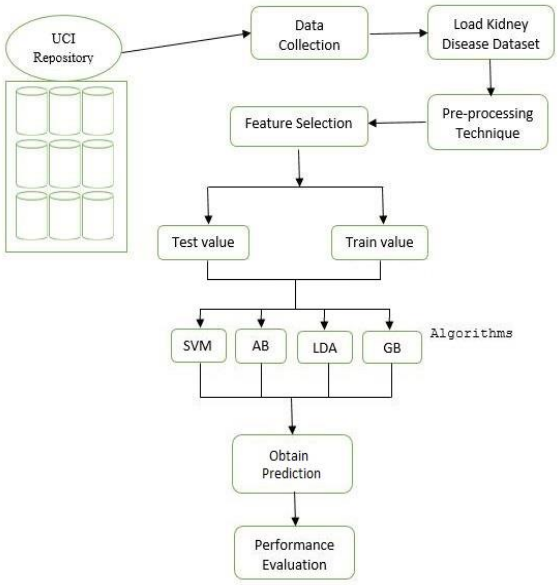


Fig. 1. The proposed model of kidney disease detection

B. Evaluation Criteria on Performance Measure Indices:

There are four different learning techniques used to represent the model together with some performance indices such as precision, recall, F1-score, specificity. The outcomes of performance measure indices are dependent on TP, TN, FP, and FN. [16-17]

True Positive = A list of reported cases exactly classified with CKD.

False Positive = A list of confirmed incidents incorrectly classified with CKD.

True Negative= A list of reported instances exactly classified with CKD.

False Negative= A list of confirmed instances exactly classified with CKD.

IV. Implementation

A. Different Machine Learning Libraries:

The proposed model implemented via Jupiter Notebook using basic python libraries are coded in the Python programming language including Panda - an open-source library that performs a superior function [8], Pyplot - generates the same view Matplotlib as similar to MatLab, Seaborn - helps to draw interesting and informative statistical graphics. A variety of machine learning techniques is used to solve real-world [9] problems used by Sklearn Python libraries [10].

B. Dataset Collection:

Data is thought of as the first and global parts of the research field. As a large number of patient records are collected from one of the most popular sites so as to get the most efficient outcomes. In this study, CKD disease dataset has been applied to predict deadly diseases from the UCI Machine Learning Repository [11]. We picked 25 individual features with 400 entries taken in a CSV format [12] from their database in where 250 are Kidney disease, KD class and 150 is a not-KD class. There are three types of data included such as float64(11), int64(1), and object(14). All features of categorical data such as objects have been converted into numbers by label encoding [13]. Take an example, we do label "not-KD" and "KD" with respect to 0 and 1.

C. Data Pre-processing:

After collecting a number of raw data from one of the most popular repositories, the dataset is to be applied in the preprocessing section. In this area, a number of missing values is detected to fulfill the data demand. Almost every column in the dataset has missing values that observe in the Table I. As the following dataset contains a number of missing values, it creates a number of complex situations in order to predict an accurate outcome. This major problem must be solved by two well-known methods in order to handle missing values. Among them, the median technique, the average of the two middle numbers, has been used to solve this vital problem. After applying this method, there have been observed no missing values depicted in Fig. 2.

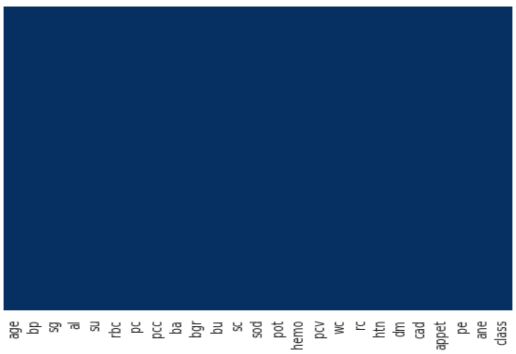


Fig. 2. No missing values in chronic kidney disease dataset

After the successful completion of the preprocessing technique, the proposed dataset attributes have been used for selection. To reduce the dimensionality, the process of feature selection has been employed. To get a better prediction rate [14-15], the narrow subsections of the appropriate features are extracted from its dataset.

D. Spliting Training and Testing set:

The dataset is partitioned into two parts: training and testing. In the training part, more than 70% of data is given to get predicted attributes value where the rest of the values is assigned for testing data. After finishing the training as well as the testing process, our machine is ready for doing classification [30]. To achieve an accurate outcome, four separate machine learning tools such as SVM, AB, LDA, and GB classifiers have been driven to predict disease stages.

TABLE I. VARIOUS ATTRIBUTES WITH MISSING VALUES

Features	Features Code	Number of missing values
Age	age	9
blood pressure	bp	12
specific gravity	sg	47
albumin	al	46
sugar	su	49
red blood cells	rbc	152
pus cell	pc	65
pus cell clumps	pcc	4
bacteria	ba	4
blood glucose random	bgr	44
blood urea	bu	19
serum creatinine	sc	17
sodium	sod	87
potassium	pot	88
hemoglobin	hemo	52
packed cell volume	pcv	70
white blood cell count	wc	105
red blood cell count	rc	130
hypertension	htn	2
diabetes mellitus	dm	2
coronary artery disease	cad	2
appetite	appet	1
pedal edema	pe	1
anemia	ane	1
class	class	0

E. Support Vector Machine:

Support vector machine [18] is also known as support vector networks and supervised learning models associated with machine learning algorithms. It also analyzes data using for regression and classification and the working process has been described by Bhagile et al. [19].

$$Y = sign\ (\sum_{i=1}^{N} y_i \alpha_i\ (x * x_i) + b) \qquad (1)$$

From equation 10, $(x * x_i)$ is known as labeled training and works as an input vector. Fig. 3 shows the working process of the implemented algorithm.

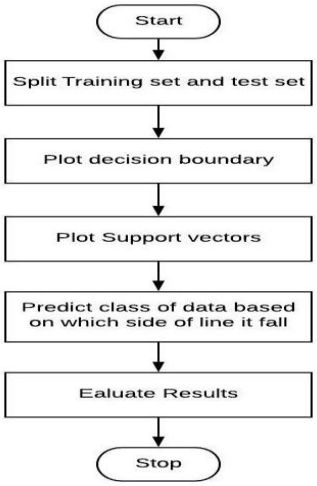


Fig. 3. The working process of SVM algorithm.

F. AdaBoost:

The algorithms of Boosting merge different weak classifiers to form strong classifiers to enhance the classification accuracy [20]. Another practical algorithm, Adaptive Boosting, had been suggested by Friedman et al. in 1997 by Fried and Schicher, although later in 2000. It was shown that LogitBoost overcame this situation through better generalizations. Boosting algorithms solves a variety of medical issues, namely the protein structure class detection in [21], cancer detection in [22] and breast cancer identification in [23]. The process of Adaptive Boost is shown in Fig. 4.

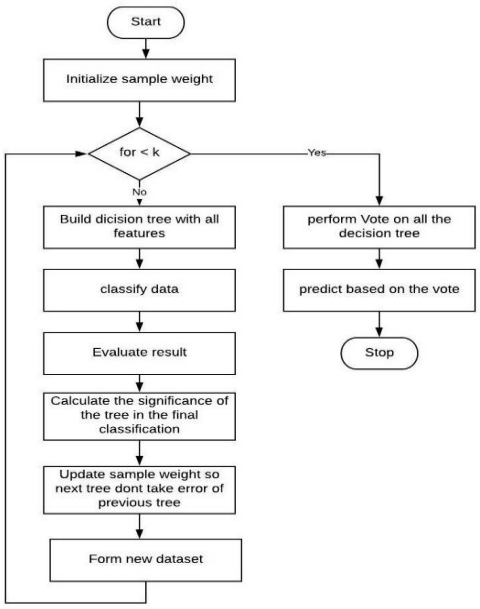


Fig. 4. The illustration process of AdaBoost algorithm.

G. Linear Discriminant Analysis

Linear Discriminant Analysis [24] has been a general form of Fisher's linear discriminant, a process included in statistics, pattern recognition, and machine learning to discover a linear combination of attributes that describes more than two groups of events. Each of C categories has a mean like μi and the same covariance like ∑. Then the scatter between variability

of class might be described by the sample means of covariance class [25]. Fig. 5 and equation 2 show the steps of LDA.

$$\sum b = \frac{1}{C}\sum_{i-1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T \quad (2)$$

H. Gradient Boosting:

Gradient boosting addresses the issues regarding a classification and regression [26]. That form a diagnostic model in the shape of an ensemble of weak forecasting analytics, usually trees of decisions. The model is structured in a phase-wise manner, congruous with other boosting method, and by allowing optimization of an arbitrary differentiable loss function. Our algorithm should be able to handle the addition of some new estimator hm (x) [27] in equation 3.

$$F_{m+1}(x) = F_m(x) + h_m(x) = y \quad (3)$$

The flow of the GB machine learning technique has been portrayed in Fig. 6.

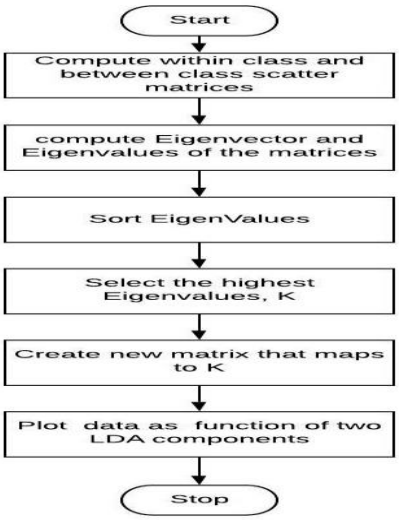


Fig. 5. The process of LDA algorithm

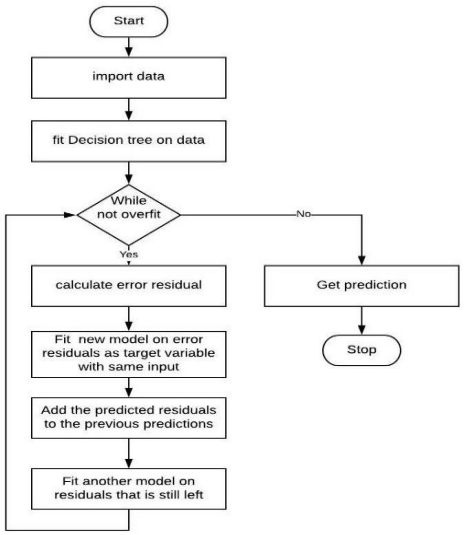


Fig. 6. The flow of the Gradient classifier.

V. Results And Discussion

A. Experimental Outputs Among Different Methods:

After the fruitful evaluation technique on the dataset, a large amount of data has been divided into training and testing. To detect a patient has a KD or not, four approaches of classification with regression were used. Performance metrics are used to justify diverse algorithm methods. Positive classification occurs if and only if a person has symptoms of kidney disease, a person does not have a kidney disease (KD), negative classification occurs. GB Classifier shows the highest outcome among all algorithms. The predictive outcomes have been depicted in Table II.

TABLE II. Performance Measurement Criteria

Dimension	Support Vector Machine	AdaBoost	Linear Discriminant Analysis	Gradient Boosting
ACC	99.56%	97.91%	97.91%	99.80%
SEN	99%	98%	98%	99%
SPE	99%	98%	98%	98%
PRE	99%	99%	99%	98%
NPV	97%	92.30%	92.30%	99%
FPR	0%	0%	0%	0%
FDR	0%	0%	0%	0%
FNR	0%	2.77%	2.77%	0%
F1	99%	98%	98%	99%
SD	0%	17.05%	17.05%	0%
MAE	0%	2.08%	2.08%	0%
MSE	0%	2.08%	2.08%	0%
RMSE	0%	14.43%	14.43%	0%

From the tables II, the accuracy rate among the four algorithms are compared. Here it is seen that the GB give the accuracy of 99.80% with 99% recall score. , and AB and LDA gives accuracy of 97.91%. The visualization of comparison is shown in Fig. 7.

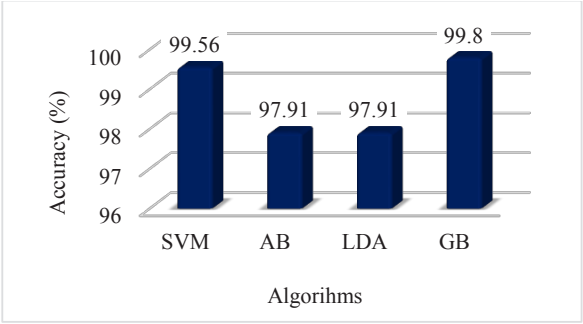


Fig. 7. Accuracy comparisons among different algorithms

B. Execuion Time Measurement over the Models:

The prediction rate is totally dependent on the dataset along with the model preprocessing technique. Besides, time takes a prominent role compared to others. As we know from Fig. 8 that the lowest predictable ratio of run time comes from SVM, On the other hand, GB takes the highest time to achieve a predictable score. Other popular algorithms generate lower time periods to find out the represented outcomes.

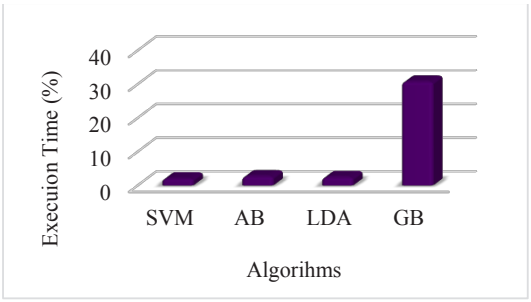


Fig. 8. A Prediction Time Comparison Occurred on Performed Algorithms.

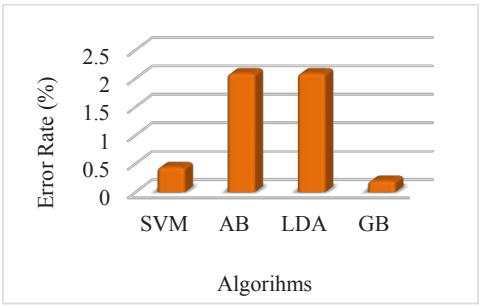


Fig. 9. The error rates of the introduced models.

C. Error Rate Among Various approaches:

Evaluating error rate [28] , accuracy of any algorithm can be measured. This helps gauging the performance of the algorithms. The four algorithms gave low errors rate altogether but the GB gave only 0.20% error after implementation on the dataset. However, AB and LDA generate the mid-lowest accuracy of 97.91% with 1.0 specificity score. The error comparison of the models are shown in Fig. 9.

D. Detection Using ROC and AUC Curves:

The diagnostic ability of the classifiers can be confused by using the confusion matrix and the receiver working property (ROC) curve [29]. The matrix of confusion is also referred to as the contingency matrix in the machine learning studies area. True Positive (TP) happens when the classifiers identified the data successfully while in False Positive (FP) the classifiers cannot identify the data. To measure the performance, AUC and ROC metrics were generated in Fig. 10.

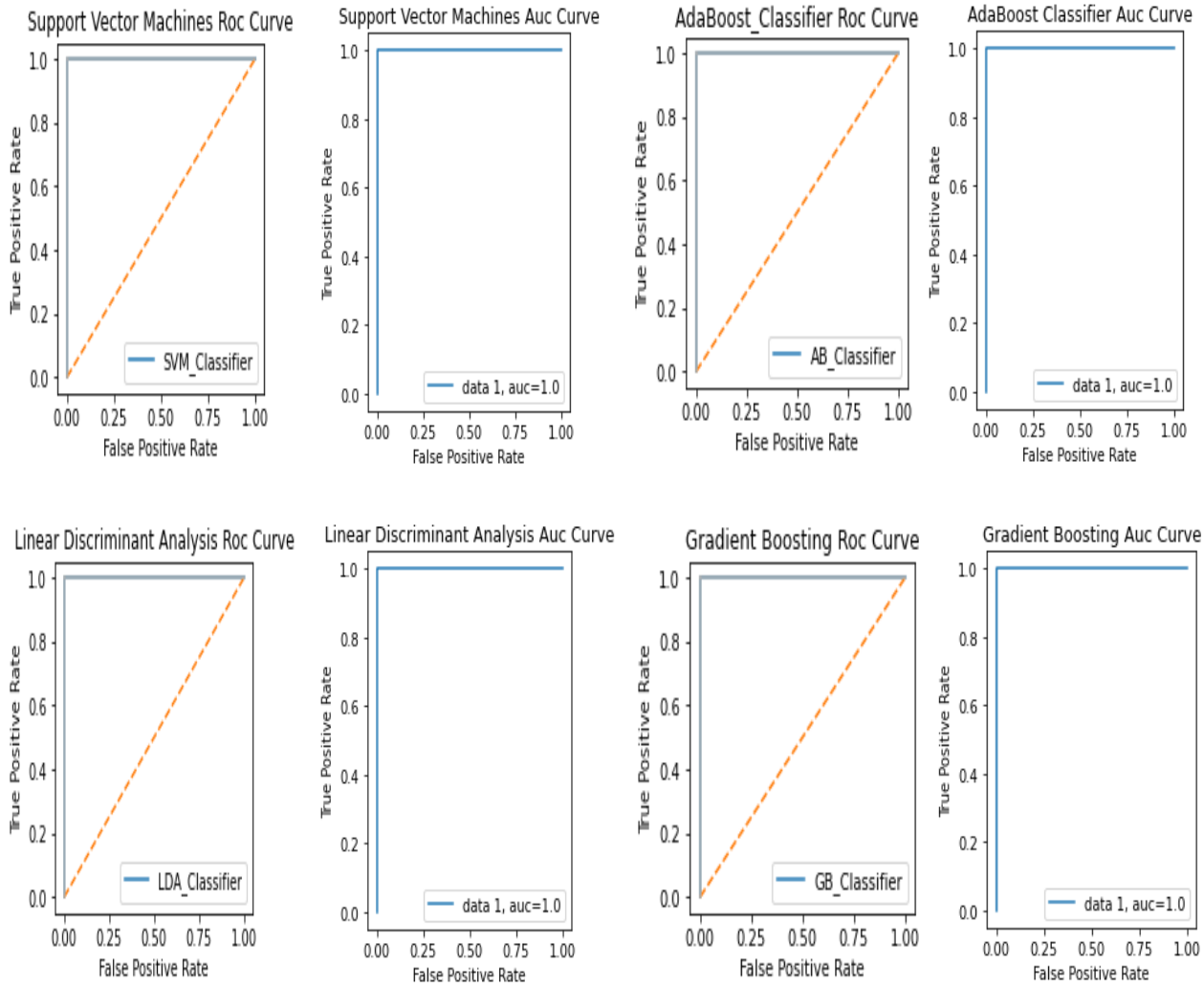


Fig. 10. ROC and AUC curves of all inroduced models.

ROC curve is diagnosed because the receiver working characteristic curve in which AUC is the vicinity under the ROC curve. If the rating of AUC is excessive, the performance of the version must be excessive, and vice versa. The ratings of Support Vector system, Linear Discriminant Analysis, and Gradient Boosting Classifier provide the best rating all of them in each ROC and AUC curves. Support Vector Machine that represents fourth model in the carve and provides 1.0 score in both curves. The score of Linear Discriminant Analysis (LDA) gives the mid-lowest score in both ROC and AUC curves. The score of Gradient Boosting Classifier gives the highest predictable score in both ROC and AUC curves.

VI. CONCLUSION

In this paper, four distinct algorithms were selected to get a precise expectation rate over the introduced dataset. Contrasting all presented approaches, the fruitful results have been gotten from GB classifier. These models effectively generate a 99.80% accuracy rate while AB, and LDA (97.91%) provides a low score. Besides the GB classifier requires more time compared to others to give a prediction and highest predictable score in both ROC and AUC curves. Since an exact pace of expectation is without a doubt reliant on the pre-processing strategy, the methods of the pre-processing must deal with cautiously to accomplish recognized outcomes precisely.

ACKNOWLEDGMENT

The authors are grateful and pleased to all the researchers in this research study.

REFERENCES

[1] "Chronic kidney disease symptoms, treatment, causes & prevention-american kidney fund, " http://www.kidneyfund. org/kidney disease/ chronic-kidney-disease-ckd/, (Accessed on 31/07/2020).

[2] F.M. Javed Mehedi Shamrat, Md. Asaduzzaman, A.K.M. Sazzadur Rahman, Raja Tariqul Hasan Tusher, Zarrin Tasnim "A Comparative Analysis Of Parkinson Disease Prediction Using Machine Learning Approaches" International Journal of Scientific & Technology Research, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 2576-2580.

[3] F. M. Javed Mehedi Shamrat, Md. Abu Raihan, A.K.M. Sazzadur Rahman, Imran Mahmud, Rozina Akter, "An Analysis on Breast Disease Prediction Using Machine Learning Approaches" International Journal of Scientific & Technology Research, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2450-2455.

[4] Alghamdi, Raghad A. Makawi, Eman A. Albiety, Tayeb Brahimi, Akila Sarirete: Sensor Based Human Activity Recognition Using Adaboost Ensemble Classifier, In: Procedia Computer Science, Volume 140, (2018), Pages 104-111, ISSN 1877-0509, https://doi.org/10.1016/j.procs. 2018.10.298.

[5] M. E. Alam, N. Smith, D. Watson, T. Hassan and K. Neupane, "Early Screening of DDH using SVM Classification," 2019 SoutheastCon, Huntsville, AL, USA, 2019, pp. 1-4, doi: 10.1109/SoutheastCon42311.2019.9020565.

[6] Varatharajan, R., Manogaran, G. & Priyan, M.K. A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. Multimed Tools Appl 77, 10195–10215 (2018). https://doi.org/10.1007/s11042-017-5318-1

[7] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, C. P. Fulton, Classification of Alzheimer's Disease with and without Imagery Using Gradient Boosted Machines and ResNet-50. Brain Sci. 2019, 9, 212.

[8] H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, pp. 1351-1356, 2018.

[9] J. Sivapriya, V. K. Aravind , S. S. Siddarth and S. Sriram, " Breast Cancer Prediction Using Machine Learning," International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, Issue. 4, Nov 2019.

[10] P. Ghosh, M.Z. Hasan, O.A. Dhore, A.A. Mohammad and M. I. Jabiullah, "On the Application of Machine Learning to Predicting Cancer Outcome", Proceedings of the International Conference on Electronics and ICT – 2018, organized by Bangladesh Electronics Society (BES), Dhaka, Bangladesh on 25-26 November, 2018, pp: 60.

[11] "UCI machine learning repository: Early stage of Chronic kidney disease dataset, https://archive.ics.uci.edu/ml/datasets/Chronic KidneyDisease, (Accessed on 05/06/2020).

[12] P. Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, pp. 193-198, 2017.

[13] P. Ghosh, M. Z. Hasan and M. I. Jabiullah, "A Comparative Study of Machine Learning Approaches on Dateset to Predicting Cancer Outcome", Journal of the Bangladesh Electronic Society, Vol. 18, Num. 1-2, June, December 2018, ISSN: 1816-1510, pp:81-86.

[14] Zheng, Lijuan, H. Wang, and S. Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews", International Journal of Machine Learning and Cybernetics, Vol. 6, March, 2015.

[15] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, pp. 1-5, 2016.

[16] F. M. J. M. Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi, S. Shultana "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease" 2020 IEEE International conferenvce for Innovation in Technology, 2020.

[17] S. Manikandan, "Heart attack prediction system," International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, pp. 817-820, 2017.

[18] Y. Amirgaliyev, S. Shamiluulu and A. Serek, "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods," 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, pp. 1-4, 2018.

[19] S. A. Hannan, V. D. Bhagile, R. R. Manza, R. J. Ramteke," Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feed forward Backpropagation Technique, " ,International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2150-2159, 2010.

[20] A. U. Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, pp. 1-6, 2019.

[21] Y. Cai, K. Feng, W. Lu and K. Chou, "Using LogitBoost classifier to predict protein structural classes", Journal of Theoretical Biology, vol. 238, no. 1, pp. 172-176, 2006.

[22] M. Dettling, and P. Buhlmann, "Boosting for tumor classification with gene expression data", Bioinformatics, vol. 19, no. 9, pp. 1061-1069, 2003.

[23] W. Zhang, F. Zeng, X. Wu, X. Zhang and R. Jiang, "A Comparative Study of Ensemble Learning Approaches in the Classification of Breast Cancer Metastasis," 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, Shanghai, pp. 242-245, 2009.

[24] A. Nishanth and T. Thiruvaran, "Identifying Important Attributes for Early Detection of Chronic Kidney Disease," in IEEE Reviews in Biomedical Engineering, vol. 11, pp. 208-216, 2018.

[25] "A simple overview on Linear Discriminant Analysis," https://en.wikipedia.org/wiki/Linear_discriminant_analysis, (Accessed on 08/06/2020).

[26] F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Pronab Ghosh, Anup Majumder, Md. Zahid Hasan "Personalization of Job Circular Advertisement to Candidates Using Decision Tree Classification Algorithm" 2020 IEEE International conferenvce for Innovation in Technology, 2020.

[27] F. M. J. M. Shamrat, M. Asaduzzaman, P. Ghosh, M. D. Sultan, and Z. Tasnim, "A Web Based Application for Agriculture: "Smart Farming System," International Journal of Emerging Trends in Engineering Research, July 2020, ISSN: 2347-3983.

A Pressure Sensor- and Depth Camera-Based Monitoring and Alarming System for Bed Fall Detection

Prakrankiat Youngkong
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, THAILAND
e-mail: prakrankiat.you@mail.kmutt.ac.th

Worawit Panpanyatep
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, THAILAND
e-mail: worawit.pan@mail.kmutt.ac.th

***Abstract*—The proposed monitoring and alarming system combined pressure sensors and depth camera, applying four different decision tree techniques for bed fall detection. Trying to avoid personal privacy issues, the pressure sensor performed perfectly to comprehend what position and location users were doing on bed. The depth camera was used to capture off-bed movement of the users right after leaving the bed. Movement data obtained from the sensor and camera was sent and computed online. Any undesirable events were notified and displayed on mobile device screens of caregivers. Total 6 normal subjects (males and females equally) aged 27-46 years old, height 154-186 cm, and weight 44-108 kg. involved in the experiment. 10-fold cross validation was used to evaluate and select the final model. Logistic Model Trees (LMT) yielded the highest accuracy.**

Keywords: On-Bed and Off-Bed Movement, Decision Tree, Pressure Sensor, Depth Sensor and Fall Detection

I. INTRODUCTION

Global population aged 65 and over is growing faster than all other age groups. Over the next three decades, the number of ageing population in all regions is reaching over 1.5 billion in 2050 [1]. As this number becomes bigger, an increase of fall event has been witnessed. In fact, falling is a very serious issues for the elderly. It is a factor of great social relevance for public health, as it is a major cause of injuries, trauma, hospitalizations and death. In addition, these events contribute to functional decline and decrease autonomy, with direct consequences on the quality of life [2]. Fall is defined as unintentionally coming to rest on the ground or other levels with or without consciousness [3]. It was listed into three classes based on its characteristics as fall from sleeping (or bed), fall from sitting (or chair) and fall from working or standing [4].

Based on these gigantic demand, global trend, potential market and social values, series of technologies have been researched, developed and evolved from manual to fully automatic over the past decades to prevent fall. Most of the manual fall detection devices and systems require user intervention for triggering the alarm while the automation monitoring and alarming systems are self-learning and response in real-time to minimize the damage caused by the fall [5]. These technologies were generally categorized into three main classes based on the sensor technology deployed to obtain the input data: wearable, camera-based and ambient sensors. Nowadays, the wearable devices, for instance smart watches and wristbands, become widely used. People who care about his or her personal health normally use these wearable devices to monitor daily activities which, in fact, reflect his or her wellness or fitness. Some of these commercial products were used to monitor weight loss, calorie consumption and meditation. For fall detection, Zhao et.al. [6] proposed a method based on a tri-axial gyroscope for fall event recognition. Decision tree was chosen to be the classifier and the accuracy of 99.52% was achieved. Saadeh et.al. [7] proposed another system equipped with a single tri-axial accelerometer. The fall detection algorithm utilized a 1-second sliding frames classification with a liner regress-based offline training to identify a single and optimal threshold for each patient. This system performed up to >95% accuracy.

However, many users, especially elderly, misused or tended to forget wearing the wearable devices. Camera-based and ambient sensor-based systems seemed to have advantages. Vision-based fall detection system called "ViFa" based on video surveillance provide an efficient solution for automatic detection of fall events [8]. Actually, it was mentioned that there must be a trade-off between classification accuracy and computational efficiency. Alonso et.al. [9] introduced the identification of the optimal background subtraction algorithm in indoor night-time environments to detect falls using data obtained from video camera. LBAdaptiveSOM algorithm yielded the best performance in this experiment. For those who were familiar with the WiFi-based connectivity, fall detection was also proposed using this technology. A system called DeFall, standing for "Detect Falls" consisted of an offline template-generating stage and an online decision-making stage, could achieve a detection rate of 96% [10].

In this paper, an IoT-based monitoring and alarming system consisted of pressure sensor and depth camera was proposed. The main objective was to focus on the fall from bed characteristics which were defined as 1) a fall is a process lasting 1-2 seconds 2) the person is lying, sitting or standing on the bed at the beginning of the fall and 3) the lying body on the floor is nearby the bed.

.

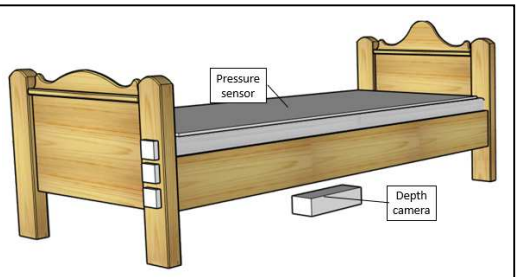


Figure 1 The proposed system setup

II. SYSTEM OVERVIEW

Previously, the system included only on-bed pressure sensors and machine learning algorithms which effectively and accurately provided on-bed movement information, for instance what kind of movement at what location the user was doing. But without having sensors outside the bed, further movements out of the bed were not yet, in fact, comprehended. Thus, the depth camera was added into the system and placed under the bed to cover the bedside movement. Hence, the personal privacy was still preserved.

A. Hardware Development

The hardware of the proposed system was illustrated in Figure 1. It consisted of one pressure sensor mattress and one depth camera. If the user moved out of bed on both sides, two depth cameras were required. The dimension of the on-bed pressure sensor mattress was 90cm x 190cm, containing 144 units pressure sensors inside, to cover whole bed movements. The signal obtained from each pressure sensor node was wired and sent to the signal processing controller (SPC) which was basically consisted of signal amplifier, voltage divider and signal transmitter. Figure 2 illustrated the SPC in further detail. After the signal from

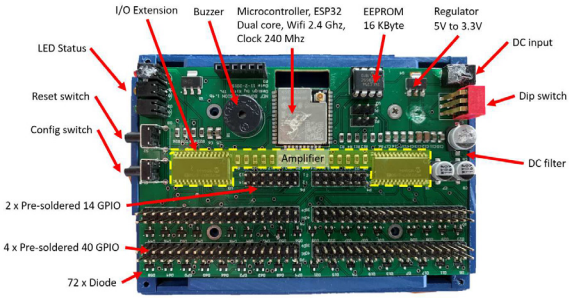


Figure 2 Layout of the SPC

the pressure sensors was amplified and processed, the data was then rearranged in 144-array pattern and transmitted to a cloud platform as illustrated in Figure 3.

A commercial depth camera RealSense™ D435i from Intel® was used during the experiment without accessing the video and picture modes. The depth information 640 x 185 pixels^2 acquired from RealSense™ library, was down sampled by 5. Then, the depth information which became 128 x 37 (4,736) pixels^2, was combined to the pressure one, labeled as different scenarios, and sent to the cloud for further analyzing.

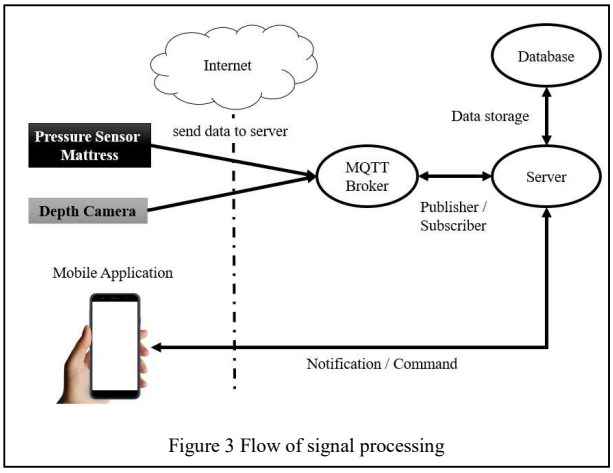


Figure 3 Flow of signal processing

B. Software Development

Heatmap using RGB model was implemented for the users to visualize how much pressure was applied on the mattress. Red and blue were maximum and minimum, respectively. Other colors were ranged as follows,

if (*value* < .10 * *maxValue*) → R = 0; G = 17; B = 255,
else if (*value* < .20 * *maxValue*) → R = 0; G = 97; B = 255,
else if (*value* < .30 * *maxValue*) → R = 0; G = 212; B = 255,
else if (*value* < .40 * *maxValue*) → R = 0; G = 255; B = 102,
else if (*value* < .50 * *maxValue*) → R = 0; G = 255; B = 0,
else if (*value* < .60 * *maxValue*) → R = 95; G = 255; B = 0,
else if (*value* < .70 * *maxValue*) → R = 223; G = 255; B = 0,
else if (*value* < .80 * *maxValue*) → R = 255; G = 159; B = 0,
else if (*value* < .90 * *maxValue*) → R = 255; G = 63; B = 0
else → R = 255; G = 0; B = 0,

where *value* is an input data obtained from the pressure sensors and *maxValue* is the maximum input value. The complete range of color was illustrated in Figure 4.

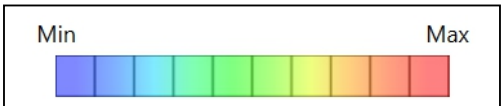


Figure 4 Color Range of Pressure Display

Six different body movement patterns were investigated and described as follows.

On-bed lying (ONLY) included on-bed supine which the user was flat on his or her back, lateral on either left or right and prone, illustrated in Figure 5, in which the user lied flat with the chest down and back up.

On-bed sitting (ONSI) was on-bed patterns that the user was asked to sit casually and freely on bottoms, hands and feet, or pose as prefer.

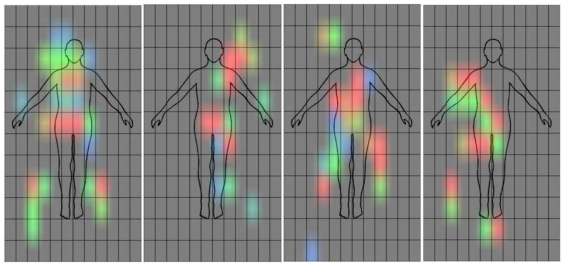


Figure 5 Examples of pressure displays from on-bed supine, lateral to the left, prone and lateral to the right respectively

On-bed standing (ONST) was the on-bed patterns that the user was asked to stand casually and freely on feet without raising any foot off the ground.

The definitions of off-bed patterns were similar to those on-bed patterns but occurred on the floor next to the bed and labeled as *Off-bed lying* (OFLY), *Off-bed sitting* (OFSI) and *Off-bed standing* (OFST). Nevertheless, these off-bed patterns were acquired from the commercial depth camera.

C. Decision Tree Technique

Data from the on-bed pressure sensors mattress and the depth camera were intensively investigated using four different decision tree techniques including bootstrap aggregating or bagging (BAG), decision tree (DT), logistic model trees (LMT) and random forest (RF) to classify 9 different scenarios as shown in Table 1.

TABLE I SEQUENCES OF MOVEMENTS FROM ON-BED TO OFF-BED

Class	Scenario		
	From on-bed	***To off-bed***	***Code***
a	ONLY	OFLY	ONLY2OFLY
b	ONLY	OFSI	ONLY2OFSI
c	ONLY	OFST	ONLY2OFST
d	ONSI	OFLY	ONSI2OFLY
e	ONSI	OFSI	ONSI2OFSI
f	ONSI	OFST	ONSI2OFST
g	ONST	OFLY	ONST2OFLY
h	ONST	OFSI	ONST2OFSI
i	ONST	OFST	ONST2OFST

D. Experiment Setup

Three normal male subjects, 24-45 years old, height 171-185 cm., and weight 67-108 kg., and three normal female subjects, 25-29 years old, height 145-163 cm., and weight 40-55 kg. participated in this experiment. Each subject was instructed to move from the center of the bed toward either bedside starting from on-bed standing, sitting and lying positions to on-floor standing, sitting and lying positions, randomly and repeatedly for 2 minutes with different velocities depended on each individual style. In fact, soft mattresses were provided on the floor to soften up the landing. Each subject was carefully described of how the experiment was conducted before beginning. None of unusual or dangerous movement was allowed.

III. RESULT AND DISCUSSION

After implementing four different decision tree techniques with 10-fold cross-validation, the overall correctly classified percentage that BAG, DT, LMT and RF obtained were 99.05, 98.78, 99.88 and 99.80, respectively.

Table II Confusion matrix of BAG (%)

a	b	c	d	e	f	g	h	i	<-- classified as
97.444	0.500	0.111	0.667	0.111	0.111	0.333	0.667	0.056	a = ONLY2OFNY
0.056	99.611	0.222	0.000	0.000	0.000	0.000	0.056	0.056	b = ONLY2OFSI
0.000	0.278	99.056	0.000	0.000	0.056	0.000	0.056	0.556	c = ONLY2OFST
0.222	0.000	0.000	98.944	0.056	0.000	0.556	0.167	0.056	d = ONSI2OFLY
0.000	0.056	0.000	0.000	99.389	0.000	0.056	0.333	0.167	e = ONSI2OFSI
0.000	0.000	0.333	0.000	0.000	99.444	0.000	0.000	0.222	f = ONSI2OFST
0.111	0.000	0.000	0.444	0.111	0.000	99.222	0.056	0.056	g = ONST2OFLY
0.111	0.000	0.000	0.000	0.389	0.000	0.167	98.778	0.222	h = ONST2OFSI
0.000	0.111	0.167	0.000	0.000	0.000	0.000	0.056	99.556	I = ONST2OFST

Table III Confusion matrix of DT (%)

a	b	c	d	e	f	g	h	i	<-- classified as
98.944	0.167	0.056	0.444	0.111	0.111	0.111	0.111	0.000	a = ONLY2OFNY
0.056	99.389	0.222	0.000	0.111	0.056	0.056	0.111	0.000	b = ONLY2OFSI
0.000	0.333	99.167	0.000	0.000	0.111	0.000	0.000	0.389	c = ONLY2OFST
0.389	0.056	0.000	98.056	0.111	0.000	1.222	0.056	0.111	d = ONSI2OFLY
0.000	0.111	0.056	0.167	98.333	0.000	0.056	1.111	0.167	e = ONSI2OFSI
0.000	0.385	0.333	0.056	0.056	98.944	0.000	0.000	0.333	f = ONSI2OFST
0.111	0.000	0.000	0.889	0.056	0.000	98.667	0.278	0.000	g = ONST2OFLY
0.222	0.167	0.000	0.000	0.556	0.000	0.000	98.944	0.111	h = ONST2OFSI
0.111	0.111	0.278	0.000	0.056	0.444	0.000	0.167	98.611	I = ONST2OFST

Table IV Confusion matrix of LMT (%)

a	b	c	d	e	f	g	h	i	<-- classified as
99.722	0.111	0.000	0.111	0.000	0.000	0.056	0.000	0.000	a = ONLY2OFNY
0.000	100	0.222	0.000	0.000	0.000	0.000	0.000	0.000	b = ONLY2OFSI
0.000	0.000	99.944	0.000	0.000	0.000	0.000	0.000	0.056	c = ONLY2OFST
0.000	0.000	0.000	99.611	0.222	0.000	0.167	0.000	0.000	d = ONSI2OFLY
0.000	0.000	0.000	0.056	99.889	0.000	0.056	0.000	0.000	e = ONSI2OFSI
0.000	0.000	0.000	0.000	0.000	100	0.000	0.000	0.000	f = ONSI2OFST
0.000	0.000	0.000	0.000	0.000	0.000	100	0.000	0.000	g = ONST2OFLY
0.000	0.000	0.000	0.000	0.056	0.000	0.056	99.889	0.000	h = ONST2OFSI
0.000	0.000	0.000	0.000	0.056	0.056	0.000	0.056	99.833	I = ONST2OFST

Table V Confusion matrix of RF (%)

a	b	c	d	e	f	g	h	i	<-- classified as
99.556	0.000	0.056	0.167	0.000	0.111	0.000	0.056	0.056	a = ONLY2OFNY
0.000	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	b = ONLY2OFSI
0.000	0.000	100	0.000	0.000	0.000	0.000	0.000	0.000	c = ONLY2OFST
0.111	0.056	0.000	99.556	0.222	0.000	0.000	0.000	0.056	d = ONSI2OFLY
0.000	0.000	0.000	0.167	99.833	0.000	0.000	0.000	0.000	e = ONSI2OFSI
0.000	0.000	0.000	0.000	0.000	99.944	0.000	0.000	0.056	f = ONSI2OFST
0.000	0.000	0.000	0.111	0.000	0.000	99.889	0.000	0.000	g = ONST2OFLY
0.000	0.000	0.000	0.056	0.000	0.000	0.000	99.944	0.000	h = ONST2OFSI
0.000	0.000	0.111	0.000	0.000	0.444	0.000	0.000	99.444	I = ONST2OFST

Regarding to the characteristics of the fall specified previously, class *a*, *d* and *h* might be considered the major consideration. LMT had the highest percentages on those classes in the confusion matrix. And, RF seemed to be the second best alternative.

TABLE VI BAG – DETAIL ACCURACY

TP Rate	*FP Rate*	*Precision*	*Recall*	*F-Measure*	*Class*
0.974	0.001	0.995	0.974	0.985	*a*
0.996	0.001	0.991	0.996	0.993	*b*
0.991	0.001	0.992	0.991	0.991	*c*
0.989	0.002	0.986	0.989	0.988	*d*
0.994	0.001	0.993	0.994	0.994	*e*
0.994	0.000	0.997	0.994	0.996	*f*
0.992	0.001	0.989	0.992	0.991	*g*
0.988	0.002	0.986	0.988	0.987	*h*
0.996	0.002	0.986	0.996	0.991	*i*

TABLE VII DT – DETAIL ACCURACY

TP Rate	*FP Rate*	*Precision*	*Recall*	*F-Measure*	*Class*
0.989	0.001	0.991	0.989	0.990	*a*
0.994	0.002	0.988	0.994	0.991	*b*
0.992	0.001	0.991	0.992	0.991	*c*
0.981	0.002	0.982	0.981	0.981	*d*
0.983	0.001	0.990	0.983	0.987	*e*
0.989	0.001	0.993	0.989	0.991	*f*
0.987	0.002	0.986	0.987	0.986	*g*
0.989	0.002	0.982	0.989	0.986	*h*
0.986	0.001	0.989	0.986	0.987	*i*

TABLE VIII LMT – DETAIL ACCURACY

TP Rate	*FP Rate*	*Precision*	*Recall*	*F-Measure*	*Class*
0.997	0. 000	1.000	0.997	0.999	*a*
1.000	0.000	0.999	1.000	0.999	*b*
0.999	0. 000	1.000	0.999	1.000	*c*
0.996	0.000	0.998	0.996	0.997	*d*
0.999	0.000	0.997	0.999	0.998	*e*
1.000	0.000	0.999	1.000	1.000	*f*
1.000	0.000	0.997	1.000	0.998	*g*
0.999	0.000	0.999	0.999	0.999	*h*
0.999	0.000	0.999	0.999	0.999	*i*

TABLE IX RF – DETAIL ACCURACY

TP Rate	*FP Rate*	*Precision*	*Recall*	*F-Measure*	*Class*
0.996	0. 000	0.999	0.996	0.997	*a*
1.000	0.000	0.999	1.000	1.000	*b*
1.000	0. 000	0.998	1.000	0.999	*c*
0.996	0.001	0.995	0.996	0.995	*d*
0.998	0.000	0.998	0.998	0.998	*e*
0.999	0.001	0.994	0.999	0.997	*f*
0.999	0.000	1.000	0.999	0.999	*g*
0.999	0.000	0.999	0.999	0.999	*h*
0.994	0.000	0.998	0.994	0.996	*i*

Looking into further detail accuracy from Table VI to IX, LMT and RF were comparable, but LMT was a bit better when focusing on the characteristics of the fall. Nevertheless, FP Rate was very minor in models obtained from LMT.

IV. CONCLUSION

A hybrid sensors-based combining pressure sensors and depth camera monitoring and alarming system for fall prediction was proposed. The pressure sensor mattress was placed on bed to monitor on-bed movements while the commercial depth camera was placed under the bed to monitor off-bed movements. The data obtained from both sensors were combined and analyzed to detect the falls from bed which were characterized as mentioned previously. Four different decision tree machine learning techniques were benchmarked and LMT was the best classifier.

Even though the correctly classified percentage was 99.9, approximately, but further investigation may be worth to be explored in order to obtain better results. Such experiments shall be conducted in multi-center with bigger number of subjects and/or longer time trials.

ACKNOWLEDGMENT

This project is partially funded by "NXPO (Office of Higher Education Science Research and Innovation Policy Council, Thailand) " and part of the project "Promotion of Robotics for All and Development of Robotics Innovators, Researchers, Engineers, and Startups", grant No. B04F630007.

REFERENCES

[1] *World Population Ageing 2019*, United Nations, Department of Economic and Social Affairs, Population Division, June 2020. [Online]. Available: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf

[2] S.B. Giacomini, J.R.Fhon and R.A. Rodrigues, "Frailty and Risk of Falling in the Older Adult Living at Home," *Acta Paul Enferm*. 2020; 33:1-8. doi: 10.37689/acta-ape/2020AO0124.

[3] E.J. Shende, P.M. Gosavi, S. Anandh and Y.A. Pawar, "Effect of Exercise Program in Reducing Risk of Fall in Elderly People," *Indian Journal of Public Health Research & Development* . 2020; 11 Issue 1: 359-362. doi: 10.37506/v11/il/2020/ijphard/193844.

[4] X. Yu, "Approaches and principles of fall detection for elderly and patient," *HealthCom 2008 - 10th International Conference on e-health Networking, Applications and Services*, Singapore, 2008, pp. 42-47, doi: 10.1109/HEALTH.2008.4600107.

[5] A. Singh, S. U. Rehman, S. Yongchareon and P. H. J. Chong, "Sensor Technologies for Fall Detection Systems: A Review," in *IEEE Sensors Journal*, vol. 20, no. 13, pp. 6889-6919, 1 July1, 2020, doi: 10.1109/JSEN.2020.2976554.

[6] S. Zhao, W. Li, W. Niu, R. Gravina and G. Fortino, "Recognition of human fall events based on single tri-axial gyroscope," 2018 *IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, Zhuhai, 2018, pp. 1-6, doi: 10.1109/ICNSC.2018.8361365.

[7] W. Saadeh, S. A. Butt and M. A. B. Altaf, "A Patient-Specific Single Sensor IoT-Based Wearable Fall Prediction and Detection System," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 995-1003, May 2019, doi: 10.1109/TNSRE.2019.2911602.

[8] Ezatzadeh, S., Keyvanpour, M.R. ViFa: an analytical framework for vision-based fall detection in a surveillance environment. *Multimed Tools Appl* **78,** 25515–25537 (2019). https://doi.org/10.1007/s11042-019-7720-3.

[9] M. Alonso, A. Brunete, M. Hernando and E. Gambao, "Background-Subtraction Algorithm Optimization for Home Camera-Based Night-Vision Fall Detectors," in *IEEE Access*, vol. 7, pp. 152399-152411, 2019, doi: 10.1109/ACCESS.2019.2948321.

[10] E. Frank, M.A. Hall and I.H. Witten The WEKA Workbench Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann 4th Edition, 2016.

A Study of Three Statistical Machine Translation Methods for Myanmar (Burmese) and Shan (Tai Long) Language Pair

Nang Aeindray Kyaw
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
nangaeindraykyaw@utycc.edu.mm

Ye Kyaw Thu
Language and Semantic Technology Research Team (LST)
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
yktnlp@gmail.com

Hlaing Myat Nwe
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
hlaingmyatnwe@utycc.edu.mm

Phyu Phyu Tar
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
phyuphyutar@utycc.edu.mm

Nandar Win Min
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
dr.nandarwinmin@utycc.edu.mm

Thepchai Supnithi
Language and Semantic Technology Research Team (LST)
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
thepchai.supnithi@nectec.or.th

***Abstract*—Shan is said to be the second-largest ethnic group of Myanmar. The main motivation is to break down the communication barrier between Shan people and Myanmar people. This paper contributes to the first evaluation of the quality of machine translation between Myanmar (Burmese) and Shan (Tai Long). We also built a Myanmar-Shan parallel corpus (around 11K sentences) based on the Myanmar language of the ASEAN MT corpus. In this research, three different statistical machine translation approaches were used to carry out the experiment: phrase-based, hierarchical phrase-based, and the operation sequence model. Furthermore, two different segmentation schemes were studied, these were syllable segmentation and word segmentation. Translating with syllable segmentation achieved higher quality machine translation for both Myanmar and Shan languages. BLEU and RIBES scoring techniques are used to measure the performance of the machine translations. The operation sequence model gave the highest scores (41.85 BLEU and 0.88031 RIBES) for Shan to Myanmar syllable translation. For Myanmar to Shan syllable translation, hierarchical phrase-based machine translation gave the highest BLEU score of 34.72 and the operation sequence model gave the highest RIBES score of 0.87012. Our experimental results with syllable segmentation produced promising results even with low data resources and we expect this can be developed into a useful translation system as more data comes available in the future.**

***Index Terms*—Statistical Machine Translation, Low-resourced languages, Myanmar, Shan, Syllable segmentation**

I. Introduction

Shan is one of the main eight ethnic groups of Myanmar, with a population of nearly 6 million in Shan state according to the 2017 Myanmar population. Hence, Shan is said to be the second-largest ethnic group of Myanmar. Translation systems between Shan language and Myanmar language are indeed useful for communication between Shan (Tai Long) people and Burmese (Myanmar) people. As Shan and Myanmar languages are completely different in word orders: Myanmar has Subject-Object-Verb (SOV) order and Shan has Subject-Verb-Object (SVO) order, it is a bit challenging in building machine translation systems between these two languages. Our main motivation for this research is to investigate machine translation performance with statistical machine translation (SMT) approaches for Myanmar (Burmese) and Shan (Tai Long) language pair.

This paper contributes to the first implementation and evaluation of automatic machine translation between Myanmar (Burmese) and Shan (Tai Long) language pair in both directions. One more contribution is that we are developing a Myanmar-Shan parallel corpus applicable for machine translation. Moreover, we updated regular expression (RE) of the RE based Myanmar syllable segmentation tool developed by Ye Kyaw Thu [1] for syllable segmentation of the Shan language.

The structure of this paper is organized as follows. A brief literature review of statistical machine translation systems for different language pairs is presented in the upcoming section. In Section III, Shan language along with its common phenomena and nature is introduced. Section IV describes the methodologies used in the machine translation experiments. Later, in Section V, we present Myanmar-Shan parallel corpus preparation for machine translation together with the information about word and syllable segmentation. Section VI presents the used evaluation schemes in measuring machine translation performances. The experimental results along with some discussions are described in Section VII. The error analysis on translated sentences is presented in Section VIII. Eventually, in Section IX, we conclude our works showing promising results for future research.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

II. Literature Review

There have been some studies on the SMT of the Myanmar language. Ye Kyaw Thu et al. (2016) [2] presented the first large-scale study of the translation of the Myanmar language. A total of 40 language pairs were used in the study that included languages both similar and fundamentally different from Myanmar. The results show that the hierarchical phrase-based SMT (HPBSMT) [3] approach gave the highest translation quality in terms of both the BLEU [4] and RIBES scores [5].

Win Pa Pa et al. (2016) [6] presented the first comparative study of five major machine translation approaches applied to low-resource languages: PBSMT, HPBSMT, tree-to-string (T2S), string-to-tree (S2T), and OSM translation methods for the translation of limited quantities of travel domain data between English and Thai, Laos, Myanmar in both directions. The experimental results indicate that in terms of adequacy (as measured by BLEU score), the PBSMT approach produced the highest quality translations. Here, the annotated tree is used only for the English language for S2T and T2S experiments. This is because there is no publicly available tree parser for Lao, Myanmar, and Thai languages. According to our knowledge, there is no publicly available tree parser for the Shan language and thus we cannot apply S2T and T2S approaches to the Myanmar-Shan language pair. From their RIBES scores, we noticed that the OSM approach achieved the best machine translation performance for Myanmar to English translation. Moreover, we learned that the OSM approach gave the highest translation performance in translation between Khmer (the official language of Cambodia) and twenty other languages, in both directions [7].

Ye Kyaw Thu et al. [8] (2019) contributed the first statistical machine translation evaluations between Kachin and Rawang. The experiment used the 10K Kachin-Rawang parallel corpus. Experiments have shown comparable results for translation in both directions. BLEU scores (43.973 for Kachin-Rawang and 46.597 for Rawang-Kachin) and RIBES scores (0.80079 for Kachin-Rawang and 0.81180 for Rawang-Kachin) were achieved.

Neural Machine Translation (NMT), a data-hungry technology, suffers from the lack of bilingual data in low-resource scenarios. With our developing Myanmar-Shan parallel corpus, it is difficult to get comparable results with NMT. Predominantly, we would like to investigate SMT performance for the Myanmar-Shan language pair. Based on the experimental results of previous works, in this paper, the machine translation experiments were carried out using PBSMT, HPBSMT, and OSM.

III. Shan Language

The Shan language (Shan written: လိၵ်ႈတႆး, pronounced [lik táj]; Shan spoken: ၵွၼ်းတႆး, pronounced [kwá:m táj]; Burmese: ရှမ်းဘာသာ) is the native language of the Shan people and is mostly spoken in Shan State, Burma. There are also Shan speakers in Kachin, Kayah, Mandalay and in the Sagaing regions.

In 2017, there were about 4.59 million speakers of Shan in Myanmar [9]. The alternative names of "Shan" are Burmese Shan, Great Thai, Sam, Sha, Shan Bama, Shan Gyi, Tai, Tai Long, Tai Luang, Tai Shan, Tai Yai, Tai-Lon and Thai Yai.

Shan is a tonal language with five tones, plus a sixth which is used for emphasis. (တူၺ်းသိၼ် ၎ တူၺ်း ၊ "ႇ ၊ ◌ၸ ၊ ◌း ၊ ◌ႉ ၊ ◌ႈ" in Shan). The basic components of Shan language are 19 consonants (တူၺ်ဢမ်းလိၵ်း ၱၼ် တူၺ်ဝ် in Shan), 10 basic vowels (ဢမ်းၵပ်းငဝ်း ၱ၀ တူၺ်ဝ် in Shan), 15 diphthongs (ဢမ်းၵပ်းသွၼ်ႇ ၱ�support တူၺ်ဝ် in Shan), 3 medial diacritics (ဢမ်းသိၼ်သွၼ်ႇ ၼ် တူၺ်ဝ် ၊ "ျ" ၊ "ြ" ၊ "ွ" in Shan), 60 final vowels (တူၺ်ဝ်ၽိတ်းသိၼ်လင်း ၊ မ-ဆ-ဇ ၼ၀ တူၺ်ဝ် and တူၺ်ဝ်ၽိတ်းသိၼ်ခၢတ်, ၊ ပ-တ-ၵ ၼ၀ တူၺ်ဝ် in Shan). Nineteen Shan Consonants (ၵ to ၯ) are described in Table I. Ten basic vowels are shown in Table II. Shan numbers (zero to nine: 10 digits) are shown in Table III. Writing the Shan numbers is very easy. For example, 1 equals ၱ , 10 equals ၱ၀, 100 equals ၱ၀၀, and etc.

Word order of Shan sentence is Subject-Verb-Object (SVO). Grammatical order of Shan language is like English. The followings are example Shan sentences:

Shan: သူ ၵမ်, ယွၼ်းဆွမ်း မၼ်းဆၢင်း ။
Myanmar: ခင်ဗျား သူမ ကို မ တောင်းပန် ခဲ့ဘူး။
English: You didn't apologize her.

Shan:ၵွပ်းခၢး ပွင်,လွင် တီး သူ ယူ, ။
Myanmar:ကျွန်တော်တို့ ခင်ဗျား ကို နားလည် ကြပါတယ် ။
English:We understand you.

IV. Methodology

A. Phrase-Based Statistical Machine Translation (PBSMT)

A PBSMT translation model is based on phrasal units [10]. Here, a phrase is simply a contiguous sequence of words and generally, not a linguistically motivated phrase. A phrase-based translation model typically gives better translation performance than word-based models. We can describe a simple phrase-based translation model consisting of phrase-pair probabilities extracted from the corpus and a basic reordering model, and an algorithm to extract the phrases to build a phrase-table [11].

The phrase translation model is based on a noisy channel model. To find the best translation $\hat{e}$ that maximizes the translation probability $P(e|f)$ given the source sentences; mathematically. Here, the source language is French and the target language is English. The translation of a French sentence f into an English sentence e is modeled as equation 1.

$$\hat{e} = argmax_e P(e|f) \tag{1}$$

TABLE I
Shan Consonants

ၵ	ၶ	ၷ
ၸ	သ	ၺ
တ	ထ	ဆ
ပ	ၽ	ၾ
မ	ယ	ရ
လ	ဝ	ၻ
	ၿ	

TABLE II
SHAN BASIC VOWELS

10 Basic Vowels of Shan Language									
ၵ	ၵၢ	ၵီ	ၵေ	ၶ	ၵူ	ၵူဝ်	ၵော	ၵိုဝ်	ၵိူဝ်

TABLE III
SHAN NUMBERS

Shan Numbers (zero to nine: 10 digits)									
၀	၁	ၢ	ၣ	ၤ	ၥ	ၦ	ၧ	ၨ	ၩ

Applying the Bayes' rule, we can factorized the $P(e|f)$ into three parts.

$$P(e|f) = \frac{P(e)P(f|e)}{P(f)} \tag{2}$$

The final mathematical formulation of phrase-based model is as follows:

$$argmax_e P(e|f) = argmax_e P(f|e)P(e) \tag{3}$$

We note that denominator $P(f)$ can be dropped because for all translations the probability of the source sentence remains the same. The $P(e|f)$ variable can be viewed as the bilingual dictionary with probabilities attached to each entry to the dictionary (phrase table). The $P(e)$ variable governs the grammatically of the translation and we model it using n-gram language model under the PBSMT paradigm.

B. Hierarchical Phrase-based Statistical Machine Translation (HPBSMT)

The hierarchical phrase-based SMT approach is a model [3] based on synchronous context-free grammar. The model is able to be learned from a corpus of unannotated parallel text. The advantage this technique offers over the phrase-based approach is that the hierarchical structure is able to represent the word reordering process. The reordering is presented explicitly rather than encoded into a lexicalized reordering model (commonly used in purely phrase-based approaches). This makes the approach particularly applicable to language pairs that require long-distance reordering during the translation process [12]. An example of hierarchical phrase-based grammar rules between Myanmar and Shan from the HPBSMT model is as follows:

ရောက် [X][X] [X] ||| ၸွတ်း [X][X] [X]
ရောက် ဖူး [X] ||| ယၢမ်း ၸွတ်း [X]
ရောက် ဖူး တယ် [X] ||| ယၢမ်း ၸွတ်း ၵိုဝ်း [X]
ရောက် ဖူး တယ် ။ [X] ||| ယၢမ်း ၸွတ်း ၵိုဝ်း ။ [X]

C. Operation Sequence Model (OSM)

The operation sequence model can combine the benefits of two state-of-the-art SMT frameworks named n-gram-based SMT and phrase-based SMT. This model simultaneously generates source and target units, and does not have spurious ambiguity that is based on minimal translation units [13], [14]. It is a bilingual language model that also integrates reordering information. OSM motivates a better reordering mechanism that uniformly handles local and non-local reordering, and strong coupling of lexical generation and reordering. It means that OSM can handle both short and long-distance reordering. The operation types are such as generate (generation of a sequence of source and target words), insert gap (insertion of gaps as explicit target positions of reordering operations), and jump (forward and backward jump which performs the actual reordering) [15].

V. EXPERIMENTS

A. Corpus Statistics

We used 11,372 Myanmar sentences (without name entity tags) of the ASEAN-MT Parallel Corpus [16], which is a parallel corpus in the travel domain. It has six main categories such as people (greeting, introduction, and communication), survival (transportation, accommodation, and finance), food (food, beverage and, restaurant), fun (recreation, traveling, shopping, and nightlife), resource (number, time, and accuracy), and special needs(emergency and health). Manual translation to Shan language was done by native Shan students from University of Technology, Yatanarpon Cyber City (UTYCC), and the translated corpus was checked by the editor of Shan magazine. Word segmentation for Shan was done manually and there are exactly 1,265,163 words in total. We used 10,000 sentences for training, 800 sentences for tuning process, and 572 sentences for evaluation respectively.

B. Word Segmentation

In both Myanmar and Shan text, spaces are used to separating phrases for easier reading. It is not strictly necessary, and these spaces are rarely used in short sentences. There are no clear rules for using spaces, and thus spaces may (or may not) be inserted between words, phrases, and even between a root word and their affixes. Although Myanmar sentences of the ASEAN-MT corpus is already segmented, we have to consider some rules for manual word segmentation of Shan sentences.

We defined Shan "word" to be meaningful units and affix, root word and suffix(s) are separated as "ဢိၼ်ၶဵဝ်း". Here, "ဢိၼ်" ("eat" in English) is a root word and suffix "ၶဵဝ်း". Different from the Myanmar language, Shan plural nouns are not followed by particle. Shan plural nouns are written similarly to English plural words. For example, a Shan word "ၵၼ်ၽိတ်း" ("mistakes" in English) is a single word unlike Myanmar word "အမှားတွေ" which is segmented as two words "အမှား" and the particle "တွေ".

Moreover, in our manual word segmentation rules, compound nouns are considered as one word and thus, Shan compound word "ထူင် + ငိၼ်း" ("bag" + "money" in English) is written as one word "ထူင်ငိၼ်း" ("wallet" in English). Shan adverb words, for example, "ၽင်းၽင်း" ("quickly" in English) are also considered as one word. The following is an example of word segmentation for a sentence in our corpus and the meaning is "which mistakes did you make."

Unsegmented Shan sentence:
မိုးဵတ်းၵၼ်ၽိတ်းၵီးသင်။

Word Segmented Shan sentence:
မိုး ဵတ်း ၵၼ်ၽိတ်း ၵီး သင် ။

In this example, "ၵၼ်ၽိတ်း" ("mistakes" in English)

is a single word which is a compound word of "အမှား + တွေ"("အမှားတွေ" in Myanmar).

C. Syllable Segmentation

Generally, similar to the Myanmar language, Shan words are composed of multiple syllables, and most of the syllables are composed of more than one character. Syllables are then composed and Shan words are formed. If we only focus on consonant-based syllables, the structure of the syllable can be described with Backus normal form (BNF) as follows:

Syllable := CMV[CK][D]

Here, "C" stands for consonants, "M" for medials, "V" for vowel, "K" for vowel killer character, and "D" for diacritic characters. Myanmar syllable segmentation can be done with a rule-based approach, finite state automaton (FSA) or regular expressions (RE) [1]. In our experiments, we used RE based Myanmar syllable segmentation tool named. For Shan syllable segmentation we performed an update on RE of the Myanmar syllable segmentation tool. The following is an example of syllable segmentation for a Shan sentence in our corpus and the meaning is "which mistakes did you make."

Unsegmented Shan sentence:
မိုးဵတ်းၵၼ်ၽိတ်းၵီးသင်။

Syllable Segmented Shan sentence:
မိုး ဵ်တ်း ၵၼ် ၽိတ်း ၵီး သင ။

D. Moses SMT System

We used the PBSMT, HPBSMT, and OSM provided by the Moses toolkit [17] for training the PBSMT, HPBSMT, and OSM statistical machine translation systems. The word segmented source language was aligned with the word segmented target languages using GIZA^{++} [18]. The alignment was symmetrized by grow-diag-final-and heuristic [10]. The lexicalized recording model was trained with the *msd-bidirectional-fe* option [19]. We use KenLM for training the 5-gram language model with interpolated modified *Kneser-Ney discounting* [20], [21]. Minimum error rate training (MERT) [22] was used to tune the decoder parameters and the decoding was done using the Moses decoder (version 2.1.1) [17]. We used the default settings of Moses for all experiments.

VI. Evaluation

We used two automatic criteria for the evaluation of the machine translation output. One was the de facto standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [4] and the other was the Rank-based Intuitive Bilingual Evaluation Measure (RIBES) [5]. The BLEU score measures the precision of n-gram (overall n 4 in our case) concerning to a reference translation with a penalty for short translations [4]. Intuitively, the BLEU score measures the adequacy of the translation and large BLEU scores are better. RIBES is an automatic evaluation metric based on rank correlation coefficients modified with precision and special care is paid to the word order of the translation results. The RIBES score is suitable for distance language pairs such as Myanmar and English. Large RIBES scores are better.

VII. Results and Discussion

The BLEU and RIBES score results for statistical machine translation experiments with PBSMT, HPBSMT, and OSM (using word segmentation) are shown in Table IV. Bold numbers indicate the highest scores among the three SMT approaches. The RIBES scores are inside the round brackets. Here, "my" stands for Myanmar, "sh" stands for Shan, "src" stands for source language and "tgt" stands for target language respectively. To compare with syllable results, the translation results were decomposed into their constituent syllables to ensure that the results are cross-comparable. The BLEU and RIBES score results for machine translation experiments with PBSMT, HPBSMT and OSM using word segmentation (evaluation with syllable units) between Myanmar and Shan languages are shown in Table V. Finally, The BLEU and RIBES score results for machine translation experiments with PBSMT, HPBSMT, and OSM (using syllable segmentation) between Myanmar and Shan languages are shown in Table VI. According to the results, with word segmentation, OSM method outperforms the other two methods for both Myanmar-Shan and Shan-Myanmar translations. Looking at the results in Table IV, V and VI, it is clear that the syllable segmentation scheme was by far the most effective for both Myanmar-Shan and Shan-Myanmar translations. With syllable segmentation, the HPBSMT method achieved the highest BLEU score and the OSM method achieved the highest RIBES score for Myanmar to Shan translation. And, OSM performed best in Shan to Myanmar translation (with syllable segmentation).

Our results with both word and syllable segmentation also indicate that Shan to Myanmar machine translation is better performance than Myanmar to Shan translation direction. Our investigation clearly proved that syllable segmentation for bi-directional Myanmar to Shan machine translation has higher scores than others.

VIII. Error Analysis

We also used the SCLITE (score speech recognition system output) program from the NIST scoring toolkit SCTK version 2.4.10 [23] for making dynamic programming based alignments between reference and hypothesis strings for detail analysis on translation errors (WER: Word Error Rate). The formula for WER can be stated as equation 4:

$$WER = \frac{(I + D + S)100}{N} \tag{4}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, C is the number of correct words and N is the number of words in the reference ($N = S + D + C$). Note that if the number of insertions is very high, the WER can be greater than 100%. Table VII presents the WER percentages of translation between Myanmar and Shan (with syllable segmentation). Table VII results show that OSM gave the

TABLE IV
BLEU and RIBES scores for PBSMT, HPBSMT and OSM using word segmentation

src-tgt	PBSMT	HPBSMT	OSM
my-sh	7.54 (0.64454)	6.94 (0.64525)	**7.79 (0.65029)**
sh-my	8.57 (0.63632)	7.88 **(0.64525)**	**8.76** (0.63669)

TABLE V
BLEU and RIBES scores for PBSMT, HPBSMT and OSM using word segmentation (Evaluation with syllable unit)

src-tgt	PBSMT	HPBSMT	OSM
my-sh	18.14 (0.78623)	17.36 (0.78676)	**19.79 (0.79519)**
sh-my	17.98 (0.76975)	16.64 (0.75738)	**18.49 (0.77456)**

TABLE VI
BLEU and RIBES scores for PBSMT, HPBSMT and OSM using syllable segmentation

src-tgt	PBSMT	HPBSMT	OSM
my-sh	33.45 (0.86646)	**34.72** (0.86372)	34.04 **(0.87012)**
sh-my	40.12 (0.87998)	40.51 (0.88023)	**41.85 (0.88031)**

lowest WER for Myanmar and Shan translation in both direction. WER of (46.3%) for Myanmar-Shan translation and (42.8%) for Shan-Myanmar translation respectively.

From our studies, the top 10 confusion pairs for Myanmar-Shan and Shan-Myanmar OSM machine translation (with syllable segmentation) can be seen in Table VIII and Table IX.

We also made manual error analysis on translated outputs of the best OSM model, and we found that dominant errors are different in sentence level. Paraphrasing error is one of the common translation error patterns and they are interesting. For example:

Source: သူ တေ ၼၢ်ႉ မိူဝ်း လွဵ ။
Scores: (#C #S #D #I) 5 3 0 2
REF: ခင် ဗျား ဘယ် *** တော့ ပြောင်း ****** မှာ လဲ ။
HYP: ခင် ဗျား ဘယ် အ ချိန် ပြောင်း လိ မ့် မယ် ။
Eval: I S I S S

Source: မ်ိး ဇွိင်. ဆီ ကမ်, လံး လၢတ်; ။
Scores: (#C #S #D #I) 6 1 1 4
REF: ********** နင် ဒီ လို မ ပြော ရ ********* *** ********** ဘူး ။
HYP: မင်း အဲ ဒီ လို *** ပြော ရ မှာ မ ဟုတ် ဘူး ။
Eval: I S D I I I

One more error type which we noticed is that question sentences are wrongly translated as normal sentences and vice versa.

Source: ဢင် ပီ တၢင်း ယူ, ။
Scores: (#C #S #D #I) 6 1 1 1
REF: ဘယ် သူ ********* လမ်း လျှောက် နေ တာ လဲ ။
HYP: ဘယ် သူ တွေ လမ်း လျှောက် နေ ****** တယ် ။
Eval: I D S

TABLE VII
WER % for PBSMT,HPBSMT and OSM with syllable segmentation

src-tgt	PBSMT	HPBSMT	OSM
my-sh	46.5%	46.5%	**46.3%**
sh-my	44.7%	44.3%	**42.8%**

TABLE VIII
The top 10 confusion pairs of OSM model for Myanmar-Shan machine translation

Frequency	Reference ==> Hypothesis
15	ကျေး ==> ကိူဝ်း
14	သူ ==> မ်ိး
13	ယူ, ==> ကိူဝ်း
12	ယဝ်. ==> ကိူဝ်း
7	ဆဆ်. ==> ဆီ.
7	ဆီ. ==> ဆီ
4	ကိူဝ်း ==> ယူ,
4	ဓၢး ==> ကိူဝ်း
4	ဆၢ, ==> ကိူဝ်း
3	ယူ, ==> ၶၢ.

TABLE IX
The top 10 confusion pairs of OSM model for Shan-Myanmar machine translation

Frequency	Reference ==> Hypothesis
14	မင်း ==> ဗျား
13	မယ် ==> တယ်
13	လဲ ==> တယ်
10	ပဲ ==> တယ်
7	တာ ==> တယ်
7	အဲ ==> အဲ
7	အဲ့ ==> အဲ့
6	တယ် ==> ပဲ
6	တယ် ==> ပြီ
6	တယ် ==> လဲ

Source: ပုင်, ဇၢမ်, ကဆ် သူ ရျိူင်း ဆဆ်. မၢး လံး ။
Scores: (#C #S #D #I) 9 2 0 2
REF: ခင် ဗျား ကြိုက် တဲ့ ပုံ စံ နဲ့ လာ နိုင် ****** *** တယ် ။
HYP: ခင် ဗျား ကြိုက် တဲ့ ပုံ စံ ကို လာ နိုင် ပါ သ လဲ ။
Eval: S I I S

We also found that some of the translation errors are happening in some part of the translated sentences. For example,

Source: သူ ယူ, ဂၢင်း ဂၢင်း ဆီ. ဂူလုၵ်း လီ ယဝ်. ။
Scores: (#C #S #D #I) 6 3 2 2
REF: ခပ် ကင်း ကင်း နေ ********** ********** တာ ပဲ ပို ကောင်း ပါ တယ် ။
HYP: ********* ခင် ဗျား နေ ဂၢင်း ဂၢင်း က ပဲ ********* ကောင်း ပါ တယ် ။
Eval: D S S I I S D

Source: လဝ်း ခွင် ဆီ. တေ ဧတ, မိူဝ်း လွဲ ။
Scores: (#C #S #D #I) 3 7 0 0
REF: ရှပ် ရှင် က ဘယ် အ ချိန် စ တာ လဲ ။
HYP: တို့ ဟာ ကို ဘယ် တုန်း က တည် သ လဲ ။
Eval: S S S S S S S

Moreover, some Myanmar sentences (containing particles and post-positional makers) are incorrectly translated because those particles and post-positional makers are not used in Shan sentences. Example errors are as follows:

Source: တီႈ ၼၼ်ႉ ၽႂ်ႉ ၵေႃႈ ၵမ်ႉ ႁူႉ ။
Scores: (#C #S #D #I) 9 1 2 0
REF: အဲ့ ဒီ ကို ဘယ် သူ မှ မ သွား ကြ ပါ ဘူး ။
HYP: အဲ့ ဒီ မှာ ဘယ် သူ မှ မ သွား ****** ****** ဘူး ။
Eval: S D D

Source: တူဝ် ၸိူင်း မိူႈ မၼ်း မီး ယူႇ ႁႃႉ ။
Scores: (#C #S #D #I) 6 1 2 0
REF: သူ ကိုယ် ပူ ချိန် ကော ရှိ သ လား ။
HYP: သူ ကိုယ် ပူ *************** ********* ရှိ ရဲ့ လား ။
Eval: D D S

Where "Scores" are operation scores of the Edit Distance [24], "C" is the number of correct words, "S" is the number of substitutions, "D" is the number of deletions, "I" is the number of insertions, "REF" for reference, "HYP" for hypothesis and "Eval" is the ordered sequence of edit operations.

IX. Conclusion

This paper has presented the first study of PBSMT, HPBSMT, and OSM machine translation performances between Myanmar and Shan languages in both directions. We used our developing Myanmar-Shan parallel corpus which has around 11K sentences. We also investigated the effectiveness of two different segmentation schemes: word segmentation and syllable segmentation. The results have clearly shown that syllable segmentation is better than word segmentation. In translation with word segmentation, the OSM approach outperformed other models. In translation with syllable segmentation, the HPBSMT approach achieved the highest translation performance for Myanmar to Shan translation and the OSM approach gave the highest BLEU and RIBES scores for Shan to Myanmar translation. Afterwards, we would like to investigate the accuracy of PBSMT and HPBSMT when training with an additional monolingual language model. We also would like to explore statistical and neural machine translation between Myanmar and Shan languages with more amount of data in the future.

Acknowledgment

We would like to express our gratitude to U Sai Kham Sint, Shan freelance translator and editor of Shan magazine, for his patience, translation, and valuable advice during the period of the paper. We would like to express our special thanks to students of the Myanmar-Shan translation team from the University of Technology, Yatanarpon Cyber City (UTYCC). Namely, Sai Kwang Kham, Sai Nyan Lin, Nang Mwe Nom and Nang Vo Hein. Last but not least, thanks to National Electronics and Computer Technology Center (NECTEC), Thailand for sharing the ASEAN-MT corpus.

References

[1] https://github.com/ye-kyawthu/sylbreak
[2] Y. Kyaw Thu, A. Finch, W. Pa Pa, and E. Sumita, "A large-scale study of statistical machine translation methods for Myanmar language," in The Eleventh International Symposium on Natural Language Processing (SNLP-2016), Phranakhon Si Ayutthaya, Thailand, Feb, 2015.
[3] D. Chiang, "Hierarchical phrase-based translation," Computational Linguistics, vol. 33, no. 2, pp. 201–228, 2007.
[4] K. Papineni, S. Roukos, T. Ward, and Z. Wei Jing, "BLEU: A method for automatic evaluation of machine translation," in IBM Research Report rc22176 (w0109022), Thomas J. Watson Research Center, 2001.
[5] H. Isozaki, T. Hirao, K. Duh, K. Sudoh, and H. Tsukada, "Automatic evaluation of translation quality for distant language pairs," in Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, pp. 944–952, Association for Computational Linguistics, Oct 2010.
[6] W. Pa Pa, Y. Kyaw Thu, A. Finch, and E. Sumita, "A study of statistical machine translation methods for under resourced languages," SLTU-2016 5th Workshop on Spoken Language Technologies for Under-resourced languages (SLTU Workshop), vol. 81, 09-12 May 2016.
[7] Y. Kyaw Thu, V. Chea, A. Finch, M. Utiyama, and E. Sumita, "A large-scale study of statistical machine translation methods for Khmer language," in Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, pp. 259–269, Oct 2015.
[8] Y. Kyaw Thu, M. Hti Seng, T. Myint Oo, D. Wom, H. M. Thint Nu, S. Mai, S. Thepchai, and K. Mar Soe, "Statistical machine translation between Kachin and Rawang," in Proceedings of the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019), Chiang Mai, Thailand, pp. 329–334, 30 Oct to 1 Nov 2019.
[9] https://omniglot.com/writing/shan.htm
[10] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp. 127–133, 2003.
[11] L. Specia, "Fundamental and new approaches to statistical machine translation," Emerging Applications of Natural Language Processing: Concepts and New Research, May 2012.
[12] F. Braune, A. Ramm, and A. Fraser, "Long-distance reordering during search for hierarchical phrase-based SMT," in Proceedings of the Annual Conference of the European Association for Machine Translation (EAMT), Trento, Italy, pp. 177–184, 2012.
[13] N. Durrani, H. Schmid, and A. Fraser, "A joint sequence translation model with integrated reordering," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, (Portland, Oregon, USA), pp. 1045–1054, Association for Computational Linguistics, Jun 2011.
[14] N. Durrani, H. Schmid, A. Fraser, P. Koehn, and H. Schütze, "The operation sequence model combining n-gram-based and phrase-based statistical machine translation," Computational Linguistics, vol. 41, pp. 157–186, Jun 2015.
[15] P. Koehn," Moses, Statistical Machine Translation System, User Manual and Code Guide," University of Edinburgh: http://www.statmt.org/moses/manual/manual.pdf
[16] B. Prachya and S. Thepchai, "Technical report for the network-based ASEAN language translation public service project," in Online Materials of Network-based ASEAN Languages Translation Public Service for Members, NECTEC, 2013.
[17] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," in Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, Prague, Czech Republic, pp. 177–180, Association for Computational Linguistics, Jun 2007.
[18] F. J. Och and H. Ney, "Improved statistical alignment models," in Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, Hong Kong, pp. 440–447, Association for Computational Linguistics, Oct 2000.
[19] C. Tillmann, "A unigram orientation model for statistical machine translation," in Proceedings of HLT-NAACL 2004: Short Papers, Boston, Massachusetts, USA, pp. 101–104, Association for Computational Linguistics, 2-7 May 2004.
[20] K. Heafield, "KenLM: Faster and smaller language model queries," in Proceedings of the Sixth Workshop on Statistical Machine Translation, Edinburgh, Scotland, pp. 187–197, Association for Computational Linguistics, Jul 2011.
[21] S. F. Chen and J. Goodman, "An empirical study of smoothing techniques for language modeling," in 34th Annual Meeting of the Association for Computational Linguistics, Santa Cruz, California, USA, Association for Computational Linguistics, pp. 310–318, Jun 1996.
[22] F. J. Och, "Minimum error rate training in statistical machine translation," in Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, pp. 160–167, Association for Computational Linguistics, Jul 2003.
[23] (NIST) The National Institute of Standards and Technology, Speech recognition scoring toolkit (SCTK), version: 2.4.10, 2015.
[24] F. P. Miller, A. F. Vandome, and J. McBrewster, Levenshtein Distance: Information Theory, Computer Science, String (Computer Science), String Metric, Damerau-Levenshtein Distance, Spell Checker, Hamming Distance. Alpha Press, 2009.

Daily Health Monitoring Chatbot with Linear Regression

Konlakorn Wongpatikaseree
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakhon Pathom, Thailand
konlakorn.won@mahidol.ac.th

Arunee Ratikan
Sirindhorn International Institute of Technology,
Thammasat University
Pathum Thani, Thailand
a.ratikan@gmail.com

Chaianun Damrongrat
Speech and Text Understanding
National Electronics and Computer Technology Center
Pathum Thani, Thailand
chaianun.damrongrat@nectec.or.th

Katiyaporn Noibanngong
Department of Computer Engineering
Mahidol University International College
Nakhon Pathom, Thailand
katiyaporn.noi@student.mahidol.edu

***Abstract*— Nowadays, elderly people harm their health from daily routines such as the habits of eating junk food, lack of time for exercising, and so on. When age increases, most of elderly people often have encountered with high blood pressure disease. This disease can lead to many dangerous diseases. Additionally, population of the doctors in hospital is not so high. The doctor usually takes a short note about symptoms for quick services that might not be enough for diagnosis. Therefore, we proposed daily health monitoring chatbot for the elderly people. We need to collect information from the elderly people to create personal health record (PHR). We develop conversational chatbot to interact with the elderly people via LINE application. Outcomes of this research support the doctor's work because after reading daily PHR, the doctor diagnoses the diseases and gives advice for medical treatment more accurately. Furthermore, Linear regression technique was developed to monitor the blood pressure's trend of elderly. Thus, they can prevent or relieve some diseases from chatbot warning and taking care the health.**

Keywords—chatbot, personal health record, blood pressure disease, health monitoring

I. INTRODUCTION

Nowadays, people do not notice about their health, especially elderly people. This group of people might know that their health may is not in a good condition when it is too late. There are many reasons that people have not noticed that can harm their health based on daily routines. The fact that the pollution in the atmosphere, the daily workload of people, the habits of eating junk food, lack of time for exercising, and so on. Commonly, most of the elderly people often have blood pressure disease when they are getting old age. The problem is that people are unaware of health is on crisis or in the state of danger. The symptoms of having a high blood pressure (hypertension) [1] could be known as severe headache, fatigue or confusion, vision problems, chest pain, breathing difficulty, irregular heartbeat, blood in urine.

Although the numbers of patient in the hospital increases, the amount of the doctors is not growing with the same rate. Therefore, when the elderly people go to a hospital or a clinic, elderly's health will be diagnose just in the hospital. Historical health data such as daily blood pressure, daily food intake, or daily activity is not shown in the report. It might not be enough for identifying the disease precisely. Even though the doctor might ask the elderly people about history of health, the elderly people cannot answer correctly because they forget to notice their daily routine.

Thus, the doctor needs to know personal health record (PHR) [2] [3] that contains blood pressure record, food and beverage intake record, activity record and weight record [4] and other information. The advantages of PHR, are that it supports work of the doctor to diagnose the disease and give medical treatment more accurately. PHR could be used as supporting medical information for the elder people to inspect and understand of health's status so that they can prevent or relieve some diseases. Good taking care of the health condition can reduce the cost of medical treatment in the future. However, to have daily PHR of the elderly people, it is not an easy task because to take a daily information from the elderly, technology should encourage them to concern about the benefit of the daily information such as blood pressure, weight, food intake, daily activity, and so on.

In this paper, we take advantages of Information and Communications Technology (ICT) to support health care in the group of elderly people. We have proposed daily health monitoring chatbot for the elderly people. We aim to create daily PHR by using chatbot technology. The conversational chatbot will ask information (blood pressure, food, beverage, activity, and weight) via Line application. The chatbot will also ask a deeper question to the elderly people to collect more information to help in the further investigation. Therefore, conversation dialog is important for this research. We consider bot persona, mood and tone in designing whole conversation flow. The conversation dialog in this research gives feeling care, warm but it still is formal. The information from the elderly people will not be revealed to other parties. AI module with linear regression is implemented to monitor the blood pressure. Consequently, this chatbot will be able to notify or warn the elderly people via Line application when trend of blood pressure is abnormal.

II. BLACKGROUND AND RELATED WORKS

A. High blood pressure

Blood pressure is when heart beat squeezes and pushes the blood from the heart to the rest of the body. The normal blood pressure for adults is less than 120 mmHg for systolic and less than 80 mmHg for diastolic. The measures of the high blood pressure stage 1 could be measured around 130-139 mmHg for systolic or higher than 90 mm Hg for diastolic. The measures of high blood pressure stage 2 is to be measured at 140 mmHg or higher for systolic or 90 mmHg or higher for diastolic. In the worst case, the hypertensive crisis is measured at 180 mmHg or higher for systolic and/or 120

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

mmHg or higher for diastolic. The symptoms of having a high blood pressure could be known [1] as severe headache, fatigue or confusion, vision problems, chest pain, breathing difficulties, irregular heartbeat, blood in urine. Causes of high blood pressure could be considered from age, overweight, relatives with high blood pressure, insufficient nutrition, lack of exercise, excessive intake of coffee or alcohol, smoking, and not enough sleeping.

B. Body Mass Index

Body mass index (BMI) is the system to calculate the person's body in order to tell them that they are healthy or overweight. The rough calculation is a person's weight in kilograms divided by the height. Overweight BMI is 27.5. The obesity range is at 30 or more. Obesity [5] occurs when the body consumes more calories than burning. Not only eating is in the factor, it also includes genetic, environment, behavior, and social factors. The symptoms and illness of being in the state of obesity are breathing disorders, cancers, heart disease, depression, diabetes, high blood pressure, high cholesterol, joint disease, and stroke. Obesity also could be linked into psychosocial problems.

C. Related works

R. B. Mathew et al.[6] and S. Divya et al.[7] developed a health care chatbot application by using natural language processing and machine learning. This chatbot application could identify disease's symptom from interaction between a bot and patient. In these works, the patient could save time to consult the doctor at a hospital. The chatbot also predicted the disease and provided necessary recommendation for treatment. A. K. Tripathy et al.[8] conducted a healthcare management system. This system had a duty like a doctor, who the user could rely on. The user needed to carry a mobile heart rate measurement in order to detect the heart rate. Moreover, the user could talk with the doctor in case of urgency via video conferencing. This system would diagnose the disease based on the user's symptom. B. E. V. Comendador et al.[9] introduced a computer application, which provided a conversational chatbot to interact the user. The user could be the patient, children' parents, who did not understand about medicines. This chatbot acted as a medicine consultant. It prescribed, suggested and gave information on generic medicines for children. H. Lal et al.[10] introduced human computer interaction system that use NLP chatbot. Their works could answer the patients' question and diagnose the disease based on analyzing the data from clinical note. For the specific information, it could be searched on the system by analyzing big data.

III. METHODOLOGY

In this paper, we applied a new chatbot technology to health care. We have proposed daily health monitoring chatbot for the elderly people. We collected daily information from elderly people by chatting via Line application. Figure 1 presents the system architecture that consists of four modules, which are users, line bot, NLP module and cloud system.

A. User

User of the system is the elderly people, who will be asked for collecting personal health information such as blood pressure, food intake, or activity in every day. The elderly people will not be asked directly. We will use everyday conversations so that the elderly people are relaxed. This group of elderly people is required to use Line application.

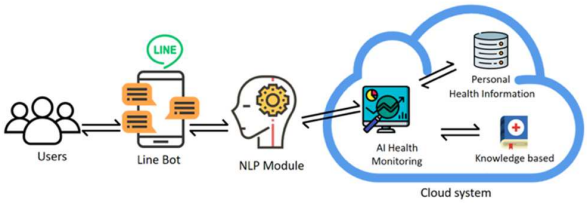


Fig. 1. Daily health monitoring chatbot for the elderly people

B. Line Bot

Line bot is a channel to communicate between the elderly people and system. We use Line application because it is the most popular application in Thailand, provide API for developer to connect this application with other system. When system found that abnormal trend of blood pressure, the system will notify or warn the elderly people via Line application. If the elderly people forgot to input the blood pressure, Line application will be activated to ask the elderly people. However, the elderly people can ask more detail about hypertension, their own PHR via this application.

C. NLP module

In this research, Dialogflow [11] will be implemented with Line application directly. Dialogflow is used as a main part of NLP module in this chatbot to categorize the message from Line application. Dialogflow could trigger the event handler to push message from bot. In this module, it composes of many sub-modules such as user, intent, fulfillment, integration and so on. The elderly people' message is analyzed in this module to choose the action to perform. The personal health information in the message will be stored in database (MongoDB) for further process in the cloud system module. Dialogflow can connect with cloud system via webhook in fulfillment module. To have smooth everyday conversation, we created a conversation flow that suitable for elderly people. Bot persona, mood and tone were used to design flow of conversation between elderly and bot.

D. Cloud system

Cloud system is backend part of the chatbot. Figure 1 shows that cloud system consists of three sub-modules: AI Health Monitoring, knowledge, and personal health information. The python server is the python code that is hosted on Google Cloud Instance. AI health monitoring of the chatbot will mainly consists of three functions. The function consisted of the AI function, report function, and Graph function. Linear regression was implemented in the AI function to check and predict the trend of blood pressure and weight, shown in Fig. 2. Mathematic model is used to setup the hypothesis function to predict the type of the trend as shown in eq (1).

$$h(x) = \theta_0 + (\theta_1 * x) \qquad (1)$$

Where hypothesis function or h(x) represents the line mathematically, θ is adjustable variable. It can be adjusted when h(x) is not good enough (This operation is called "optimization"). We can optimize h(x) by checking from

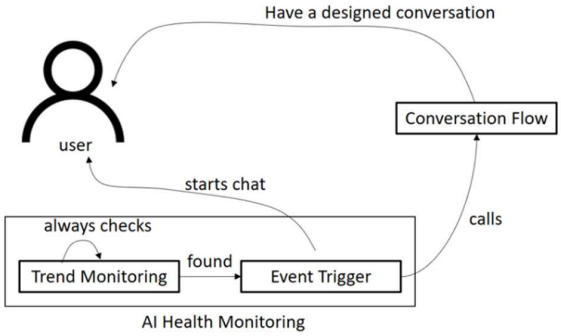


Fig. 2. Warning elderly people from AI monitoring

coefficients with the loss value from the loss function as indicated in eq 2. We also use Gradient Descent Optimizer is to adjust θ value indefinitely until the lowest loss value (optimum point) is achieved.

$$Loss_{fn} = \left(\frac{1}{(2 \times n\_sample)}\right) \times tf.reduce\_sum\left(tf.pow(y - h(x), 2)\right) \quad (2)$$

When AI function found abnormal trend of blood pressure, it will push message to the user via LINE messaging API to start warning the elderly people. Also, AI function will select set of questionnaires in NLP module to ask the elderly people for more information. The selecting questionnaire is based on the type of AI function results.

Graph function plots the blood pressure and weight data into the graph separately. This helps the doctor look at the blood pressure trend and weight trend easier. Report function is the function that generates PHR of the elderly people from the database to pdf file. The report shows both information in term of tables and graphs.

IV. Experiment and Results

Daily health monitoring chatbot for the elderly people has four main features, which are registration, personal information edition, blood pressure input, PHR report creation. Moreover, other information about food and beverage, activity, weight will be extracted in conversation during the chatting.

A. Registration

When the elderly people firstly add this chatbot as a friend in Line application, the chatbot will ask the elderly people for the personal information, including name, age, height, weight, congenital disease, gender, and telephone number. In some of the information, the user can decline to answer the chatbot about their information because privacy is concerned [3]. All of the information will be stored in the database. AI function calculates body mass index by using height and weight as basic information for the doctor. After the registration, the chatbot will show the menu that the elderly people can edit the personal information, input the blood pressure and request to see PHR report.

B. Personal information edition

This feature allows the elderly people to edit their own information. They can choose what information they want to edit such as name, age, height, weight, congenital disease, gender, and telephone number. For example, the elderly people did not answer the weight in the registration process, this feature allows him/her to fill the weight. The changed information will be replaced in the database.

C. Blood pressure input

The blood pressure input feature is the most important feature because the doctor can use all information getting from this feature for diagnose the current diseases precisely such as heart failure, stroke and so on. Moreover, the doctor gives a good advice to prevent or relieve the future diseases.

This feature allows the elderly people to input the blood pressure values anytime by choosing option in the menu. If the elderly people forget to input in each day, AI function will send some signal to the chatbot. Then, it calls to Line application to start asking the elderly people. Before measuring the blood pressure values, the elderly people should stay in a quiet place and relax. Additionally, they are advised to sit and rest at least 30 minutes early if they have previously been smoking or drinking coffee and caffeinated beverages. Next, the chatbot will ask the elderly people to input their current blood pressure values both systolic blood pressure and diastolic blood pressure. These two values will be stored in the database and AI function will process the blood pressure trend by using linear regression and other works as indicated in Fig. 3.

- **Inputting blood pressure**

The elderly people might input the blood pressure in different formats such as 130/80, up 130 down 80, 130 80 and so on. Thus, it needs to format this input into two values: systolic and diastolic further process in the AI function.

- **Analysis by linear regression and Questionnaire about high blood pressure**

AI function will calculate these two values and find the trend by using linear regression as explained in section III. However, there is a cold start problem because the regression requires enough data to find trend. This research cannot diagnose the diseases as a doctor, but it helps the elderly people screen the risk of diseases, which trend to occur in the current situation and future. It also supports the doctor's work to prepare PHR from past until present for diagnosing the diseases. There are three types of blood pressure results from the linear regression's prediction: high blood pressure, normal blood pressure and low blood pressure. In this research, we focus on only high blood pressure result because it is more dangerous than other two results. We classify the high blood pressure results into three types as shown in Fig 4., which are continuous increase of blood pressure, peak of blood pressure, none increasing of high blood pressure. For the continuous increase of blood pressure, the elderly people have high blood pressure continuously (systolic over 120 mmHg and/or diastolic over 80 mmHg) as shown in Fig 4 (a).

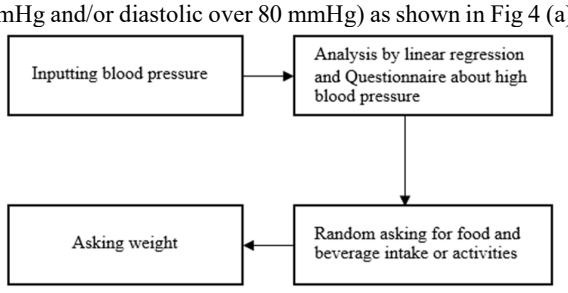
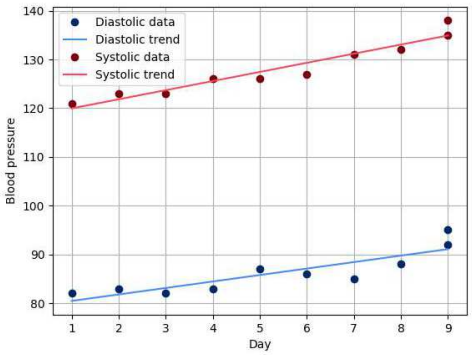


Fig. 3. AI function process

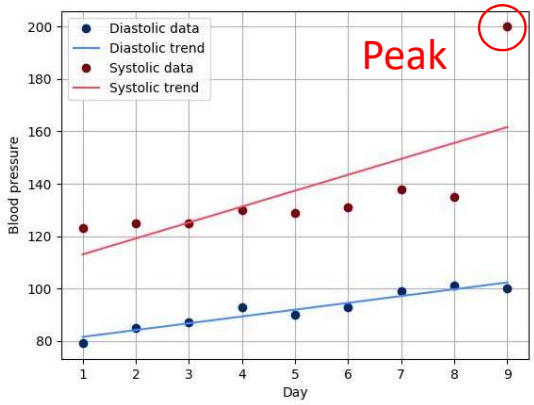
It is dangerous for this group of the elderly people. They might suffer from severe headaches, fatigue or confusion, vision problems and so on. AI function will choose the questionnaire to ask symptom to collect in PHR. The elderly people might answer by themselves or allow another person to answer. Sometimes, the elderly people do not care about high blood pressure. They have symptom, but they think that taking a rest is enough. This cause the severe symptom in the future.

For the peak of blood pressure, the elderly people have a high blood pressure noticeably when compared to the previous pressure as indicated in Fig 4 (b). This case can occur in real situation because before measuring the blood pressure, the elderly people might do exercise, feel shock or sad. This causes heart beat so much squeezes and pushes the blood from the heart to the rest of the body. Nonetheless, taking a rest for a while, the blood pressure will back to a normal state. AI function also asks the elderly people in the questionnaire.

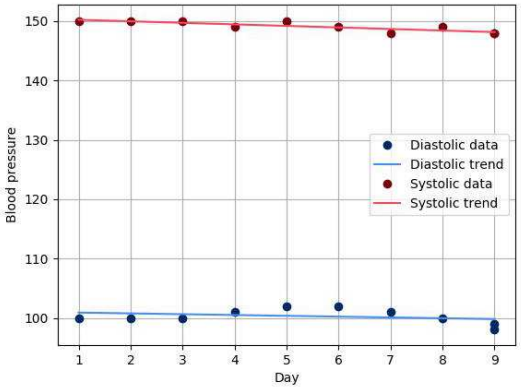
For the none increasing of high blood pressure, the elderly people have high blood pressure as indicated in Fig 4 (c). When the time passed, the blood pressure does not go increase or decrease so much. This is possible that the elderly people forget to take medicine. The medicine might not suit for the elder people' diseases anymore. Moreover, they might not know have high blood pressure as congenital disease. This group of people has high risk because they do not how to take care themselves or control the disease before severe symptom occurs. Thus, AI function choose the questionnaire to ask about family background about high blood pressure, what kind of medicine they have taken, what kind of food and beverage they have eaten, and so on.



a. Continuous increase of blood pressure



b. Peak of blood pressure



c. None increasing of high blood pressure

Fig. 4. High blood pressure types

- **Random asking for food and beverage intake or activities**

To get more information in period of the day, AI function randomly chooses a set of questionnaires either food and beverage intake or activities. The questions in the questionnaire will not repeat. A set of questions in the questionnaire is prepared, because we do not want the elderly people feel bore to answer same questions every time. The number of questions is just two or three questions. For example, the chatbot start asking general questions, then asks about "what food did you eat?", "what dessert did you eat", "what beverage did you drink". The conversation style will be natural and smooth.

Questionnaire about food and beverage intake and activities are important. This is because they affect controlling the high blood pressure disease. If the elderly people, who have high blood pressure, do exercise every day or eat healthy food, this can relieve or control symptom of high blood pressure. Additionally, the elderly people will be healthy and save the money for medical treatment. When the doctor reads the daily activities and food and beverage intake, he/she can understand and diagnose precisely [12].

- **Asking weight**

The high blood pressure relates to body weight. Around 50% of the elderly people, who have BMI more than 25 or overweight, trend to have high blood pressure around 20-30%. Moreover, the obesity is a factor that relates to many diseases such as chronic diseases, diabetes, cardiovascular diseases and cancer. The elderly people should have BMI between 18.5-24.9. Therefore, this research asks the elderly people' weight to measure BMI [12][13].

D. PHR report creation

Report creation is one of key features of daily health monitoring chatbot for the elderly people. the report will be created anytime that the elderly people requests via Line application. It shows daily PHR, which is collected from the elderly people in each period of time in a day. There is three main information in daily PHR as indicated in Fig 5.: registered information, table and graphs. The registered information is personal information that the elderly people registered such as name, gender, weight and so on. The tables in Fig. 5 indicates date and time, food, symptom, activities and others. This information will be retrieved from databased. The doctor can see that which date the elderly people has symptom and find cause of that symptom. There

are two graphs for blood pressure and weight. The graphs illustrate the real data and trend line in order to be easy to see. Therefore, if the elderly people pay attention to input and answer the questionnaires, the doctor will have enough information to diagnose and give advice precisely. In addition, daily PHR helps the elder people be aware their health's condition and find the way to control the disease.

V. Conclusion and Future Improvement

In this research, we adopted a new chatbot technology to health care. We have proposed daily health monitoring chatbot for the elderly people. We focus on collecting daily PHR of elderly people by chatting via Line application. This is because when age increase, some diseases, especially the high blood disease, come to attack the elderly people, some elderly people do not take care themselves until health condition is not so good. This research provides the elderly people to interact with the chatbot. Four main features: registration, personal information edition, blood pressure input, PHR report creation were proposed to help the elderly to keep in touch with their health. Linear regression was proposed in AI module to monitor the trend of elderly's blood pressure. This module can warn to the elderly when high blood pressure trend is going to occur. AI module can classify the high blood pressure results into three types as shown in Fig 4., which are continuous increase of blood pressure, peak of blood pressure, none increasing of high blood pressure. Furthermore, related information such as food intake, weight, or daily activity is also record for giving the additional information to the doctor. Thus, daily PHR supports work of the doctor to diagnose the disease and give medical treatment accurately. PHR helps the elder people be aware their health condition before disease became severe level.

For improvement, we plan to extend style of chatting from text to voice because some elderly people is familiar with voice rather than typing. Furthermore, we plan to integrate system that can extract nutrition of food such as calories, nutrient, vitamin and so on from text input. This can evaluate how useful the food is.

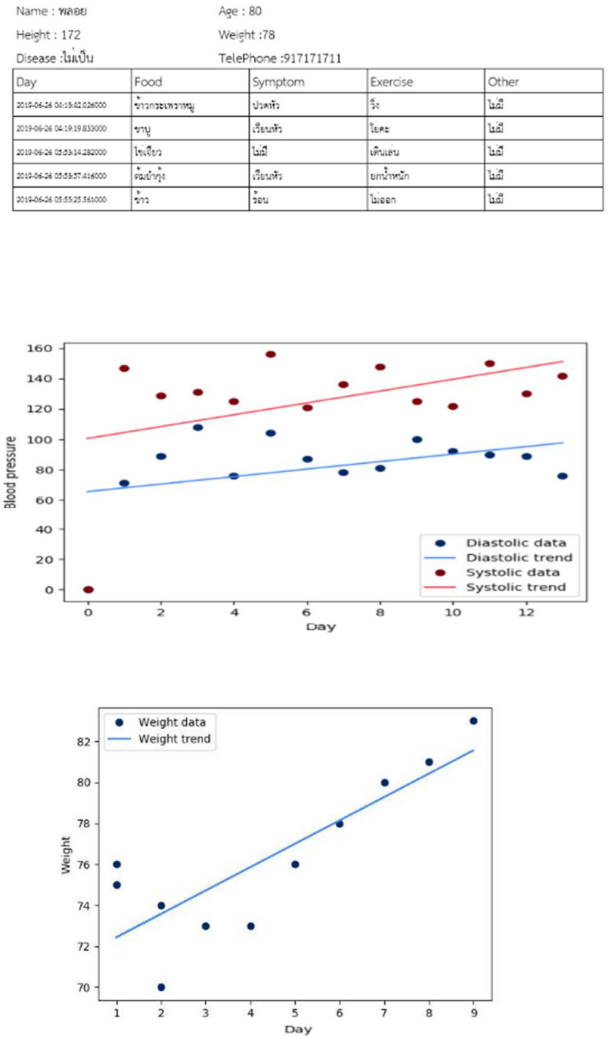


Fig. 5. PHR report

Acknowledgment

This research was supported by Thammasat University research grant TU2017015 for the project "Digital Platform for Sustainable Digital Economy Development", based on the RUN Digital Cluster collaboration scheme. This research was also supported by the Department of Computer Engineering, Faculty of Engineering at Mahidol University.

References

[1] P. Kaushik and P. Sethi, "A Comprehensive Study on Blood Pressure Measurement Techniques," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-3,

[2] V. Ved, V. Tyagi, A. Agarwal and A. S. Pandya, "Personal Health Record System and Integration Techniques with Various Electronic Medical Record Systems," 2011 IEEE 13th International Symposium on High-Assurance Systems Engineering, Boca Raton, FL, 2011, pp. 91-94.

[3] D. Daglish and N. Archer, "Electronic Personal Health Record Systems: A Brief Review of Privacy, Security, and Architectural Issues," 2009 World Congress on Privacy, Security, Trust and the Management of e-Business, Saint John, NB, 2009, pp. 110-120.

[4] R. Shrestha, R. Mahabir and L. Di, "Healthy food accessibility and obesity: Case study of Pennsylvania, USA," 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Fairfax, VA, 2013, pp. 329-333

[5] World Health Organisation. Obesity & Overweight Fact Sheet. February 2018; Available at: http://www.who.int/mediacentre/factsheets/fs311/en/.

[6] R. B. Mathew, S. Varghese, S. E. Joy and S. S. Alex, "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 851-856, doi: 10.1109/ICOEI.2019.8862707.

[7] S. Divya, V. Indumathi, S. Ishwarya, M. Priyasankari, S. Kalpana Devi, "A Self-Diagnosis Medical Chatbot Using Artificial Intelligence", J. Web Dev. Web Des., vol. 3, no. 1, pp. 1-7, 2018.

[8] Amiya Kumar Tripathy, Rebeca Carvalho, Keshav Pawaskar, Suraj Yadav,"Mobile based healthcare management using artificial intelligence" in International Conference on Technologies for Sustainable Development (ICTSD), 2015, 4-6 Feb, 2015

[9] Benilda Eleonor V. Comendador, Bien Michael B. Francisco, Jefferson S. Medenilla, Sharleen Mae T. Nacion, and Timothy Bryle E. Serac, "Pharmabot: A Pediatric Generic Medicine Consultant Chatbot", Journal of Automation and Control Engineering Vol. 3, No. 2, April 2015

[10] H. Lal and P. Lal, "NLP chatbot for Discharge Summaries," 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2019, pp. 250-257, doi: 10.1109/ICCT46177.2019.8969045.

[11] Dialogflow, Available at: https://dialogflow.cloud.google.com

[12] W.D.McArdle,F.I.Katch,andV.L.Katch,ExercisePhysiology: Energy, Nutrition, and Human Performance, 7 ed. Philadelphia: Lippincott Williams & Wilkins, 2010.

[13] R. R. Wing and S. Phelan, "Long-term weight loss maintenance," The American Journal of Clinical Nutrition, vol. 82, no. 1, pp. 222S-225S, 2005

Face Recognition System for Financial Identity Theft Protection

Thidarat Pinthong[1], Worawut Yimyam[2], Narumol Chumuang[3], Mahasak Ketcham[4]

[1,2]Dep. of Business Information Management, Phetchaburi Rajabhat University Phetchaburi , Thailand.
[3]Dep. of Digital Media Technology, Muban Chombueng Rajabhat University, Ratchaburi , Thailand.
[4]Dep. of Management Information System,King Mongkut's University of Technology North Bangkok.

[1]thidarat.pin@mail.pbru.ac.th, [2]worawut_yimyam@hotmail.com, [3]lecho20@hotmail.com, [4]mahasak.k@it.kmutnb.ac.th

***Abstract*— the paper propose a part of image processing for applying on security and authentication. This system looks to be about image processing techniques applied to work in security and data access authentication information through application on mobile. In regard to financial transactions by using Eigen face for face recognition to process and verify its validity. The paper is an idea for prototype applications, the system may not be complete and correct, it can be used practically. However, our work will provide the technical know-only.**

Keywords— Face Recognition, Haar-like model, Eigen face

I. INTRODUCTION

Nowadays, it is more convenient to process financial transactions because there are technologies that provide convenience through all channels such as on computers or smart phones [1], [2] . However, these technologies come with danger since they allow access to important data about financial transactions [3]. For instance, on 31 July 2019 a thief transferred money to Mr. Phansuthee's Kasikorn Bank account by requesting to change a new SIM Card at Truemove [4]. The old password was changed to access K-Mobile Banking application and the amount of 986,700 baht was transferred to other account. This situation has revealed security problems of application on mobile access when processing financial transactions. Accessing data requires Log-in process of OTP or receiving SMS that informs OTP password to verify a transaction [5]-[7]. When a mobile number that receive OTP or SIM Card is changed, a signal at a new SIM card will be registered and a user of the new SIM card will receive OTP password to verify access to mobile application [8]. As a result of such incident, the idea to increase security when accessing financial transactions through mobile application was initiated. BioMatrix technology was applied to verify access [9]-[11], using a combination of face recognition system and a camera on smart phone. Cloud system was used to store master initial image data to be compared and every data processing was recorded [12], [13]. This system will increase complexity and security of data access [14]. This paper aimed to create a security system for processing financial transaction through a user's face recognition.

The organized of our work start with related research give a describe about concept and literatures are related in section II. Next main topic shows the styling system design for understanding of overview workflow and the solutions about our system with step by step. Section IV., the experimental and results and the last is conclusion.

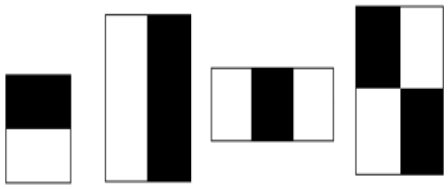
II. RELATED RESEARCH

CCTV face detection and tracking system [9] by the Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang 2018 used knowledge-based methods, Feature Invariant Approaches, Template matching methods, Appearance-based methods, face detection of Viola-Jones [14], [15] . It functions through a program and can monitor via android operating system. The advantage of this system is the accurate data processing and suits for a close building that requires authentication [16]. However, there is an error in reflectivity and the system cannot be used with android operating system. Y. Zhang, et al., [17] proposed a security controlled system on raspberry pi with Eigen face technique and illumination change. They proposed Eigen face technique and illumination change for facial authentication for automatic door opening-closing system. The system consists of a computer and raspberry pi, applying face detection, face recognition and Eigen face for face recognition techniques. PCA algorithm or Eigen faces is studied since it is a less complex algorithm and is appropriate for such context. [18] The advantage of this system is that it used understandable technique and the processing was not complex [19]. However, the system could cause an error when there is a disturbing signal and different illumination. In addition, a study of R. Ranjan et al [20], proposed an intelligent model for system access control by human face verification which applies the biometric technology. The system focuses on human face detection which uses the skin color model to detect the skin color and human face verification which recognizes important face features. The advantage of this system is that it can be used with the disabled, gives accurate results and similar features can be analyzed nearly 100%. However, the system is expensive and it can be used with non-moving devices such as CCTV or a camera which directly scans. It also functions as a program only.

A. Haar-Like Model

Haar-like feature proposed by Viola and Jones [21]. Haar-like feature extraction is a popular technique implemented for finding the difference of the intensity between white and black areas [22]- [24].

978-0-7381-3068-2/20/$31.00 ©2020 IEEE



(a) (b) (c) (d)

Fig. 1. (a) haar-like feature characteristics, (b) two-rectangle features, (c) three-rectangle features, (d) four-rectangle features.

B. Eigen face

Face recognition by Eigen face is a system that analyzes basic facial features through statistical analysis [25], [26]. Eigen vector is a set of vectors derived from covariance matrix [9], [25]-[27]. The Eigen face forms a model that includes outstanding features of all facial images to identify specific values on faces. Individual person has individual value [28].

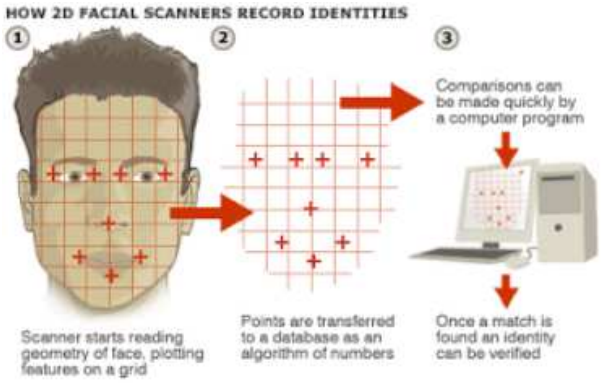


Fig. 2. Face Recognition by Eigen Face.

Eigen face is a techniques for improving face recognition are achieved by converting two-dimensional images of human faces into one-dimensional vectors [29]. And stored in the database and when wanting to compare images of faces, people who are interested will convert that face to a one-dimensional vector. Then compare the vector with the image in the database to find the result For example, the first face comparison is compared with the basic set. It is the same as Eigenface 1 10% same as Eigenface 2 55% same as Eigenface 3 minus 3% when comparing the second face. And then get the percentage of% in the same direction, then page 1 and page 2 are the same page [26],[30].

Eigen face creation steps 1) Prepare images for a variety of tests. Eigen face characteristic, which is an analysis of the basic elements of the face. With a variety of facial statistical processes Therefore, the images must be prepared for the database for review. Which is a set of images for practice This should be a picture that was taken in the light conditions, the position of the eyes and mouth in the same position. And all pictures must have the same resolution By having the same pixel array characteristics The vector of this image is stored as a matrix. 2) Calculate the characteristic page or eigen Vecto. A characteristic face is a set of eigen vectors that can be obtained from the covariance matrix, forming a model of a face that combines the characteristics of a sample face image. Let's come together to find the specific value of the elements on the face. In which each person will be unique to that person. 3) Convert the grayscale image to vector to determine the eigenvalue and apply the eigenvalue of the portrait preview. Let's create a unique face model to find the position of the face. (Distribution-Base Methods) show the distribution of the sample data model with page and non-page to be used as a basis for decision making. 4) Take the Gaussian function to estimate the distribution of the sample mean.

III. STYLING SYSTEM DESIGN

The development of financial transaction processes images received from a mobile phone camera by face recognition of a user. Confidence is consistently monitored when accessing the data. The system displays the access through Real Time mode as shown in Fig 3.

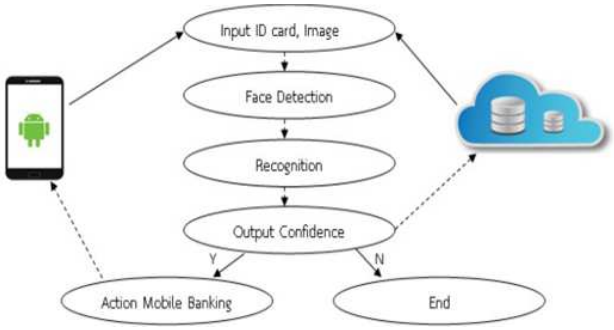


Fig.3. System Functioning.

Fig 3. shows the system functioning. The system receives ID card and images captured by a smart phone camera as well as the data stored on cloud base in order to proceed detection and recognition to find Confidence value. The Confidence values from master images on Cloud Base recorded when a user first opened a bank account are compared. If the Confidence value is close to 0.5, it is True. If not, it is False. The system is divided into 2 sides: 1) trade and: 2) test.

A. Trade Functioning

Trade records the data of Haar Like used for comparison. The trade records master images of users and proceeds Detection to identify face values including eye, nose and mouth. Then, the values are converted into number to be compared with a user's data. The process is shown in Fig 4.

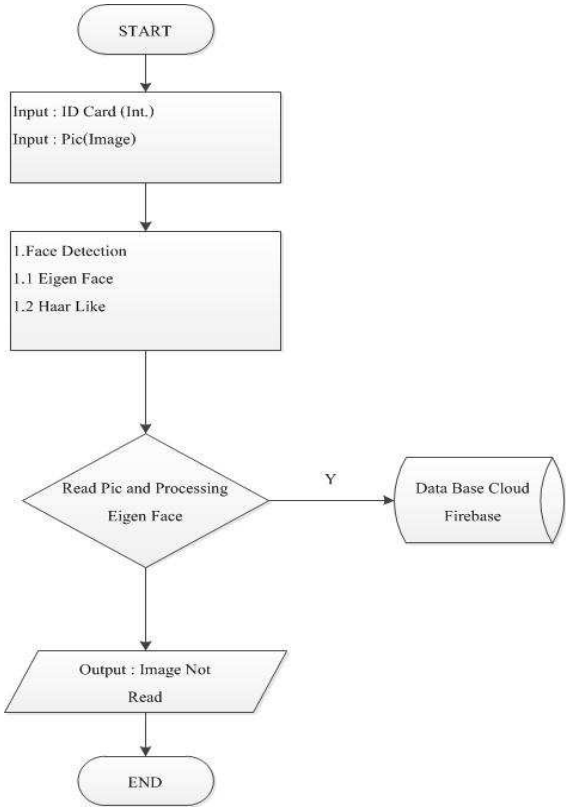


Fig. 4. Trade Functioning.

- Input: Receive ID data and image captured by a smart phone camera.
- Face Detection: Use Eigen Face as follows.
    - 6 images are used in Eigen Face.
    - Use K with the maximum Eigen value (K < M to reduce data size). Each image is N x N and is in matrix format. It is converted to a n 2x 1 vector to make the image for PCA. For instance, a 120 x 120-pixel image equals 14400 vector.

- Algorithm Eigen Face
    - The image is derived from the best K Vector. Then, all images are read as $I1,2,I3,\ldots I$. Faces in all images should be at the center.
    - Each image is converted to vector from N x N matrix to n 2x 1 $\Gamma\ i$ matrix. For instance, 6 images are converted to vectors. Lengthy vectors can be reduced by PCA method.
    - Mean face of vectors of the 6 images is identified.

$$\Psi=\frac{1}{M}\sum_{i=1}^{M}\Gamma_i \qquad (1)$$

Ψ is Mean Face

- Eigenvector of each image is minus Mean Face and new Eigenvector is identified.

$$\Phi_i=\Gamma_i-\Psi \qquad (2)$$

The concept is that a new Eigenvector (Φi) records differences of mean. That is, similar values to mean values are deleted because similar data cannot distinguish individual faces. The values rarely different from the Mean Face value are stored in the database. Therefore, it is important to delete these values by using PCA method.

- Eigenvector is sequenced as A matrix as shown below.

$$A=\{\Phi_1,\Phi_2,\ldots,\Phi_M,\} \qquad (3)$$

- Covariance matrix (C) is identified as shown below.

$$C=A^TA \qquad (4)$$

A is N2 x M matrix and C is M x M matrix.

- The maximum value of Eigenvectors in the 6 images is identified. Each vector is used to form Eigen Face, using linear combination of all faces. It is a set of vectors showing face directions in the database with the most different value.

- Haar Like: Use vector values to process Machine Learning called “Adaptive Boost” or “Adaboost” to identify Weak Classifier. Then, weighted average is identified to reduce errors which can create Strong Classified.
- Read Image and Processing :After Eigen Face process and the best images are identified, the system examines images with constant values again. Then, images read are recorded and converted to numbers to give accurate Confidence values. The data is uploaded on Cloud Base known as Firebase.

B. Test Functioning

- Test functioning is shown in Fig 5.

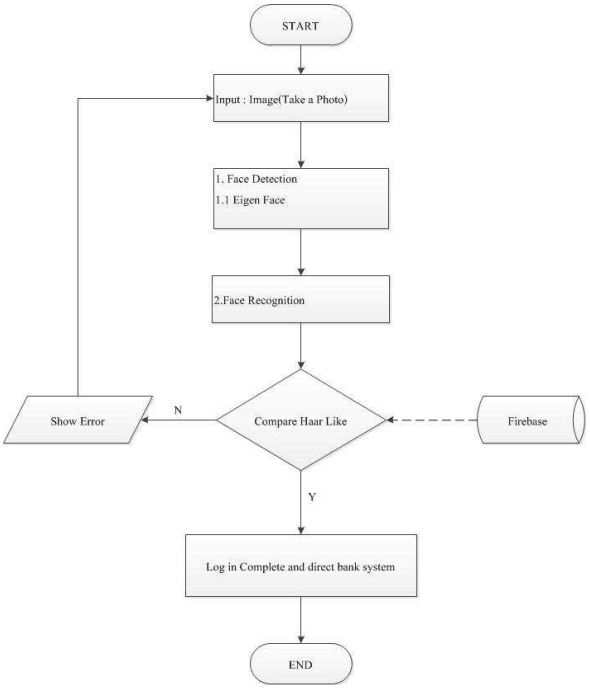


Fig. 5. Test Functioning.

- Input: Receive an image from Mobile Application.

Fig. 6. Receiving an image.

- Face detection: The system uses Eigen Face which is similar to what the trade uses.
- Face Recognition consists of 2 steps.
    - 1) Identify vector identify Ω from Γ input image.
    - 2) Identify differences between Ω and vector identities of the 6 images in the database. The vector identities in the database with least differences compared with Ω can be the same face.
    - Vector identity of input can be calculated as follows.
        - Identiy can be calculated by:

$$\Phi = \Gamma - \Psi \quad (5)$$

- Weight value can be calculate by:

$$w_i = u_i^T \Phi \quad (6)$$

- Weight values of K form a vector identiy of input

$$\Omega_J = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} \quad (7)$$

When vector identity of input is identified, differences between input and vector in the database is are calculated through the following equation.

$$S_J = \|\Omega - \Omega_J\|_2 = \sqrt{\sum_{i=1}^{K} (w_i - w_{Ji})^2} \quad (8)$$

Through the equation, the differences between indentities in the database are identified. The least SJ identity is selected for verification.

- Compare Haar Like: Input received is not face images recorded in the database. Therefore, images are screened out, using Threshold values. If the least different value is less than Threshold value, that face is the face of the test in the database.

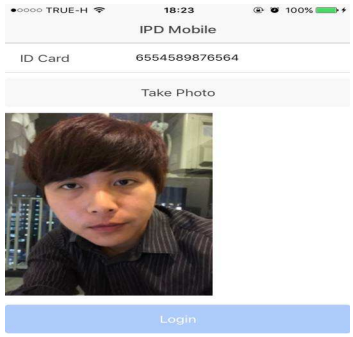


Fig. 7. Face dentity detection on Haar Like in Firebase.

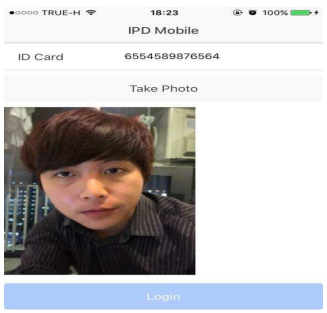


Fig. 8. The system compares face identity with the data on Haar Like. The system screens Threshold value. The least value or the value in Firebase range is correct.

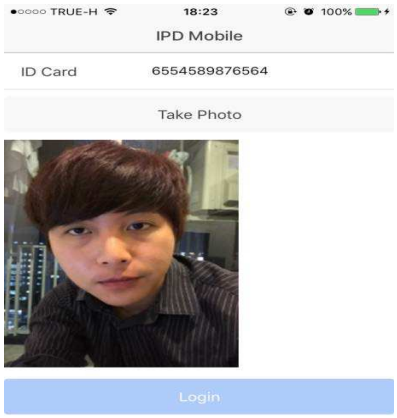


Fig. 9. The system is loading the dat on Firebase.

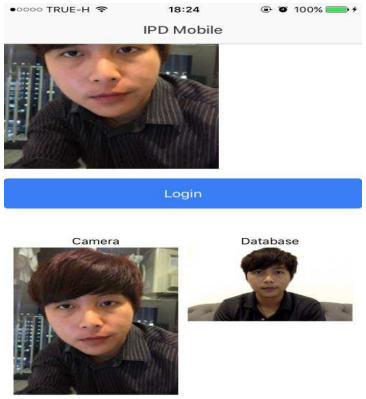


Fig. 10. Data derived from Compare Haar Like process.

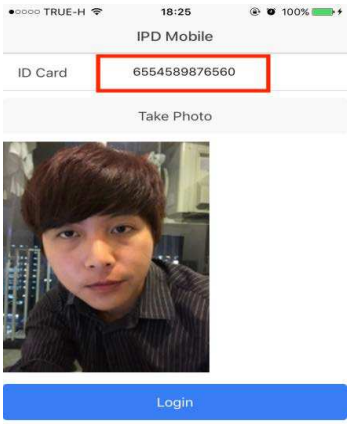


Fig. 11. Data not derived from Compare Haar Like process.

IV. Experimental Results

The system was tested in both trade and test for 10 times. The test details are as follow.

A. *Experimental Results of Trade*

Face detection test: 6 images were captured for 10 times so it included 60 images in total. However, only 45 images could be processed. These images included straight face with clearly-seen eyes, nose and mouth. Another 15 images could not be used. These images included those with sufficient illumination, inaccurate details and objects blocking faces. Also, the images were captured at too-high or too-low angles. The experimental results showed that 85% of images passed the face detection process.

B. *Experimental Results of Test*

Face recognition test: 10 tests were carried out and yielded the same Confidence values as Haar Like did for 6 times. Images that straight face, eyes, nose, mouth were clearly detected gave Confidence values. Four images with Confident values did not match Haar Like included images that eyes, nose and mouth could not be captured. When proceeding recognition, the values to be compared were inaccurate and were not in the range of Haar Like on Firebase. The experimental results showed that 85% of images passed the face recognition whereas 15% of the images failed to pass the test.

V. *CONCLUSION*

This research aimed to create a security system for processing financial transaction on a smart phone through recognition and face detection techniques. The researcher proposed Eigen Face and Haar Like methods in detecting face and recognizing face. The results yielded accurate results of face detection at 85% and of face recognition at 85%.

References

[1] Matheson, M. T. (2011). Taxing financial transactions: Issues and evidence (No. 11-54). International Monetary Fund.

[2] Lerner, J., & Schoar, A. (2005). Does legal enforcement affect financial transactions? The contractual channel in private equity. The Quarterly Journal of Economics, 120(1), 223-246.

[3] Chang, R., Ghoniem, M., Kosara, R., Ribarsky, W., Yang, J., Suma, E., ... & Sudjianto, A. (2007, October). Wirevis: Visualization of categorical, time-varying data from financial transactions. In 2007 IEEE Symposium on Visual Analytics Science and Technology (pp. 155-162). IEEE.

[4] Laurie, S., & Mortimer, K. (2019). How to achieve true integration: the impact of integrated marketing communication on the client/agency relationship. Journal of Marketing Management, 35(3-4), 231-252.

[5] Wang, X., Wang, Y., Zhu, X., & Luo, C. (2020). A novel chaotic algorithm for image encryption utilizing one-time pad based on pixel level and DNA level. Optics and Lasers in Engineering, 125, 105851.

[6] Li, G., Zhang, Z., Zhang, J., & Hu, A. (2020). Encrypting wireless communications on the fly using one-time pad and key generation. IEEE Internet of Things Journal.

[7] Argyris, A., Pikasis, E., & Syvridis, D. (2016). Gb/s one-time-pad data encryption with synchronized chaos-based true random bit generators. Journal of Lightwave Technology, 34(22), 5325-5331.

[8] Chen, F. L., Liu, W. F., Chen, S. G., & Wang, Z. H. (2018). Public-key quantum digital signature scheme with one-time pad private-key. Quantum Information Processing, 17(1), 10.

[9] W. Yimyam, T. Pinthong, N. Chumuang and M. Ketcham, "Face Detection Criminals through CCTV Cameras," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 351-357, doi: 10.1109/SITIS.2018.00061.

[10] Chumuang N., Ketcham M., Sawatnatee A. (2019) Criminal Background Check Program with Fingerprint. In: Theeramunkong T. et al. (eds) Advances in Intelligent Informatics, Smart Technology and Natural Language Processing. iSAI-NLP 2017. Advances in Intelligent Systems

and Computing, vol 807. Springer, Cham. https://doi.org/10.1007/978-3-319-94703-7_16

[11] N. Chumuang and M. Ketcham, "Intelligent handwriting Thai Signature Recognition System based on artificial neuron network," TENCON 2014 - 2014 IEEE Region 10 Conference, Bangkok, 2014, pp. 1-6, doi: 10.1109/TENCON.2014.7022415.

[12] Q. Z. Trepte et al., "Global Cloud Detection for CERES Edition 4 Using Terra and Aqua MODIS Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 9410-9449, Nov. 2019, doi: 10.1109/TGRS.2019.2926620.

[13] Yimyam W., Ketcham M. (2017) The Electroencephalography Signals Using Artificial Neural Network for Monitoring Fatigue System. In: Numao M., Theeramunkong T., Supnithi T., Ketcham M., Hnoohom N., Pramkeaw P. (eds) Trends in Artificial Intelligence: PRICAI 2016 Workshops. PRICAI 2016. Lecture Notes in Computer Science, vol 10004. Springer, Cham. https://doi.org/10.1007/978-3-319-60675-0_14

[14] W. LU and M. YANG, "Face Detection Based on Viola-Jones Algorithm Applying Composite Features," 2019 International Conference on Robots & Intelligent System (ICRIS), Haikou, China, 2019, pp. 82-85, doi: 10.1109/ICRIS.2019.00029.

[15] K. Candra Kirana, S. Wibawanto and H. Wahyu Herwanto, "Facial Emotion Recognition Based on Viola-Jones Algorithm in the Learning Environment," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, 2018, pp. 406-410, doi: 10.1109/ISEMANTIC.2018.8549735.

[16] Y. Xu, G. Yu, X. Wu, Y. Wang and Y. Ma, "An Enhanced Viola-Jones Vehicle Detection Method From Unmanned Aerial Vehicles Imagery," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1845-1856, July 2017, doi: 10.1109/TITS.2016.2617202.

[17] Y. Zhang, X. Xiao, L. Yang, Y. Xiang and S. Zhong, "Secure and Efficient Outsourcing of PCA-Based Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1683-1695, 2020, doi: 10.1109/TIFS.2019.2947872.

[18] M. B. S. Divya and N. B. Prajwala, "Facial Expression Recognition by Calculating Euclidian Distance for Eigen Faces Using PCA," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0244-0248, doi: 10.1109/ICCSP.2018.8524332.

[19] Y. Shatnawi, M. Alsmirat and M. Al-Ayyoub, "Face Recognition using Eigen-Faces and Extension Neural Network," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 2019, pp. 1-7, doi: 10.1109/AICCSA47632.2019.9035343.

[20] R. Ranjan et al., "A Fast and Accurate System for Face Detection, Identification, and Verification," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 1, no. 2, pp. 82-96, April 2019, doi: 10.1109/TBIOM.2019.2908436.

[21] Y. Kong, S. Liu, J. Wang and B. Zhang, "Person following based on Haar-like feature and HOG feature in indoor environment," 2016 35th Chinese Control Conference (CCC), Chengdu, 2016, pp. 6345-6349, doi: 10.1109/ChiCC.2016.7554354.

[22] J. Dai and N. Yan, "A Kernelized Correlation Filter Tracker with Random Sampling Haar-Like Feature," 2019 12th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2019, pp. 176-179, doi: 10.1109/ISCID.2019.10123.

[23] J. Ai, R. Tian, Q. Luo, J. Jin and B. Tang, "Multi-Scale Rotation-Invariant Haar-Like Feature Integrated CNN-Based Ship Detection Algorithm of Multiple-Target Environment in SAR Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 12, pp. 10070-10087, Dec. 2019, doi: 10.1109/TGRS.2019.2931308.

[24] S. Suwannakhun, N. Chumuang and M. Ketcham, "Identification and Retrieval System by Using Face Detection," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 294-298, doi: 10.1109/ISCIT.2018.8587856.

[25] E. Rekha and P. Ramaprasad, "An efficient automated attendance management system based on Eigen Face recognition," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, 2017, pp. 605-608, doi: 10.1109/CONFLUENCE.2017.7943223.

[26] Y. Shatnawi, M. Alsmirat and M. Al-Ayyoub, "Face Recognition using Eigen-Faces and Extension Neural Network," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 2019, pp. 1-7, doi: 10.1109/AICCSA47632.2019.9035343.

[27] L. Lei, S. Kim, W. Park, D. Kim and S. Ko, "Eigen directional bit-planes for robust face recognition," in IEEE Transactions on Consumer Electronics, vol. 60, no. 4, pp. 702-709, Nov. 2014, doi: 10.1109/TCE.2014.7027346.

[28] Yehong Liao and Xueyin Lin, "Blind image restoration with eigen-face subspace," in IEEE Transactions on Image Processing, vol. 14, no. 11, pp. 1766-1772, Nov. 2005, doi: 10.1109/TIP.2005.857274.

[29] N. Chumuang, M. Ketcham and T. Yingthawornsuk, "CCTV based surveillance system for railway station security," 2018 International Conference on Digital Arts, Media and Technology (ICDAMT), Phayao, 2018, pp. 7-12, doi: 10.1109/ICDAMT.2018.8376486.

[30] A. Das, "Face recognition in reduced Eigen-plane," 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS), Kolkata, 2012, pp. 620-623, doi: 10.1109/CODIS.2012.6422279.

A Machine Learning Approach for the Classification of Methamphetamine Dealers on Twitter in Thailand.

Punnavich Khowrurk
Department of Computer science
Faculty of Science and Technology, Thammasat University
Bangkok, Thailand
punnavich.k@sci.tu.ac.th

Rachada Kongkachandra
Department of Computer science
Faculty of Science and Technology, Thammasat University
Bangkok, Thailand
krachada@staff.tu.ac.th

***Abstract*— This research presents a method to classify messages from Twitter (tweet) related to Methamphetamine. The messages are classified into three classes: normal, seller, buyer. The models presented in this research are Multinomial Naive Bayes, Multi-Class LSTM, and Hierarchical LSTM. Model training uses a balanced and imbalanced dataset. The text used for Model training is tokenized from four tokenizers: Tlex+ , Lexto+, Attacut, and Deepcut. To study the model performance's effect, we divide the data with a different dataset and tokenizer. The results showed that all models could classify the messages into the three classes. The most effective model built from a balanced dataset is the Hierarchical LSTM model using the Lexto+ Tokenizer provides the highest Accuracy, and the most effective model build from an imbalanced dataset is the Multi-Class LSTM model using the Lexto+ Tokenizer. This model gave the highest Accuracy, but the F1-Score of the Hierarchical LSTM model gave better Accuracy in each class.**

The creation of a text classification model related to Methamphetamine uses Twitter messages. Most of them are Thai grammatical errors and has many slang usage. We found that Lexto+ is the best tokenizer to build a model. However, it is not much different from other tokenizers. On the other hand, the best dataset to build the model is a balanced dataset that significantly affects model performance.

Keywords— text classification, Methamphetamine, LSTM, Natural Language Processing, twitter classification

I. INTRODUCTION

The use of Twitter in Thailand is widespread today. Usage is reported to have grown 35% in 2019 compared to the previous year. (This is the highest growth rate in Southeast Asia), but behind this growth, some businesses are using Twitter to expand their marketing into selling illegal items. That is the drug trade.

Surveillance reports found that more than 200 Twitter users sold drugs such as Methamphetamine, marijuana, kratom, and Tramadol. Most of these sales use various slang or hashtags, which symbolize the drug trade group and are mostly related to its name, its use, and addiction symptoms.

The widespread use of social media provides an opportunity for merchant groups to use as a communication channel to facilitate drug dealing. According to a report by the Office of the Narcotics Control Board (ONCB) of Thailand, this has led to increased drug dealing rates through social media. The trading rate of Methamphetamine is higher compared to the past. Moreover, there are also more online trading channels, such as Twitter.

Currently, Thai tweets related to methamphetamine trading are detected by searching for relevant keywords directly on Twitter's website or app. This method requires a large number of detection personnel to achieve its goal.

This research aims to classify which Thai tweets are likely to be sellers, buyers, or normal tweets. By using tweets that contain keywords or slang that are related to Methamphetamine in Thai, such as "ไฮ", "ไอซ", "บิน" or "งานถือน", etc. Twitter Streaming API will automatically crawl all tweets used for research. All tweets labeled will be tokenized by 4 Tokenizers (Tlex+, Lexto+, Attacut, and Deepcut) to create classifiers.

Classifiers are created with three techniques: Multinomial Naive Bayes, Multi-class LSTM, and Hierarchical LSTM, created with balanced and unbalanced datasets and four different Tokenizers. There will be 24 classifiers to compare and conclude in sections 4 and 5, respectively.

II. RELATED WORKS

Currently, the drug epidemic has been found on social media, Twitter. It is used as a channel through which sellers and drug users make it easy to access drugs[3]. There are various tools used to find drug-related information on Twitter, including

One of the most straightforward tools to find information is keyword [4], which helps distinguish Tweet messages from other groups. However, Classifying a group of sentences by keyword is not good at classifying the text.

The researcher then chose to use the Naive Bayes Classifier, which classifies sentences into multiple classes [5]. Moreover, the LSTM tool is suitable for sentiment analysis and can classify messages [6]. As a result, the text can be classified more accurately.

S.Yuenyong et al. (2018) studied the classification of illegal messages, including Pornography, sex toys, Prostitute, gambling, and user drugs in the Thai language by Deep learning. The accuracy of the drug message was 89% [7].

Therefore, it inspired the researcher to study the classification of methamphetamine trading further. Which will be experimented with in this research.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

III. Proposed Methods

A. Keyword selection for searching sentences from twitter

A study was conducted to create a classifier to classify methamphetamine dealer and buyer in the Thai context from the Twitter API by using the keyword that related to the methamphetamine in Thai language. Consisting of these words

TABLE I. METHAMPHETAMINE KEYWORD

no	Methamphetamine keyword		
	word	*IPA*	*meaning*
1	ไอซ์	ʔaj	Nickname of Methamphetamine
2	เก็ต	kèt	Nickname of Methamphetamine
3	น้ำแข็ง	nám kʰěŋ	Nickname of Methamphetamine
4	ไฮคูล	haj kʰuːn	Nickname of Methamphetamine
5	งานก□อน	ŋaːn kɔ̂ːn	Unit of measure
6	จี	tɕiː	Unit of measure
7	ไฮ	haj	symptoms of drug addiction
8	ลอย	lɔːj	symptoms of drug addiction
9	บิน	Bin	symptoms of drug addiction
10	ดึง	dɯŋ	How to take drugs

The researcher was concerned that a classifier built on an imbalanced dataset would cause poor classifier performance, so the researcher decided to divide the data into two types: imbalanced dataset and balanced dataset (using the Under-sampling handles balanced dataset), which includes.

TABLE II. CHARACTERISTICS OF THE INFORMATION

Class label	characteristics of the information		
	class	*imbalanced total*	*balanced total*
0	normal	2491	416
1	seller	476	416
2	buyer	416	416

B. Data Cleaning

Sentence cleaning before tokenization to get the cleanest text, the researcher needs to eliminate the unwanted parts of the sentence using regular expressions to omit the text or special symbols, as shown in Table III.

However, in this research, the hashtag # was not taken for special significance, but the words in #{word} are used for processing as well as other words in the sentence.

TABLE III. UNWANTED PARTS OF THE SENTENCE

no	unwanted parts of the sentence	
	unwanted parts	*description*
1	RT	Word RT At the beginning
2	@{username}	@ Sign and following username
3	…	Three dot sign after all sentences
4	555*	555 more than 3 times represent the laugh
5	ๆ	Re-reading symbols
6	“(“ and “)”	Open and close parenthesis
7	.	The full stop
8	,	comma
9	-	dash
10	;	semicolon
11	“”	Quotation marks
12	http://* and https://	The URLS
13	/	slash
14	=	Equal sign
15	Space more than 1	Space including tab and other

C. Tokenizations

The tokenizing process for pre-processing in this research, using 4 Tokenizers consisted of:

1.Thai language tokenizer system (LexTo+) Tokenize with dictionary-based technique using Longest matching algorithm.

2.Thai language tokenizer system (TLex+) Tokenize with machine learning techniques using Conditional Random Fields algorithm.

3.AttaCut-Fast and Reasonably Accurate Word Tokenizer for Thai by Pattarawat Chormai

4.Deepcut is a Thai word segmentation library using Deep Neural, specifically, 1D Convolution Neural Network.

The tokens received from the tokenizer are arranged sequentially separated by spaces and then stored in the database in the fields as "sentence_lexto" ," sentence_tlex",," sentence_attacut" and ,"sentence_deepcut" for further processing.

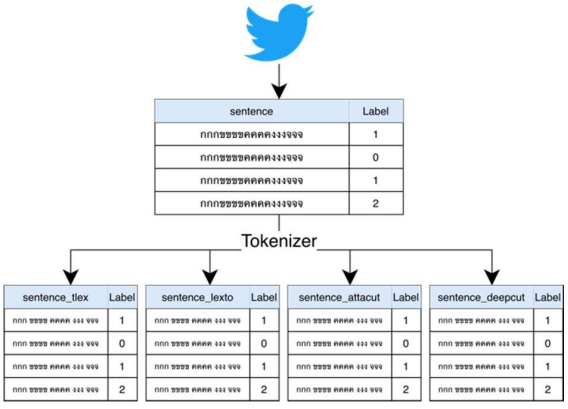


Fig. 1. Procedures for tokenizing messages with four types of Tokenizer.

D. Modeling for classification

The classifiers' creation with messages from the Corpus has been cleaned up, and those messages are divided into four groups based on their Tokenizer algorithm. Every classifier creation process requires that the training set and testing set data be used 75:25 of the total.

In this research, a classification model consisting of three main methods, which use different tokenizers and different datasets. (balanced and imbalanced dataset).

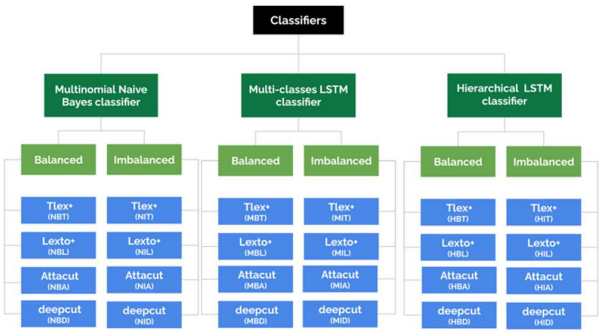


Fig. 2. All models used in the experiment.

1) Naive Bayes Classifier

The Naive Bayes Classifier is one of the popular methods of text classification that can classify multiple classes of text [5]. It is one of the probability equations. Creating a classifier is done by preparing the Naive BAYES network with a hand tool called Scikit-learn and then taking the sentence from the database, including: "sentence_lexto", "sentence_tlex", "sentence_attacut", and "sentence_deepcut" to create a classifier from both the balanced and imbalanced dataset. This classifier can predict all three classes: normal, seller, and buyer.

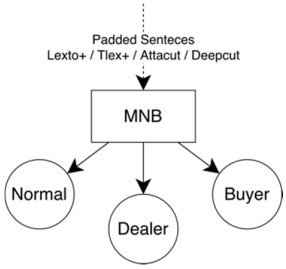


Fig. 3. The structure of multinomial Naive bayes classifier.

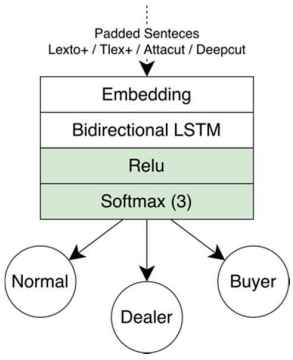


Fig. 4. The structure of multi-classes LSTM classifier.

2) Multi-classes LSTM classifier

This method is easy to provide because it creates only one network classifier to identify three classes.[6] However, in the sentence used to train about Methamphetamine, both buyers and sellers tend to be very similar. Also, the sentence in class 0 (Normal class) are considered together in the network. Therefore, the three-class prediction results were not accurate enough to predict buyers and sellers. Consequently, the researcher was inspired by [7] to separate the classifier's level to The third model that is a Hierarchical LSTM classifier

3) Hierarchical LSTM classifier

The Hierarchical LSTM classifier is made up of two models. The first model classifies text sentences that are related or unrelated to methamphetamine. The prediction results from model 1, which were related to Mmethamphetamine, were used to classify buyers or sellers using Model 2, respectively. By separating the hierarchy of unrelated text (class0), the second level classifier can better predict the sentences related to the buyer or seller.

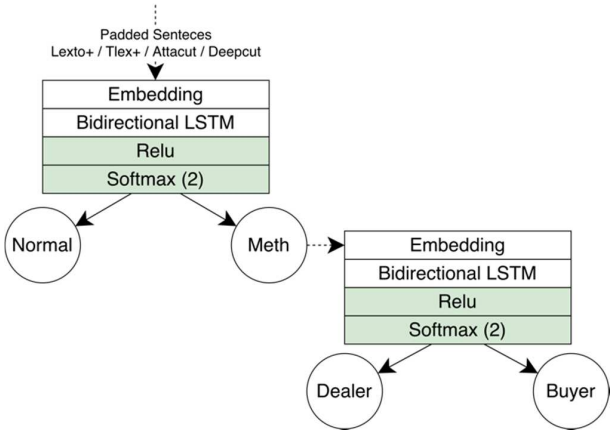


Fig. 5. The structure of 2-levels LSTM classifier.

IV. Experiment Result

A. Experiment Setting

This experiment used sentences from Twitter that matched the keyword from the table. 1, and then labeled those sentences into three classes, class 0 was the normal sentence, class 1 was the seller sentence, and class 2 was the buyer sentence. according to the numbers as shown in the table. 2

Subsequently, the experiment was divided into two parts: Experimenting with the balanced and imbalanced dataset, 75% used for model training, and 25% for testing.

An experiment with a balanced dataset extracted from a corpus using the Under-sampling handles balanced dataset (416 messages)

Both balanced and imbalanced data selected as model training data are passed through four tokenizer types: the first is the Thai tokenizer system (LexTo+) tokenize with the dictionary-based technique using the Longest matching algorithm. Second Thai language tokenizer system (TLex+) Tokenize with machine learning techniques using the Conditional Random Fields algorithm. Third AttaCut tokenizer and DeepCut tokenizer.

Tokens from 4 tokenizer use for created three models, multinomial Naive Bayes Classifier, Multi-Classes LSTM classifier, and Hierarchical LSTM classifier. Finally, 24 models are obtained, as shown in figure 2.

B. Evaluation Measurement

The model performance is measured from the accuracy to compare each model's overall performance. Each class's F1_score is measured to compare the performance in the Normal class, buyer class and seller class of each model.

C. Results

After building and testing the model, the researcher presented the results in two parts based on the data balance characteristics used to create the model: balanced dataset and imbalanced dataset, showing performance based on the accuracy and F1_score of each model, which uses a different tokenizer. The performance details are shown in Tables IV and V. The details are as follows.

1) Result of the Balanced Dataset

When a balanced dataset was used to create three models and measure their performance with Accuracy and measure the performance of each prediction class with F1-Score, the model with the highest Accuracy for Balanced Dataset was Hierarchical LSTM using a Lexto+ Tokenizer with Accuracy 0.94 is show in table IV.

Therefore, the Hierarchical LSTM model using the Lexto+ Tokenizer provides the highest Accuracy in the Balanced dataset.

TABLE IV. PERFORMANCE OF EACH BASED ON VARIATIONS OF TOKENIZER WITH BALANCED DATASET.

	Performance of each based on variations of tokenizer with balanced dataset.									
	Performance Measure	**Multinomial Naive Bayes**			**Multi-class LSTM**			**Hierarchical LSTM**		
		Normal	Seller	Buyer	Normal	Seller	Buyer	Normal	Seller	Buyer
Tlex+	Accuracy	0.84	0.84	0.84	0.76	0.76	0.76	0.93	0.93	0.93
	Precision	0.92	0.83	0.80	0.77	0.90	0.64	0.87	0.99	0.99
	Recall	0.81	0.93	0.80	0.83	0.71	0.73	0.84	0.99	0.99
	F1	0.86	0.88	0.80	0.80	0.80	0.68	0.85	0.99	0.99
	Time	0.019s			330.252s			1,758.750s		
Lexto+	Accuracy	**0.87**	**0.87**	**0.87**	**0.76**	**0.76**	**0.76**	**0.94**	**0.94**	**0.94**
	Precision	0.97	0.83	0.82	0.78	0.91	0.65	0.87	0.99	0.99
	Recall	0.86	0.94	0.81	0.86	0.69	0.74	0.91	0.99	0.99
	F1	0.86	0.94	0.81	0.82	0.79	0.69	0.89	0.99	0.99
	Time	0.019s			340.781s			1,980.604s		
Attacut	Accuracy	0.83	0.83	0.83	0.76	0.76	0.76	0.93	0.93	0.93
	Precision	0.93	0.82	0.74	0.76	0.91	0.65	0.87	0.99	0.99
	Recall	0.80	0.90	0.78	0.81	0.76	0.71	0.88	0.99	0.99
	F1	0.86	0.86	0.76	0.78	0.83	0.68	0.87	0.99	0.99
	Time	0.020s			328.791s			1,900.984s		
Deepcut	Accuracy	0.85	0.85	0.85	0.74	0.74	0.74	0.93	0.93	0.93
	Precision	0.92	0.73	0.88	0.76	0.92	0.62	0.87	0.99	0.99
	Recall	0.78	0.95	0.81	0.78	0.70	0.75	0.89	0.99	0.99
	F1	0.85	0.89	0.81	0.78	0.80	0.68	0.88	0.99	0.99
	Time	0.020s			338.792s			2,037.081s		

TABLE V. PERFORMANCE OF EACH BASED ON VARIATIONS OF TOKENIZER WITH IMBALANCED DATASET.

	Performance of each based on variations of tokenizer with imbalanced dataset.									
	Performance Measure	**Multinomial Naive Bayes**			**Multi-class LSTM**			**Hierarchical LSTM**		
		Normal	Seller	Buyer	Normal	Seller	Buyer	Normal	Seller	Buyer
Tlex+	Accuracy	0.83	0.83	0.83	0.88	0.88	0.88	0.86	0.86	0.86
	Precision	0.92	0.73	0.88	0.94	0.90	0.58	0.93	0.81	0.77
	Recall	0.89	0.93	0.69	0.99	0.55	0.65	0.98	0.80	0.79
	F1	0.90	0.82	0.77	0.96	0.68	0.62	0.95	0.80	0.78
	Time	0.024s			992.718s			3,134.634s		
Lexto+	Accuracy	**0.86**	**0.86**	**0.86**	**0.89**	**0.89**	**0.89**	**0.87**	**0.87**	**0.87**
	Precision	0.95	0.76	0.87	0.94	0.82	0.63	0.94	0.91	0.72
	Recall	0.91	0.91	0.76	0.98	0.71	0.58	0.98	0.70	0.92
	F1	0.93	0.83	0.81	0.96	0.76	0.60	0.96	0.84	0.81
	Time	0.026s			884.088s			3,015.825s		
Attacut	Accuracy	0.82	0.82	0.82	0.88	0.88	0.88	0.86	0.86	0.86
	Precision	0.93	0.70	0.88	0.93	0.80	0.63	0.94	0.88	0.72
	Recall	0.87	0.94	0.67	0.98	0.62	0.60	0.97	0.69	0.89
	F1	0.90	0.80	0.76	0.95	0.70	0.61	0.95	0.77	0.79
	Time	0.023s			779.918s			3,325.945		
Deepcut	Accuracy	0.85	0.85	0.85	0.87	0.87	0.87	0.86	0.86	0.86
	Precision	0.93	0.75	0.88	0.94	0.83	0.52	0.93	0.89	0.72
	Recall	0.90	0.94	0.71	0.96	0.67	0.58	0.98	0.70	0.90
	F1	0.91	0.84	0.79	0.95	0.74	0.55	0.95	0.78	0.80
	Time	0.025s			908.318s			3,237.583s		

2) Results of imbalanced dataset

When the imbalanced dataset was used to create in t create three models. the model's performance was evaluated by Accuracy, and F1-Score evaluated each prediction class's performance. The model with the highest Accuracy for the Imbalanced Dataset was Multi-Class LSTM using a Lexto+ Tokenizer that gives an Accuracy of 0.89 is show in table V

However, when considering the F1-Score of the Multi-Class LSTM, which consisted of a normal class, seller class, buyer class, the results were 0.96, 0.76, 0.60, respectively, which less effective than the F1-Score of the Hierarchical LSTM model, which consisted of class normal, seller, buyer is 0.96, 0.84, 0.81, respectively.

Therefore, the Multi-Class LSTM model using the Lexto + Tokenizer gave the highest Accuracy, but the F1-Score of the Hierarchical LSTM model gave better Accuracy in each class.

V. CONCLUSION

This research created three models (Naive Bayes, Multi-Class LSTM, Hierarchical LSTM) for classifying Twitter messages related to Methamphetamine. We try to build models that can classify those Twitter messages into three classes: normal, seller, buyer. In the process of model building, we use different tokenizers and datasets. Furthermore, we compared the performance of each model, using Accuracy and F1-Score.

The study results showed that all models could classify the messages into the three above groups. The most effective model built from a balanced dataset is the Hierarchical LSTM model using the Lexto+ Tokenizer provides the highest Accuracy, and the most effective model build from an imbalanced dataset is the Multi-Class LSTM model using the Lexto+ Tokenizer. This model gave the highest Accuracy, but the F1-Score of the Hierarchical LSTM model gave better Accuracy in each class.

The creation of a text classification model related to Methamphetamine uses Twitter messages. Most of them are Thai grammatical errors and has many slang usage. We found that Lexto+ is the best tokenizer to build a model. However, it is not much different from other tokenizers. On the other hand, the best dataset to build model is a balanced dataset that significantly affects model performance.

FUTURE WORK

Future work will be on testing other drug trading platforms on Twitter and other platforms. To be an alternative tool for monitoring online drug purchases.

REFERENCES

[1] Ms.นกยูง. คนไทยใช Twitter เถิง! 2 ปี เดิบโตสุดในภูมิภาค กลายเป็น “Entertainment Hub” ครองใจวัยมิลเลนเนียล, Jan2019. [Online]. Available:https://www.marketingoops.com/mediaads/social-media/year-on-twitter-in-thailand/

[Accessed: 09-FEB- 2020]

[2] กรุงเทพธุรกิจ นักค้ายาเปิดตลาดทวิตเตอธิจัดโปรไฟไหม,Aug2019.[Online].

Available:https://www. bangkokbiznews.com /news/detail/843563. [Accessed: 10-FEB- 2020]

[3] Seaman, Iris & Giraud-Carrier, Christophe. (2016). Prevalence and Attitudes about Illicit and Prescription Drugs on Twitter. 14-17. 10.1109/ICHI.2016.98.

[4] Shah, Bhavya & Agarwal, Varun & Dubey, Utkarsh & Correia, Stevina. (2018). Twitter Analysis for Disaster Management. 1-4. 10.1109/ICCUBEA.2018.8697382.

[5] Xu, Shuo. “Bayesian Naïve Bayes Classifiers to Text Classification.”

Journal of Information Science 44, no. 1 (February 2018): 48–59.

[6] B. Chen, Q. Huang, Y. Chen, L. Cheng and R. Chen, "Deep Neural Networks for Multi-class Sentiment Classification," 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, United Kingdom, 2018, pp. 854-859, doi: 10.1109/HPCC/SmartCity/DSS.2018.00142.

[7] S. Yuenyong, N. Hnoohom, K. Wongpatikaseree and T. P. N. Ayutthaya, "Classification of Tweets Related to Illegal Activities in Thai Language," *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Pattaya, Thailand, 2018, pp. 1-6, doi: 10.1109/iSAI-NLP.2018.8692858.

Design and Construct Logistic Robot Using Camera to Detect Line Combine with Lidar Sensor (Mecanum Wheel)

Tanaporn Anurakwongsri
Mahidol University International Demonstration School
Nakornpathom, Thailand
inginganurakwongsri@gmail.com

Burapha Srichanya
King Mongkut's Institute Technology Ladkrabang International Demonstration School
Ladkrabang, Thailand
meteemaker@gmail.com

***Abstract*— The researcher has an interest with implementing robotics technology to solve problems using engineering methods to create a project that can be developed for use. Since the work of a family had a problem with transporting goods in a warehouse, the researcher had an idea to design and construct a logistic robot that uses a Mecanum wheel system which can move in many directions. The structure is designed using Computer Aid Design (CAD). A motion experiment was held while the structure had a 40kg of load to test the movement of the logistic robot. Next is an experiment using the pixy camera to detect a black line which is the route of the robot in three directions, straight line, right curve and left curve to find the discrepancy value of movement at the x and y axis. The experiment includes the robot with no load, 20kg of load and 40kg of load with the result as an increase of discrepancy value as the weight of load increases. At the distance of 2 meter during the left curve, the x axis has the most discrepancy at 0.31 meter and at y axis the distance of 2 meter during the right curve, the discrepancy was at 0.32 meter. Another experiment was to detect a blacking obstacle using a lidar sensor. The robot was able to stop before crashing into the obstacle at the test distance of 0.20 meter, 0.40 meter and 0.60 meter by having a small problem because the structure was blocking the signal of the sensor. Moving and adjusting the angle of the pixy camera to detect the line more appropriately will make the robot be able to move accordingly to the experiment.**

Keywords—Robot, Mecanum Wheel, Pixy Camera

I. INTRODUCTION

Developing technology in these days is accelerating in a highly manner. Most developments are able to be used and apply to real life situations, especially robotics technology which is concerned with the lifestyles of humans because robots make life more convenient. They help with human work by embellishing it with higher quality, less production time and reduce human labor. Based on the benefits of robotic technology that was mentioned, robots can be used to solve problems in the area of transportation. For example, using robot technology to control a vehicle to move on its own, using robot technology to transport equipment. In this way, the researcher agreed to study and develop a robot prototype that can be used for transporting goods. The robot can move in many directions, can use camera technology for processing and has a technology that protects the robot from damage. So, the researcher decided to develop a logistic robot that uses a proportional system of Mecanum wheels that has a pixy camera as the as the main control for movement. It will detect the line that is in the laid-out path by using a technique of processing images through detecting line module in the pixy camera. As for avoiding crashing, the researcher used a lidar sensor to protect the robot from bumping into obstacles. Lidar sensor is a sensor that spins 360 degrees and shooting out light beams to detect objects ahead. This project is to help learn to design and construct logistic robot by moving from detecting the line through the pixy camera and has a protective system from the lidar sensor. In the future, the robot will use ROS (robot operating system) to help evaluate so it can become an automatic robot. The benefit of this research is the application of knowledge of different fields that is laid out through the project and can be used to solve problems guided by the engineering design process illustrating that Thai children can create a fully developed innovation.

II. OBJECTIVE

A. Design and construct logistic robot (Mecanum Wheel)

B. To study the motion of the logistic robot by detecting line using pixy camera

C. To study the Lidar sensor to use as protection for the logistic robot.

III. METHOD

A. Design

Designing. The researcher divided the design into 3 parts including structural design, control design and control program design.

1) Structural Design

The structural design. The researcher divided the structure into 3 parts which are the upper structure, the central structure and the lower structure. As shown in figure 1

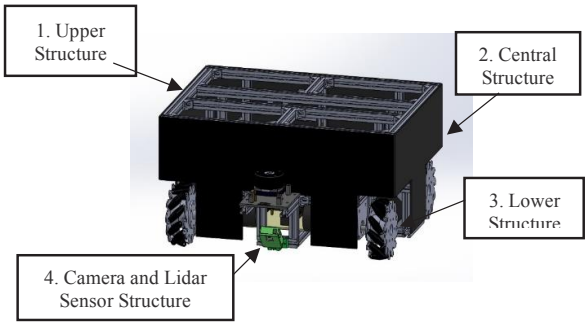


Fig. 1. Structure of the robot.

The upper structure has the purpose of supporting the weight of the equipment. The design uses 20x20 mm aluminum profiles to act like a support and column and is the structure and frame to protect the transportation robot that is made of acrylic. The central structure supports the base that is

made of an acrylic plate and it locates the control module, sensors and voltage level conversion module to be used as a power source for the control unit. The lower structure and frame supports another base with an acrylic sheet so that the motor driver, motors and batteries can be installed in a stable manner. As shown in figure 2.

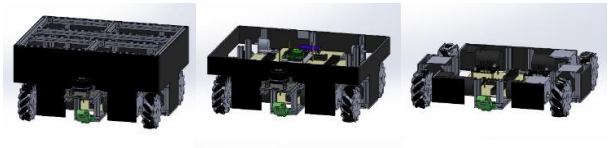


Fig. 2. Upper, central and lower structure of the robot.

The camera and lidar sensor structure is the structure that helps the logistic robot detect the line and be aware of its surroundings. The lidar sensor is placed on a base attached to an aluminum profile as support and is located above the pixy camera. The pixy camera has a cover that is made by 3D printing for protection. The cover is designed with a saddle joint at the edge so the pixy is able to rotate 90 degrees. The rotation is necessary for the purpose of the camera finding the line in order to have movement. As shown in figure 3.

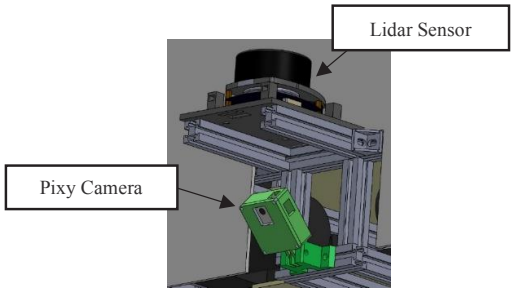


Fig. 3. Camera and Lidar Sensor Structure.

2) Control Design

The control design is divided into 2 sections: control design the proportional system of Mecanum wheels and control design of pixy camera and lidar sensor.

The researcher designed the control system that consists of Arduino Due microcontroller board as the main control board. The 4 encoder sensor receives the rotation signals of the 4 Mecanum wheels and commands the motor through the motor controller board, also known as, drive motor board. As shown in figure 4.

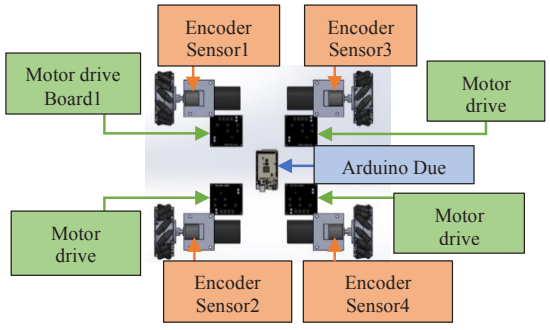


Fig. 4. Control system of the transportation robot.

To design the control system of the camera and lidar sensor, the researcher used a microcontroller board, Arduino Due as the main board. The lidar sensor sends the signal and Arduino Uno receives it. The signal is sent from Arduino Uno to the main board, Arduino Due. The signal from the lidar sensor is used as protection for the logistic robot to avoid crashing into obstacles ahead. The pixy camera helps the logistic robot move accordingly to the line. It detects if the line is in the middle or if the robot is off the course and sends the signal to Arduino Due. All the information stored in Arduino Due is used to determine when and how the logistic robot should move. See in figure 5.

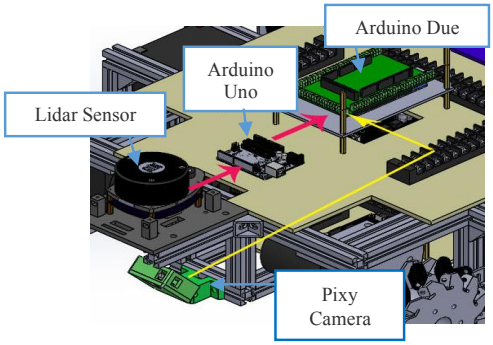


Fig. 5. Control design of the pixy camera and lidar sensor.

3) Control Program Design

Designing the control program. The researcher used Arduino IDE program that requires C programming language to instruct the drive motor to control the rotations of the gear motor and receiving the motor rotation value from the encoder sensor. To control the motors, the researcher used a pulse width modulation (PWM) signal to control the speed and direction. Programming interrupts are needed for receiving the values from the encoder sensors because it can keep up with the changes in signals from the values of encoder sensors.

B. Action

The researcher has applied the design of each part to create a workpiece to be the transportation robot as specified. The robot structure has the dimensions of Width x Length x Height, 440x580x280 millimeters. The main structure is assembled with parts of aluminum profiles. The base of each layer is made up from acrylic sheets. The Mecanum wheels of the propulsion system is 4 inches wide and is connect to a 12-volt DC gear motor. The exterior is covered with a black acrylic frame. The motor shaft has a conveyer belt sent to the encoder sensor above. In the front there is a lidar sensor that is located on top of the pixy camera. At the back of the robot, there are 2 switches and 2 LED that displays the voltage of the battery stuck to the acrylic frame. See figure 6

Fig. 6. Transportation robot with mecanum wheels before frame and after attaching a frame.

The control unit uses Arduino Due board as its main processor. It makes the motor rotate by instructing the drive motor using PWM signal and receives the rotation signals from the encoder sensors. The drive motor board, and Arduino

Due board obtains power from 2 separate 12 volt batteries. The first is applied to motor driver board. The second is imported into a voltage converter module for Arduino Due control board.

The lidar sensor in front is on top of a green base which is made from the 3D printer. It receives signals by spinning and shooting laser to see if there is an obstacle close by to protect the robot from crashing and then sends the signal to Arduino Due. The pixy camera is covered by a 3D printed frame and it detects the line that makes the robot moves. The signals is send to Arduino Uno first and then to Arduino Due to lighten up the evaluation process for Arduino Due. See figure 7.

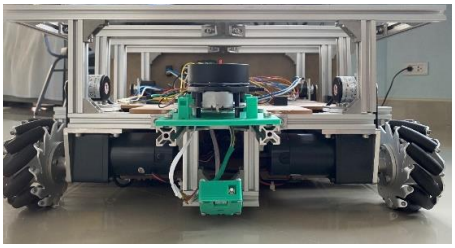


Fig. 7. Lidar sensor and Pixy camera.

IV. RESULT AND DISCUSSION

During the motion experiment of the logistic robot using Mecanum wheels, the researcher found that the robot can move with 20kg and 40kg of load without bending of the structure. The designed structure is able to perform according to plan. As for the Mecanum wheel system, the robot was able to move while carrying 20 and 40 kg of loads without problems. For the next experiment, the researcher tested the robot using a pixy camera to record the route of the robot's movement. The researcher divided the experiment using a black line into 3 cases: the logistic robot's motion in a straight line, right curve and left curve. In each case, the logistic robot has different load examples, including: no load experiment, 20 kilograms of load and 40 kilograms of load at the distance of 0.5 meter, 1.0 meter, 1.5 meter and 2.0 meter. To find the distance and movement the robot traveled on the x and y axis, as shown in figure 8 below.

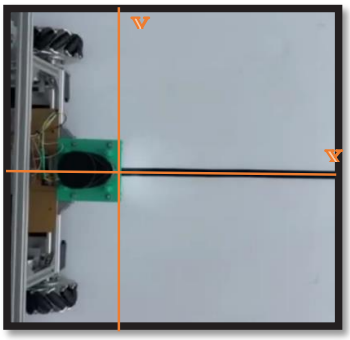
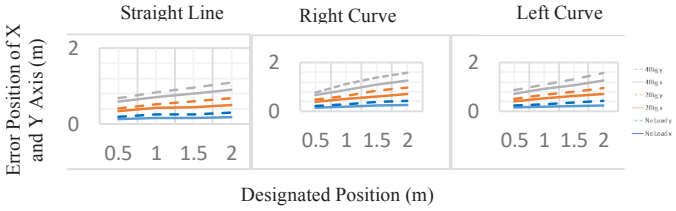


Fig. 8. Distance of the robot travelled on the x and y axis.

The testing of the logistic robot includes: testing with a straight line without load, with 20kg load and 40kg load, testing with a right and left curve movement without load, with 20kg of load and with 40kg of load. The experiments were measured in 0.5 meters, 1.0 meters, 1.5 meters and 2.0 meters of distance. The results of the movement of discrepancy on the x and y axis is shown in the graph 1 below



Graph. 1. Discrepancy of movement

Graph number 1 shows the movement in a straight line, right curve and left curve with no load, 20kg of load and 40 kg of load displayed by a x and y axis using a pixy camera.

The results of the movement of the logistic robot using a pixy camera to detect the route in the direction of a straight line, right curve and left curve and measuring the movement at x and y axis. When the robot carries no load, 20kg of load and 40kg of load at the distance of 0.5 meters, 1.0 meters, 1.5 meters and 2.0 meters, it is found that the average discrepancy of straight movement on the x axis is 8.83 percent and the average discrepancy of movement on the y axis is 6.04 percent. The average discrepancy of a left curve movement on the x axis is 11.71 percent and the average discrepancy of a left curve movement on the y axis is 8.50 percent. The average discrepancy of a right curve movement on the x axis is 11.25 percent and the average discrepancy of a right curve movement on the y axis is 9.00 percent. Based on the results on the logistic robot using a pixy camera to detect the route for movement, the discrepancy in the x and y axis will increase in proportion to the testing distance and load. The discrepancy of the x axis will be larger than the discrepancy of the y axis by at most 15.50 percent when the robot is carrying 40kg of load.

The next experiment is to detect object to protect the logistic robot from crashing by using Lidar sensor. The experiment included a white rectangle object the size of 0.20x0.20x0.40 as an obstacle located on the left side, right side and on top at the measured distance of 0.20 meter, 0.40 meter and 0.60 meter. If the sensor can detect the obstacle, the robot will stop moving. The results can be seen in the chart 1.

Distance of Obstacle (m)	Left Side		On Top Side		Right Side	
	Stop	*Doesn't Stop*	*Stop*	*Doesn't Stop*	*Stop*	*Doesn't Stop*
0.20	✓		✓		✓	
0.40	✓		✓		✓	
0.60	✓		✓			✓

Chart. 1. Result of Lidar sensor.

Based on the results, the robot can stop before running into the obstacle on the top and left side but on the right side, the robot couldn't detect the obstacle at the distance of 0.60 meter. The reason behind the error can be because the reflected signal of the lidar sensor is not enough for detection because the structure is blocking the signal of the sensor.

V. CONCLUSION

According to this research, it is found that the design of the structure still has problems. During the movement of the Mecanum wheeled robot, the structure vibrates when moving at a long distance. It is the reason why screws for the structure and motor gears are loosening. Thus, checking the structure

frequently is needed during the experiment. As for the control system, the power of the battery that is send to the drive motor decreases very fast resulting in no control of the drive motors' speed. The test of detecting line for movement using the pixy camera had a discrepancy movement at both the x and y axis. The x axis had a more value of discrepancy and the most with 40kg of load. The movement of the logistic robot using a pixy camera had a problem with detecting the line since the brightness of the light from the camera is not high enough to detect and the angle position causes the movement of the robot to not be able to move smoothly. In the obstacle detection experiment of the lidar sensor can be found that the detection on the right side at the distance of 0.60 meter had a problem because the structure is blocking the signal of the lidar sensor. Changing the position of the lidar can resolve the problem. For further development, the researcher will the change the system using control boards into ROS (Robot Operating System) so the system can be controlled at all times and with perfect accuracy. The processing of camera can also be developed into Depth Sensing which is suitable for use in the future.

REFERENCES

[1] Diogo Puppim de Oliveira, Wallace Pereira Neves dos Reis, OridesMorandin Junior. (2019 Sep). "A Qualitative Analysis of a USB Camera for AGV Control" [Online]. Available: https://www.mdpi.com/1424-8220/19/19/4111/htm. [Accessed: May. 4, 2020].

Development of a Web Service to Support the Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations

Chihiro Takada
Kochi University of Technology
Kochi, Japan

Yurika Takeuchi
Kochi University of Technology
Kochi, Japan

Mari Kinoshita
University of Kochi
Kochi, Japan
kinoshita@cc.u-kochi.ac.jp

Mikifumi Shikida
Kochi University of Technology
Kochi, Japan
shikida.mikifumi@kochi-tech.ac.jp

***Abstract*—Japan is prone to natural disasters, which pose a major threat to the safety and security of the country [1]. Shelters will be opened to protect evacuees from such disasters. A shelter is not only a place for physical protection, but also a place to provide a variety of services [2]. However, during the 2011 Great East Japan Earthquake and Tsunami, problems with the environment and management of shelters were identified. Many of the evacuees chose to evacuate from their homes, and issues were also raised regarding consideration for those people [1].**

Therefore, we have launched a novel information management system COACHES (Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations), which enables the efficient and quality provision care and relief to the all evacuees. The present study proposes a prototype of web service for the COACHES.

***Index Terms*—Disaster support, Evacuee support, Rapid healthcare assessment, Information sharing, Web service**

I. Introduction

Japan is located in the Circum-Pacific Mobile Belt where seismic and volcanic activities occur constantly. Although the country covers only 0.25% of the land area on the planet, the number of earthquakes and active volcanoes is quite high. In addition, because of geographical, topographical and meteorological conditions, the country is subject to frequent natural disasters such as typhoons, torrential rains and heavy snowfalls, as well as earthquakes and tsunami [1].

In 2011, more than 18,000 people died or went missing due to the Great East Japan Earthquake. There is also a high probability of the occurrence of large-scale earthquakes in the near future including impending possibilities of Nankai Trough Earthquake and Tokyo Inland Earthquake. As such, natural disasters remain a menacing threat to the safety and security of the country [1].

Shelters are established in the event of such a large scale disaster. Shelters can be defined as a displaced neighborhood of convenience, providing physical shelter as well as social and medical services [2]. In fact, records show that approximately 410,000 people in Iwate, Miyagi, and Fukushima prefectures, and 470,000 people in total across Japan, lived in shelters after the Great East Japan Earthquake [3].

However, in the Great East Japan Earthquake, there were many problems arising during the disaster: affected people suffered health problems; aged people were forced to stay home because they could not adapt themselves to the evacuation shelters in some cases, relief supplies were not provided sufficiently to home evacuees in many cases; and there were reported problems for provision of information, relief supplies, and services for widearea evacuees who evacuated to other prefectures or municipalities [1].

Mr. Fujita said that currently, many of the shelters have no privacy and often have poor sanitary conditions. Under such circumstances, the staff of the shelter need to provide medical and health care to the evacuees. In addition, when the damage is particularly severe and the central disaster hospitals and mainstay medical institutions in the affected area do not function, or are too busy, medical care in the evacuation centers becomes important for a certain period of time [4].

In addition, the Disaster Relief Law stipulates that evacuation shelters should provide food and water through cooking and other means. For this reason, many evacuation centers distribute relief supplies.

However, in the immediate aftermath of the disaster, because of the chaos on the ground, there was a situation in which the production capacity of the suppliers could not keep up with the supply of goods due to overlapping procurement, or because of the concentration of orders to specific suppliers [5].

In addition, many of the survivors have chosen to evacuate from their homes instead of using the evacuation centers. According to a survey conducted by the Cabinet Office on the victims of the Great East Japan Earthquake, about 70% of the respondents had not visited an evacuation center [6].

Evacuees in shelters have some level of support system in place, including medical personnel and relief supplies, which makes it easier for them to receive assistance. In contrast, those who choose to evacuate at home tend not to be able to receive adequate support [7].

In order to address these issues, the University of Kochi School of Nursing in collaboration with the Kochi University

The work was supported by JSPS KAKENHI Grant Number JP20K11132 and the University of Kochi Strategic Research Promotion Project (R1 Theme3) given to MK.

of technology has launched a novel information management system COACHES (Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations), which enables the efficient and quality provision care and relief to the all evacuees. The present study proposes a prototype of web service for the COACHES, which enables a secure and user-friendly data collection and reference in disaster settings.

COACHES collect individual healthcare data anonymously and record them on a cloud based database. The app runs on any personal mobile device (iPhone, iPad, Android) by scanning locally distributed unique personal identification code (QR code), which is the only item to identify individuals. The system requires a strict verification process to connect its database with any other personal information than a unique QR code for individuals.

We expect local healthcare professionals to run this system as volunteers and keep checking and recording people's health conditions. If this can be done at multiple places, COACHES will eventually cover the entire population and provide comprehensive data on the database to be shared among relief workers and the local governments.

II. Related Research

A related study is an ICT-based shelter management system [8]. This system uses ICT to collect information on evacuees, create and disseminate lists of evacuees including their relief needs, and manage the evacuees. An electronic triage system has also been proposed [9].

The system will use the evacuee's own NFC cards for personal identification, and if they don't have one, the shelter will distribute NFC wristbands provided by the shelter. However, under this system, when evacuees come to an evacuation center, staffs has to go through the procedure of asking the receptionist if they have an NFC card and handing out wristbands to those who don't have one. The procedure would not be so complicated if the number of evacuees in shelters had settled down, but immediately after a disaster occurs, the number of evacuees in shelters is large and the situation is chaotic, so there is a possibility that people may line up at the reception desk. In addition, personal information, such as name and address, is required, so consideration must be given to those who may not feel comfortable entering this information. Also, the results of the electronic triage are displayed by the color of the light on the wristband with LEDs, which is visible to other people and does not provide sufficient privacy considerations.

III. Community Oriented Approaches for Comprehensive Healthcare

The present study proposes a prototype of web service for the novel information management system COACHES (Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations), which enables a secure and user-friendly data collection and reference in disaster settings.

A. Flow of use

This system is intended to be used by evacuees who have been evacuated to shelters or evacuated at home, as well as healthcare professionals and relief workers who provide support.

QR codes for personal identification are routinely stored at evacuation centers, town halls, and disaster prevention warehouses, and are distributed to evacuees at evacuation centers and town halls in the event of a disaster. The QR codes are distributed on wristbands and are intended to be carried at all times.

Evacuees who get a QR code will record their health conditions and vital signs. A healthcare professional reads the person's QR code within the web application and records the results. The vital signs recorded at this time will be used for primary triage. Evacuees will also have their QR codes scanned when they receive meal service, bathe, receive relief supplies, and enter and exit the shelter, and the content and time of their visit will be recorded. Relief workers can scan the QR code in the web app to view information about the health condition of the evacuees and the support they have received. At this time, the personal information of the evacuees (name, address, etc.) is not included and is managed by IDs only, thus protecting their privacy.

The flow of use is illustrated in Fig. 1.

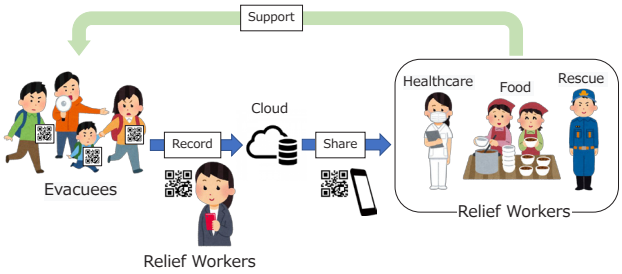


Fig. 1. Flow of use.

B. Features of the proposal methodology

The feature of the proposed method is that the usage procedure is very simple and rapid. As a concrete example, when a new evacuee arrives at a shelter, there is no need for user registration, etc.; all they have to do is hand over a QR code. Also, when handing out relief supplies, the staff need only scan the QR code of the evacuees with the terminal to hand them out smoothly. The system must be simple and rapid to use, like the proposed method, especially in the immediate aftermath of a disaster, because of the chaos in the field.

Another feature is the ability to make a comprehensive decision based on a lot of information. As a concrete example, when a healthcare professional examines a patient, he or she can view not only the health condition and vital signs records,

but also the records of eating and bathing, which can be used to make decisions based on living conditions. Therefore, we believe that a more accurate decision can be made.

C. Function overview

1) Evacuee's information recording and viewing function: In order to ensure efficient delivery of support, this feature records information such as support history and health conditions, which can be shared with other relief workers.

The information to be recorded is shown below.

- Health condition
  - Gender
  - Approximate age category
  - Physical status
  - Physical symptoms
- Vital signs
  - Body temperature
  - Pulse
  - Respiratory rate
  - Blood pressure (diastolic and systolic)
- Triage results
- Support history
  - Meal service (Time and place)
  - Receiving relief supplies (time, place, distribution group, category (foodstuffs, daily necessities, etc.), photos)
  - Bathing (Time and place)
  - Medical examination (Time, doctor in charge, remarks)
  - Legal/Welfare consultation (Time, responders, content (e.g., moving into a hypothetical home, inheritance issues, etc.)
- Shelter entry and exit records

A relief worker can access a evacuee's information by scanning the QR code that the evacuee has within the web application. Healthcare professionals can edit health conditions and vital signs. In addition, up to five of the most recent vital signs, support history, and shelter entry and exit records are available for viewing.

2) Triage support functions: It is a function that assists health care providers in making accurate decisions about health conditions, called triage.

From the recorded vital signs results, the triage results are displayed on the screen according to the algorithm. These results can be used to more accurately triage.

IV. Evaluation Experiment

A. System implementation

The current implementation situation of the system is shown below.

1) Health condition check screen: Fig. 2 is displayed when the QR code of the evacuee is loaded. This is a screen that displays the health condition of the evacuees. The information available is triage results, gender, approximate age category, physical status, and physical symptoms. Pressing the Vital signs button will move to Fig. 3, and pressing the Service Log button will move to Fig. 4.

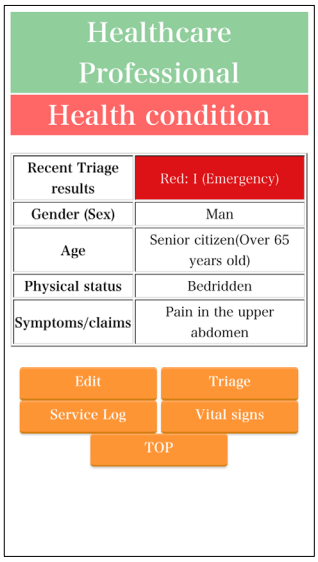


Fig. 2. Health condition.

2) Vital signs check screen: Press the Vital signs button on Fig. 2 to display Fig. 3. This is a screen that displays the vital signs of the evacuees. The table at the top of the screen shows the latest measurement results. If you scroll down the screen, you can see the results of the last five measurements in a graph.

3) Support history confirmation screen: Press the Service Log button on Fig. 2 to display Fig. 4. This is a screen that displays a history of the support received by the evacuees and a record of their entry and exit from the shelter. The support items are: meal service, relief supplies, bathing, shelter entry and exit, medical examinations, and legal and welfare consultations, which can be viewed by pressing the item you want to see, as shown in Fig. 5.

B. Survey method

In this study, we plan to demonstrate the system to people who have experienced the disaster and to medical workers, and to conduct an evaluation experiment using a questionnaire survey. The content of the questionnaire to be used in this process is shown in TABLE I and TABLE II. Choose from five options for your answer. The questionnaire will also include items to be implemented in the future.

C. Expected effects

In the questionnaire survey of this study, we believe that positive opinions about the proposed method can be obtained

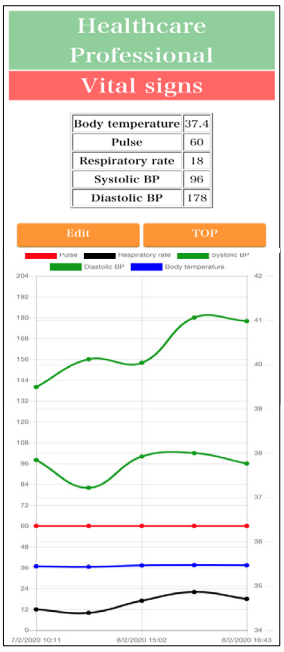


Fig. 3. Vital signs.

Fig. 4. Service Log 1.

Fig. 5. Service Log 2.

TABLE I
QUESTION ITEM (FOR PERSON WHO HAS EXPERIENCED A DISASTER)

Privacy Questions	
1	Do you think your personal privacy is being taken into consideration?
Information sharing Questions	
1	Would you be comfortable with relief workers (e.g., healthcare professionals, public administration staff, counselors, etc.) referring to the information you have entered for efficient support?
2	Would you prefer to have a limited relief workers who can see your information?
3	Do you want to set up relief workers who can see your information?
4	Do you want to set up your information that relief workers can see?
QR codes Questions	
1	Are you comfortable with wearing a QR code on your person at all times?
2	Would you be comfortable with needing to read a QR code in the following cases? · Meal service · Bathing · Medical examination · Entry and exit of the shelter · Verification of one's whereabouts
This proposed system Questions	
1	How satisfied are you overall with this method?
2	Would you be willing to use this method in a disaster?

TABLE II
QUESTION ITEM (FOR HEALTHCARE PROFESSIONALS)

Privacy Questions	
1	Do you think your personal privacy is being taken into consideration?
Information sharing Questions	
1	The information you enter is stored in the cloud and can be viewed at any time, do you think it useful?
2	Do you think it useful to be able to reference information from your own smart phone or tablet?
3	Do you think it helpful to have the following information available for reference? · Meal service · Bathing · Verification of one's whereabouts · Medical examination history · Vital signs · Local information · Latest medical information
QR codes Questions	
1	Are you comfortable with wearing a QR code on your person at all times?
2	If there is no QR code, you will have to search for the disaster victim's information by entering their names and IDs by yourself. Do you find reading a QR code useful?
This proposed system Questions	
1	How satisfied are you overall with this method?
2	Would you be willing to use this method in a disaster?

from both people who have experienced the disaster and medical professionals. Below is a list of the benefits of using this system, divided into two groups: evacuees and relief workers, including healthcare workers.

- Evacuees
  - A health professional can check the health condition of a evacuees simply by reading the QR code, allowing them to receive the correct treatment quickly.
  - Records of relief supplies will help distinguish those who received more supplies from those who received less, so that everyone can receive supplies equally.
  - Privacy is protected because the system identifies the individual by ID only, so you can use the system without registering your name, address or other personal information.
  - They do not need to have a terminal equipment to use the system, so the elderly and infants can also use the system.
- Relief workers, including healthcare workers
  - By simply scanning the QR code, the health conditions of the evacuees can be checked quickly.
  - Eating and bathing records as well as health conditions and vital signs can be used to determine overall health.
  - A record of the distribution of relief supplies is kept in a database, allowing us to manage the inventory of each supply. Based on that record, we can prevent shortages of supplies and excess inventory beyond what is needed by procuring the few remaining supplies and ceasing to procure the ones that are high.

Based on the above, we believe that the proposed method can make the activities of relief workers in disaster situations more efficient and effective. The better quality of support will also help the evacuees to live a more comfortable life.

V. CONCLUSION

In recent years, Japan has been hit by many natural disasters such as earthquakes and torrential rains, each of which has forced evacuees to evacuate to evacuation shelters or to stay at home. The most important aspect of support for the evacuees of a large-scale disaster is to maintain their health in the midst of their unstable living conditions. However, in the aftermath of the Great East Japan Earthquake in 2011, problems were identified in the environment and management of evacuation centers and issues related to consideration for those evacuated from their homes, and there are a number of issues that need to be improved in current support for the evacuees [1].

So, we have launched a novel information management system COACHES (Community Oriented Approaches for Comprehensive Healthcare in Emergency Situations), which enables the efficient and quality provision care and relief to the all evacuees. The present study proposes a prototype of web service for the COACHES, which enables a secure and user-friendly data collection and reference in disaster settings.

This system is designed to support relief workers in times of disaster through the use of a web application to provide efficient and high quality support to evacuees. In the future, we plan to conduct an evaluation experiment using a questionnaire survey, the details of which are also presented in this paper. We will continue to improve the web application in preparation for that evaluation experiment.

REFERENCES

[1] *Disaster Management in Japan*, Director General for Disaster Management Cabinet Office, Government of Japan, 2015.
[2] T. G. Veenema, A. B. Rains, M. Casey-Lockyer, J. Springer, and M. Kowal, "Quality of healthcare services provided in disaster shelters: An integrative literature review," *International Emergency Nursing*, vol. 23, no. 3, pp. 225–231, 2015.
[3] *Shelter Management Guidelines*, Director General for Disaster Management Cabinet Office, Government of Japan, 2016.
[4] T. Fujita, "Medical support (shelter)," *Journal of the Japanese Society of Internal Medicine*, vol. 99, no. 10, pp. 216–218, 2010.
[5] M. Kuwahara and K. Wada, "Recording and quantitative analysis of the flow of emergency relief supplies in the great east japan earthquake - analysis of the flow of emergency relief supplies handled by the national and prefectural governments," *Japan Transport and Tourism Research Institute*, vol. 16, no. 1, pp. 42–53, 2013.
[6] *Report on the Implementation of Comprehensive Measures for Evacuation*, Director General for Disaster Management Cabinet Office, Government of Japan, 2013.
[7] S. Sato, T. Maeno, M. Takai, Y. Owada, and M. Oguchi, "A development of an application to support distribution of relief supplies to victims in a large-scale disaster," *Proceedings of Multimedia, Distributed Collaboration and Mobile Symposium 2018*, pp. 716–720, 2018.
[8] K. Akasaka, R. Osone, Y. Amagi, T. Yamaguchi, and K. Abe, "Development and evalation of refuge management system using ict at the time of large scale disaster," *Transactions of the Information Processing Society of Japan, Consumer Devices and Systems*, vol. 7, no. 3, pp. 15–25, 2017.

[9] K. Akasaka, K. Kanisawa, T. Kanemaru, M. Isshiki, and K. Abe, "A proposal on refuge management system by electronic triage for persons requiring special assistance at large-scale disaster," *Proceedings of Multimedia, Distributed Collaboration and Mobile Symposium 2018*, pp. 277–282, 2018.

Feature Extraction with SHAP Value Analysis for Student Performance Evaluation in Remote Collaboration

Mako Komatsu
School of Information
Kochi University of Technology
Kochi, Japan
210317e@ugs.kochi-tech.ac.jp

Chihiro Takada
School of Information
Kochi University of Technology
Kochi, Japan

Chihiro Neshi
CS Project Department I
NRI System Techno, Ltd.
Kanagawa, Japan

Teruhiko Unoki
R&D center
Photron Limited
Tokyo, Japan

Mikifumi Shikida
School of Information
Kochi University of Technology
Kochi, Japan
shikida.mikifumi@kochi-tech.ac.jp

***Abstract*—In recent years, group discussions are becoming an important part of corporate recruitment examinations in Japan. Developing a remote teaching support system for group discussion will help reduce the burden of teachers. As a part of our project, this study aims to support teachers who need effective teaching method in remote group discussions by analyzing the video images. In this study, we used the features obtained from the videos. Students performances in group discussion were assessed automatically by classification, and important features were selected for teaching from the SHapley Additive exPlanations(SHAP) values.**

***Index Terms*—group discussion, remote teaching, video analysis, machine learning**

I. INTRODUCTION

Recently, group discussions are a part of corporate recruitment examinations in Japan [1]. Group discussions are important because they can measure abilities such as communication and aggressiveness that cannot be determined from documents and interviews.

At some universities, teachers of employment teach students in group discussions. If there is one discussion group, the teacher can teach sufficient advices. However, it may be difficult to teach all groups when multiple groups are discussing at the same time or when it is necessary to teach them remotely due to the effects of coronavirus.

For these reasons, we are focusing on group discussion and aim to realize a remote teaching support system for group discussion. In this research plan, we aims to develop a support system that automatically indicates to the teacher what kinds of advice are to be given to which group without using special devices or large equipment such as sensors. As a part of this project, in this study, we examine whether it is possible to estimate three grade levels: “Excellent”, “Average”, and “Failing”, using only the features obtained from the video images. In addition, we identify which features explain the most important factor for evaluating discussion participants.

II. RELATED WORK

Related studies to this project focus on group discussion and feature extraction. Aran constructs and evaluates a model that extracts features from participants’ nonverbal behaviors in group conversations to estimate personality [2]. This model uses information obtained from external devices, such as voice information, to successfully estimate with high accuracy. In a study similar to ours situation, Muralidhar analyzed the correlation between job interview ratings and nonverbal characteristics [3]. This study involved one-on-one interviews and used sensors and other devices for feature analysis. Zhang constructs a classification model for estimating participants’ roles in group discussions during job-hunting activities by using features such as language, speech turn, and speech information [4]. Naim provides a framework that automatically analyzes interview evaluations in job interviews using facial expression, audio and verbal information [5].

Thus, there are many studies related to group discussions, feature extraction, and classification estimation models. Many of them use sensors and other devices [2], [4], [6]. In our study, we do not need dedicated devices because we assume our model being used in an actual educational situation. We use a video of a group discussion that shows the upper body of the participants. We identify whether features obtained from the video images can be used to estimate the classification of the evaluation and what the importance of features.

III. EXPERIMENTAL METHOD

A. Overview

In this experiment, we analyze the number of features such as smiles and facial or direction obtained from videos during

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

group discussions. We then examine which features should be focused on when evaluating the participants. We also evaluate whether it is possible to detect participants who may need advices, such as people who do not speak up [7].

B. Capturing group discussions

The participants of this experiment consist of eight university students in their twenties, four males and four females. Two groups of four people, two males and two females, are created. Each participant has 30 minutes to discuss the theme. The task is to present single opinions in a group. The group discussion was captured by iPad. We thought that we could conduct a more natural group discussion because people are not likely to be aware that they are being captured on a tablet more than with a video camera. Figure 1 shows a diagram of the experiment. Figure 2 shows a part of the video taken in the experiment.

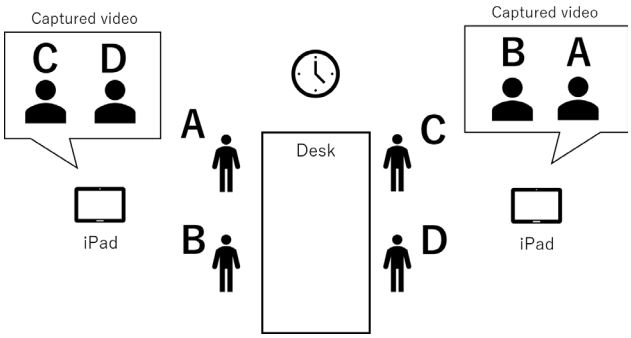


Fig. 1. State of experiment

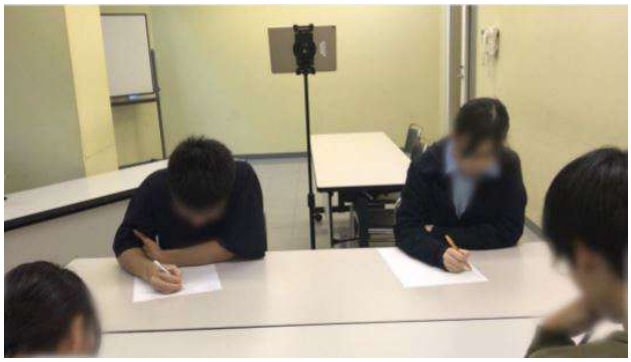


Fig. 2. Example video used

C. Features measurement and behaviors evaluation

A 30-minute video of the group discussion is divided into 10-second slots, and each slot is used to measure behaviors of the participants and evaluate them as participants. In this study, we classify the participants into "speaker" and "listener" for each slot, and measure and evaluate each feature. These processes are all performed manually.

In the measurement of features, we measure the features that can be found in the video images. The correspondence between X + numbers and the features are shown in Table I. Feature values are shown in Table II. This feature is an objective indicator that can be measured from video and audio. We set 12 features (X1∼X12) for speakers and 18 features (X1∼X10, X13∼X20) for listeners, such as "Degree of smile".

In the evaluation as a participant, each experimental participant is evaluated from the behaviors in Table III. This is an evaluation of communication skills such as cooperativeness and consideration for others from the teacher's point of view. The evaluations are based on 10 behaviors for speakers (Speaker 1∼10 in Table III) and 5 behaviors for listeners (Listener 1∼5 in Table III), such as "The participant make an effort to communicate to others". We evaluate one speaker and three listeners for each slot on a five-point scale, with 5 points for "Strongly Agree" and 1 point for "Strongly Disagree" in Table IV. A higher number indicates a higher rating.

TABLE I
FEATURES

Speaker	Listener	No.	Features
✓	✓	X1	Degree of smile
✓	✓	X2	Duration of smile
✓	✓	X3	Duration of putting on one elbow
✓	✓	X4	Duration of putting on both elbows
✓	✓	X5	Duration of touching your face and hair with your hands
✓	✓	X6	Duration of crossing your arms
✓	✓	X7	Duration of holding the pen in both hands
✓	✓	X8	Duration of spinning the pen
✓	✓	X9	Duration of moving hands
✓	✓	X10	The number of eye contact with listeners
✓		X11	Volume of speech
✓		X12	Duration of speaking
	✓	X13	Horizontal direction of the face (0 degree to the speaker)
	✓	X14	Vertical direction of the face (0 degree to the speaker)
	✓	X15	The number of nods
	✓	X16	The number of times you tilted your face
	✓	X17	Movement of the head when listening
	✓	X18	Duration of taking notes
	✓	X19	The number of times you raised your head and looked at the speaker while taking notes
	✓	X20	Duration of keeping your mouth closed

D. Data analysis method

In this study, we create a model that classifies the results of recruitment tests into three levels: "Excellent", "Average", and "Failing", based on the features obtained from the videos. First, we create a distribution map of the total values, divided into speakers and listeners. The total value is the sum of the scores of the feature ratings for each slot, 12 behaviors for speakers and 18 behaviors for listeners. We create a machine learning model that arbitrarily divides the distribution into three stages and makes classification predictions. We use XGboost to create a model with 75% of the total data as training data and 25% as testing data. We then use the SHAP values to evaluate the importance of the features. SHAP(SHapley

TABLE II
FEATURE VALUES

No.	The meaning of the numbers				
	1	*2*	*3*	*4*	*5*
X1	serious expression (not smile)	Both corners of the mouth bring up slightly	Both corners of the mouth bring up a little	Both corners of the mouth bring up	See the teeth
X2	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X3	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X4	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X5	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X6	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X7	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X8	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X9	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X10	0time	once	twice	3times	4times over
X11	cannot hear without attention	a little hard to hear	ordinary (about daily conversation)	a little louder	clear
X12	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X13	More or no speaker	± 90 degrees	± 60 degrees	30 ± degrees	0 degree
X14	± 60 degrees or no speaker		± 30 degrees		0 degree
X15	none	once	twice	3times	4times over
X16	none	once	twice	3times	4times over
X17	Not moving much	Moving a little	Moving	Moving a lot	Moving hardly
X18	0sec	0∼2sec	2∼4sec	4∼6sec	6sec over
X19	0time	once	twice	3times	4times over
X20	0∼2sec	2∼4sec	4∼6sec	6∼8sec	8sec over

TABLE III
EVALUATION BEHAVIORS FOR DISCUSSION PARTICIPANTS

	Behaviors
Speaker	1.The participant make an effort to communicate to others.
	2.The participant is logical.
	3.It is easy to understand what the participant is talking about.
	4.The participant is ignoring the flow of the discussion.
	5.The participant speaks something to propel the discussion.
	6.The participant contributes to the development of the discussion by asking for input from others.
	7.The participant is able to organize the various opinions collectively.
	8.The participant is able to revise the discussion.
	9.The participant is able to create a comfortable atmosphere to talk about, for example by making the place more relaxed.
	10.The participant is dismissing opinions without thinking.
Listener	1.The participant is trying to actively participate in the discussion.
	2.The participant is distracted.
	3.The participant is cooperative.
	4.The participant is asserting their own views.
	5.You feel good about the participant.

TABLE IV
THE EVALUATION METHODS IN TABLE III

	The meaning of the numbers
1	Strongly Disagree
2	Disagree
3	Undecided
4	Agree
5	Strongly Agree

Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model [8]. The features used in this study are not suitable for multivariate analysis because they are highly related to each other, such as the direction of the face, and therefore SHAP values are used.

IV. EXPERIMENTAL RESULTS

A. Classification model

The confusion matrix of the model created by XGboost is Table V and Table VI. The horizontal axis, A, B, and C for Predicted are the three categories predicted by the model, and the vertical axis, A, B, and C for Actual are the actual three categories. A is Excellent, B is Average, and C is Failing. Thus, taking Table V as an example, the value 3 at the location of Predicted A on the horizontal axis and A on the vertical axis is the number of correct answers that machine learning predicts to be A.

The classification reports are shown in Table VII and Table VIII. The accuracy was 0.71 for speakers and 0.65 for listeners, showing that the accuracy of the prediction was high.

B. Importance of features

The results of the analysis of the effect of each characteristic on the evaluation of the participants of the group discussion using the SHAP values are shown in Figures 3 and 4. The larger the value of the horizontal axis, the greater the impact on the evaluation, and the more important the feature is. The color of each graph shows which features were affected by each of the three levels of classification. When we look at the colors of the classifications C (for “Failing”) and A (for “Excellent”), there are no behaviors with many colors in either one of the two categories. There are many behaviors that are equal in both categories. Therefore, we can understand that there are no features that affect only good or bad evaluations.

TABLE V
CONFUSION MATRIX OF SPEAKERS

Actual \ Predicted	A	B	C
A	3	10	0
B	2	48	4
C	1	6	4

TABLE VI
CONFUSION MATRIX OF LISTENERS

Actual \ Predicted	A	B	C
A	62	40	1
B	30	62	6
C	1	19	60

V. DISCUSSION

A. Classification accuracy

The values of confusion matrix and accuracy in Table V and Table VI show that the classification model created in this study has high accuracy. This means that the model can classify data into three levels with the same degree of accuracy as when they are evaluated by hand. We think that there is a possibility to automate the evaluation of group discussions by using this model.

B. Speakers

Figure 3 shows that the X9 "Duration of moving hands" is the feature that has the greatest influence on the speaker's evaluation, and behaviors such as X2 "Duration of smile" and X1 "Degree of smile" also have an impact. The result of this study suggests that we can detect the characteristics of the students who are trying to convey their opinions by hand gestures or who are laughing at other members in the group discussion and let the teacher know this features.

On the other hand, the behaviors X7 "Duration of holding the pen in both hands", X4 "Duration of putting on both elbows", X6 "Duration of crossing your arms", and X8 "Duration of spinning the pen" had a very low impact on the evaluation. Therefore, it can be said that low impact behaviors do not need to be used to support teachers' remote instruction.

TABLE VII
CLASSIFICATION REPORT OF SPEAKER

	precision	recall	f1-score	support
A	0.50	0.36	0.42	11
B	0.75	0.89	0.81	54
C	0.50	0.23	0.32	13
accuracy			0.71	78
macro avg	0.58	0.49	0.52	78
weighted avg	0.67	0.71	0.68	78

TABLE VIII
CLASSIFICATION REPORT OF LISTENER

	precision	recall	f1-score	support
A	0.90	0.75	0.82	80
B	0.51	0.63	0.57	98
C	0.67	0.60	0.63	103
accuracy			0.65	281
macro avg	0.69	0.66	0.67	281
weighted avg	0.68	0.65	0.66	281

C. Listeners

Figure 4 shows that X1 "Degree of smile", X13 "Horizontal direction of the face", X2 "Duration of smile" are characteristics that significantly affect the ratings of listeners. From these results, we can say that characteristics such as whether or not the listeners turn toward the speaker and nods while listening to the speaker are important. Therefore, we think that informing the teacher of the characteristics of the attitudes of listeners regarding their listening attitudes can be useful as a support for evaluation.

On the other hand, X7 "Duration of holding the pen in both hands", X5 "Duration of touching your face and paper with your hands", X6 "Duration of crossing your arms" were found to be features with low influence on the evaluation. These characteristics are not considered necessary to support teachers in their instructional support.

D. Towords the realization of the support system

From this experiment, it was found that the feature values obtained from the video images could be classified into three levels of evaluation with high accuracy. In addition, we were able to identify characteristics that were highly influential in the evaluation of speakers and listeners. These evaluations and features will be useful in supporting the teacher's advice and evaluation by informing the teacher of these evaluations. These features can be automatically selected from the video images by using hand and face detection, voice recognition, etc. Thus, it is possible to realize automatic support by combining feature extraction and classification estimation models. In addition, we think that other feature behaviors should be considered. In this study, we captured the group discussion and examined whether it is possible to select and categorize the features that are useful for the evaluation of the participants using only the information obtained from the video. From the experimental results, it was shown that the feature values obtained from improve accuracy.

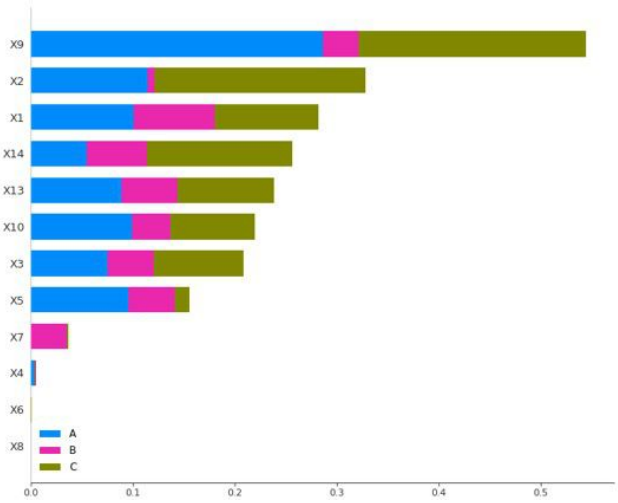


Fig. 3. SHAP values of speakers

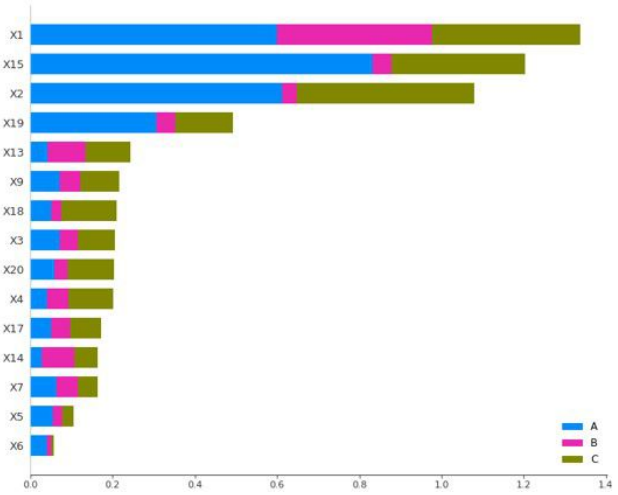


Fig. 4. SHAP values of listeners

VI. CONCLUSION

In this study, we captured the group discussion and examined whether it is possible to extract and categorize the features that are useful for the evaluation of the participants using only the information obtained from the video. From the experimental results, it was shown that the feature values obtained from the video images were estimated as accurately as they were manually and that the features with a high influence on the evaluation were identified. These results show that features obtained from video are useful in realizing a system that automatically supports teachers' instruction for group discussions.

In this paper, we have focused on group discussions among university students, but in the future, further studies are required with more data and in other situation.

REFERENCES

[1] Mynavi Corporation, "Mynavi student employment monitor survey - Activity in May-," 2019.

[2] O. Aran and D. Gatica-Perez, "One of a kind: Inferring personality impressions in meetings," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*. Association for Computing Machinery, 2013, pp. 11–18.

[3] S. Muralidhar, L. S. Nguyen, D. Frauendorfer, J.-M. Odobez, M. Schmid Mast, and D. Gatica-Perez, "Training on the job: Behavioral analysis of job interviews in hospitality," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, 2016, pp. 84–91.

[4] Q. Zhang, H.-H. Huang, S. Kimura, S. Okada, N. Ohta, and K. Kuwabara, "Analysis on the Participants' Functional Roles and their Transition in Group Discussion Conversation -Toward a Communication Skill Training System," vol. 20, no. 1. The Transactions of Human Interface Society, 2018, pp. 31–44.

[5] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated prediction and analysis of job interview performance: The role of what you say and how you say it," vol. 1. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015, pp. 1–6.

[6] M. Zancanaro, B. Lepri, and F. Pianesi, "Automatic detection of group functional roles in face to face interactions," in *Proceedings of the 8th International Conference on Multimodal Interfaces*. Association for Computing Machinery, 2006, pp. 28–34.

[7] C. Takada, C. Neshi, T. Unoki, and M. Shikida, "An attempt of Video Analysis for Remote Guidance Support of Group Discussion," vol. 34, no. 6. Japanese Society for Information and Systems in Education Reserch Report, 2020, pp. 235–239.

[8] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Statistical Machine Translation for Myanmar Language Paraphrase Generation

Myint Myint Htay
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
myintmyinthtay@utycc.edu.mm

Ye Kyaw Thu
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
Language Understanding Lab, Myanmar
yktnlp@gmail.com

Hnin Aye Thant
Department of Information Science
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
hninayethant@utycc.edu.mm

Thepchai Supnithi
National Electronics and Computer Technology Center (NECTEC)
Pathum Thani, Thailand
thepchai.supnithi@nectec.or.th

***Abstract*—In this paper, we applied a statistical machine translation (SMT) approach to generate Burmese paraphrases of input sentences and words in Burmese. The system trained 89K sentence pairs that are manually collected from Facebook Comments and daily conversation corpus and also 89K Burmese Paraphrase Words are collected from Burmese Wiktionary. We implemented three different statistical machine translation models; phrase-based, hierarchical phrase-based, and the operation sequence model. Moreover, we used two segmentation units; character and syllable segmentation for comparing the machine translation performance. The performance of machine translation or paraphrase generation was measured in terms of BLEU, RIBES, chrF^{++}, and WER scores for all experiments. However, automatic evaluation metrics are weak for judging whether the generated Burmese sentences and words "is a paraphrase" or "is not a paraphrase". And thus, we also conducted a human evaluation on both sentence-to-sentence and word-to-word paraphrase generation results. We found that the results obtained using the BLEU and RIBES automatic evaluation metrics were misleading and as the human evaluation result the machine translation approach is suitable for Burmese paraphrase generation.**

***Index Terms*—Paraphrase Generation, Burmese (Myanmar Language), PBSMT, HPBSMT, OSM, Human Evaluation**

I. Introduction

Paraphrasing is showing an input text in different ways but retaining its original meaning. Many Natural Language Generation (NLG) tasks can be viewed as generating paraphrases. Paraphrase generation is an important task in NLP, which is the main technology in many applications such as retrieval based question and answering system, semantic analyzing in NLP task, query expression in web searching, data increasing for dialogue system such as chat-bot system. However, due to the complexity of natural language and as well as one of the under-resourced languages Burmese (Myanmar Language), automatically generating accurate and different paraphrases is still very challenging.

In machine translation, phrases, words, or sentences in a source language are translated into corresponding phrases, words, or sentences in a target language. There are two types of machine translation methods statistical machine translation (SMT) and neural machine translation (NMT). Both SMT and NMT rely on parallel corpus for each sentence, word, and phrase to compare weights, semantic, word order, and probabilities of each pair of corpus translation. Generally, a large parallel corpus (source language together with translated target language) is required for machine translation experiments.

In this paper, machine translation is applied for generating Burmese paraphrase sentences and words. Although Burmese grammar is not very complex, no clear word boundary is challenging for the alignment process. Moreover, there is no publicly available Burmese word segmenter yet, and thus, we used character and syllable segmentation for all experiments.

II. Related Work

In recent years, various traditional techniques have been developed to solve the paraphrase generation problem. Manually defined rules are used in [1]. Thesaurus-based method for lexical substitution is used in [2]. Using WordNet for paraphrasing in [3]. And statistical machine translation (SMT) has also been used [4]. It trained SMT tools on a large number of sentence pairs collected from newspapers. Using latent-variable pcfgs for semantic parsing in [5]. Paraphrasing generation with automatic evaluation in [6]. Phrase-based SMT model trained on aligned news headlines proposed [7]. Multiple resources to strengthen a log-linear SMT model utilize in [8]. SMT translation between Burmese and other twenty languages from the multilingual Basic Travel Expressions Corpus (BTEC) in [9]. Recently, deep neural models have also been applied to paraphrase generation due to their great success in natural language processing tasks.

Our work focuses on three SMT techniques (PBSMT, HPBSMT, and OSM) for generating Burmese paraphrases

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

and we evaluated with three automatic evaluation metrics (BLEU, RIBES, chrF^{++}) and also conducted a human evaluation on our results for paraphrase generation.

III. Word Segmentation

Burmese sentences are composed as continuous sequences of syllables with no characters delimiting the words. In statistical machine translation (SMT), the word segmentation step is necessary for languages that do not naturally delimit words. In this research, we used syllable segmentation for Burmese sentences and character and syllable segmentation for Burmese words.

A. Character Segmentation

All of the Burmese words are composed of multiple syllables and most of the syllables are composed of more than one character. The collected words for this research are insufficient to cover the whole of the Burmese words. Thus, we also used the character segmentation for the Burmese words. We used character segmentation with the following tool.

(https://github.com/ye-kyaw-thu/tools/blob/master/bash/char-segmentation.sh)

Character breaking examples are such as (ကမ္ဘာလောက)("the world" in English) and is broken into (က မ ္ ဘ ာ လ ေ ာ က).

B. Syllable Segmentation

We used the algorithm of [10] for syllable breaking. There are three general rules to break Burmese syllables from Unicode input text where a consonant is followed by dependent vowels and other symbols. Firstly, put a word break in front of consonants, independent vowels, numbers, and symbol characters. Secondly, remove any word breaks that are in front of subscript consonants, Kinzi characters, and consonant + Asat characters.

If we only focus on consonant-based syllables, the structure of the syllable can be described with Backus normal form (BNF) as follows:

Syllable := CMV[CK][D]

Here, it defines the consonants as C, Medials as M, vowel as V, Killer character as K, and diacritic characters as D. Burmese syllable segmentation can be done with a rule-based approach, finite-state automaton (FSA), or regular expressions (RE).

(https://github.com/ye-kyaw-thu/sylbreak).

Syllable breaking is a very important first step for Burmese word segmentation since most Burmese words are written of multiple syllables and most of the syllables are written of more than one character. Moreover, syllable segmentation can solve the problem of the Out Of Vocabulary (OOV) in Burmese SMT translation.

Syllable breaking examples are such as (တော်ပါတယ်ကွာ)("So clever" in English) and is broken into (တော် ပါ တယ် ကွာ)and Pali Word (ဒါငါ့ဥစ္စာ)("This is mine." in English) is broken into (ဒါ ငါ့ ဥစ္စာ)it used orthographic segmentation.

IV. Building Burmese Paraphrase Corpus

In Burmese, we use various words and various dialogue styles for the same fact in daily conversation and writing sentences. Some of the paraphrase sentences are different only one word in that sentence. However, some are quite different for the whole sentence. Some of the sentences are collected from social media (Facebook comments). Comments are collected from the famous Myanmar news websites by extracting the Facepager Tool (version 4.2.7) [11]. And then, we choose the proper and paraphrase sentence for our work. Moreover, some of the sentences are manually constructed by using the paraphrase words from Burmese Wiktionary [12]. At first, we collected a paraphrase word list (89,446 word pairs), and based on that list we developed sentence level paraphrase corpus (89,036 sentence pairs). The Burmese paraphrase corpus is a UTF-8 plain text file and using "sentence<TAB>paraphrase_sentence" formats. The followings are some example of Burmese paraphrase words and sentences and the original sentence is denoted by "os", the paraphrase sentence is denoted by "ps", the original word is denoted by "ow" and the paraphrase word is denoted by "pw".

Some example of Burmese paraphrase words are as follows:

ow: ဖျတ် လတ် ("active" in English)
pw: သွက် လက်
pw: မြန် ဆန်

In the above example, all three different words have the same meaning.

Some Burmese paraphrase sentences are nearly the same but only using different vocabulary or phrases. Some examples are as follow:

os: သေ ချာ ဆုံး ဖြတ် ပြီး ပြီ လား ။ ("Do you make the decision definitely sure?" in English)
ps: အ သေ အ ချာ ဆုံး ဖြတ် ပြီး ပြီ လား ။
ps: ပြတ် ပြတ် သား သား ဆုံး ဖြတ် ပြီး ပြီ လား ။
ps: သေ သေ ချာ ချာ ဆုံး ဖြတ် ပြီး ပြီ လား ။

Here are some examples of paraphrasing the whole Burmese sentence:

os: နေ ကောင်း လား ။ ("How are you?" in English)
ps: ကျန်း မာ ရဲ့ လား ။
ps: နေ ကောင်း တယ် နော် ။
ps: မာ ရဲ့ လား ။

V. SMT Methodologies

In this section, we will describe the methodologies of statistical machine translation processes used in the experiments in this paper.

A. Phrase-based Statistical Machine Translation (PB-SMT)

Phrase-based translation systems work on the phrasal units [13]. A phrase-based translation model typically gives better performance than word-based models. A simple phrase-based translation model consists of phrase-pair probabilities that are extracted from the corpus and a basic reordering mode, and an algorithm to extract the phrases to build a phrase-table [14].

To find best translation $\hat{t}$ that maximizes the translation probability $\mathbf{P}(t|s)$ given the source sentences; mathematically. Here, the source language is Burmese Source Sentence (s) and the target language is Burmese Paraphrase Sentence (t). The translation of a Source sentence into a Paraphrase sentence is modeled as equation 1.

$$\hat{t} = argmax_t \mathbf{P}(t|s) = argmax_t \mathbf{P}(s|t)\mathbf{P}(t) \quad (1)$$

B. Hierarchical Phrase-based Machine Translation (HPB-SMT)

The hierarchical phrase-based SMT approach is a model based on synchronous context-free grammar (SCFG) [15]. The translation model is constructed by using SCFGs which can be learned from a corpus of unannotated parallel text. We can first extract initial phrase pairs and then obtain hierarchical phrase rules. An example of hierarchical phrase-based grammar rules between source Burmese sentence to target Paraphrase sentence for HPBSMT translation.

ချမ်း သာ ခြင်း [X] ||| ရွှင် ကြ် [X] |||
ချမ်း သာ ခြင်း [X] ||| ရွှင် လန်း ကြ် [X] |||
ချမ်း သာ ခြင်း [X][X][X] ||| ရွှင် ကြ် [X][X][X] |||
ချမ်း သာ ခြင်း [X][X][X] ||| ရွှင် လန်း ကြ် [X][X][X] |||
ချမ်း သာ ခြင်း ပါ [X] ||| ရွှင် ကြ် ပါ [X] |||
ချမ်း သာ ခြင်း ပါ [X] ||| ရွှင် လန်း ကြ် ပါ [X] |||
ချမ်း သာ ခြင်း ပါ ဇေ [X] ||| ရွှင် ကြ် ပါ ဇေ [X] |||
ချမ်း သာ ကြ် ပါ ဇေ [X] ||| ရွှင် လန်း ကြ် ပါ ဇေ [X] |||

That means ("Be Happy and Peaceful in your life" in English). These are the wishes for other people in Burmese conversation.

C. Operation Sequence Model (OSM)

The operation sequence model is a model for SMT that combines the benefits of two state-of-the-art SMT frameworks, namely n-gram-based SMT and phrase-based SMT [16]. It is a generative model that performs the translation process as a linear sequence of operations that jointly generate the source and target sentences. The operation types are such as generate (a sequence of source and target words), insert gap (reordering operations), jump (forward and backward jump which performs the actual reordering).

The following example shows generating paraphrase text with OSM translation from the original text.

Source: မင်း က အရမ်း တော် တယ်
Target: မင်း က အလွန် တော် ပါတယ်
("You are so clever." in English)
Operation 1: Generate (မင်း, မင်း)
Operation 2: Insert Gap
Operation 3: Generate (က, မင်း က)
Operation 4: Jump Back (1)
Operation 5: Insert Gap
Operation 6: Generate (အရမ်း, မင်း က အလွန်)
Operation 7: Jump Back (2)
Operation 8: Insert Gap
Operation 9: Generate (တော်, မင်း က အလွန် တော်)
Operation 10: Jump Back (3)
Operation 11: Insert Gap,Generate (တယ်, မင်း က အလွန် တော် ပါတယ်)

VI. Experimental Setup

A. Moses SMT System

We used Moses toolkit [17] for training the PBSMT, HPBSMT, and OSM statistical machine translation systems. The standard word alignment toolkit GIZA^{++} [18] aligned the word segmented source language and the word segmented target. The language modeling toolkit KenLM [19] was used for training the 5-gram language model. For decoding, Moses decoder (version 2.1.1) [17] has been used. We used the default settings of Moses for all experiments.

B. Evaluation

We used three automatic criteria for the evaluation of the machine translation output. They are the standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [20], the Rank-based Intuitive Bilingual Evaluation Measure (RIBES) [21] and Character n-gram F-score (chrF^{++}) [22]. The BLEU score measures the precision of n-gram (overall n 4 in our case) concerning a reference translation with a penalty for short translations. RIBES is based on rank correlation coefficients modified with precision and special care is paid to the word order of the translation results. chrF^{++} is represented as a very promising evaluation metric for machine translation, especially for morphologically rich target languages. The RIBES score is suitable for distance language pairs such as Burmese and English. Large BLEU, RIBES, and chrF^{++} scores are better.

C. Human Evaluation

In this section, we show that our model produces better quality (paraphrases or not, readability and reliability) for the input text checking by ten native Burmese speakers. The education level of the volunteer judges' is above the Master degree level, their age is between 31 to 45 and including 8 Females and 2 Males. Based on the comments from the judges, we found that most of the translated sentences and words are paraphrase, readability, or reliability of the input text and also understandable. However, some of the translated outputs are meaningless in Burmese. Some examples are as follows:

TABLE I
AVERAGE BLEU,RIBES AND CHRF^{++} SCORES FOR PBSMT, HPBSMT AND OSM FOR BURMESE-TO-PARAPHRASE SENTENCE TRANSLATION WITH SYLLABLE SEGMENTATION (BOLD NUMBERS INDICATED THE HIGHEST SCORES)

	BLEU		**RIBES**		**chrF^{++}**	
	Orig-Para	*Para-Orig*	*Orig-Para*	*Para-Orig*	*Orig-Para*	*Para-Orig*
PBSMT	**46.37** (0.929)	44.99 (1.052)	**0.81**	0.79	c6+w2-F2: **58.4337** c6+w2-avgF2: 53.3943	c6+w2-F2: 59.5629 c6+w2-avgF2: 53.1453
HPBSMT	44.85 (0.920)	**45.77** (1.054)	0.79	0.79	c6+w2-F2: 57.7781 c6+w2-avgF2: 52.9129	c6+w2-F2: **60.2631** c6+w2-avgF2: 54.7905
OSM	46.32 (0.950)	44.89 (1.032)	**0.81**	0.79	c6+w2-F2: 58.3851 c6+w2-avgF2: 53.6204	c6+w2-F2: 58.5999 c6+w2-avgF2: 52.9606

TABLE II
AVERAGE BLEU, RIBES AND CHRF^{++} SCORES FOR PBSMT, HPBSMT AND OSM FOR BURMESE-TO-PARAPHRASE WORD TRANSLATION WITH SYLLABLE SEGMENTATION (BOLD NUMBERS INDICATED THE HIGHEST SCORES)

	BLEU		**RIBES**		**chrF^{++}**	
	Orig-Para	*Para-Orig*	*Orig-Para*	*Para-Orig*	*Orig-Para*	*Para-Orig*
PBSMT	6.84 (0.968)	7.40 (0.996)	0.15	0.18	c6+w2-F2: 21.4053 c6+w2-avgF2: 17.2644	c6+w2-F2: **24.0165** c6+w2-avgF2: 19.5332
HPBSMT	7.21 (0.979)	7.28 (0.941)	**0.19**	0.18	c6+w2-F2: **24.1841** c6+w2-avgF2: 19.9750	c6+w2-F2: 23.4180 c6+w2-avgF2: 18.7521
OSM	**8.08** (0.990)	**7.97** (0.999)	0.17	0.18	c6+w2-F2: 23.3402 c6+w2-avgF2: 18.8879	c6+w2-F2: 24.0076 c6+w2-avgF2: 19.4794

TABLE III
AVERAGE BLEU, RIBES AND CHRF^{++} SCORES FOR PBSMT, HPBSMT AND OSM FOR BURMESE-TO-PARAPHRASE WORD TRANSLATION WITH CHARACTER SEGMENTATION (BOLD NUMBERS INDICATED THE HIGHEST SCORES)

	BLEU		**RIBES**		**chrF^{++}**	
	Orig-Para	*Para-Orig*	*Orig-Para*	*Para-Orig*	*Orig-Para*	*Para-Orig*
PBSMT	29.06 (1.001)	29.26 (0.995)	0.53	0.53	c6+w2-F2: **29.8701** c6+w2-avgF2: 25.7870	c6+w2-F2: 29.9672 c6+w2-avgF2: 26.0027
HPBSMT	**29.22** (0.999)	**29.48** (0.999)	0.53	0.53	c6+w2-F2: 29.8213 c6+w2-avgF2: 25.9188	c6+w2-F2: **30.2355** c6+w2-avgF2: 26.1226
OSM	28.93 (1.008)	29.08 (0.995)	**0.54**	0.53	c6+w2-F2: 29.6704 c6+w2-avgF2: 25.9380	c6+w2-F2: 29.8524 c6+w2-avgF2: 25.7878

TABLE IV
AVERAGE HUMAN EVALUATION SCORES OF PBSMT, HPBSMT AND OSM FOR BURMESE-TO-PARAPHRASE SENTENCE AND WORD TRANSLATIONS WITH SYLLABLE SEGMENTATION (BOLD NUMBERS INDICATED THE HIGHEST SCORES)

	PBSMT		**HPBSMT**		**OSM**	
	Orig-Para	*Para-Orig*	*Orig-Para*	*Para-Orig*	*Orig-Para*	*Para-Orig*
For Sentence	**79.5**	77.0	78.7	**79.8**	79.4	77.8
For Word	75.6	**87.5**	**88.7**	84.3	80.9	85.3

Original Word: လက် ညှိုး ထောင်
HPBSMT: ရန် လို ထောင်
PBSMT: ရန် လို ထောင်
OSM: ရန် လို ထောင်

In above words, we use လက် ညှိုး (“index finger” in English) at Burmese Poem as ရန် လို. We do not use လက် ညှိုး ထောင် (“give a sign of assent by raising one’s fore finger” in English) as ရန် လို ထောင် in conversation. Although the above words are paraphrase as well, we are also omitted in counting.

Original Sentence: စည်း လုံး တဲ့ အ သိုက် အ အုံ ကို မ ပြို ကွဲ စေ ချင် ပါ ။ (“We don’t want a united family to fall apart.” in English)
HPBSMT: ဝေး လုံး တဲ့ မိ သား စု ကို မ ပြို ကွဲ စေ ချင် ပါ ။
PBSMT: ဝေး လုံး တဲ့ မိ သား စု ကို မ ပြို ကွဲ စေ ချင် ပါ ။
OSM: ဝေး လုံး တဲ့ မိ သား စု ကို မ ပြို ကွဲ ဖို့ မ မေ့ ပါ ။

The above example outputs are not the paraphrase of the original input sentence. Thus, we omitted these lines counting in human evaluation.

Original Sentence: ဒီ အ လှူ မူ ချ အောင် မြင် ရ မယ် ။ (“This donation must be successful in surely. ” in English)
HPBSMT: ဒီ အ လှူ ဧ ဝ ကန် အောင် မြင် ရ မယ် ။
PBSMT: ဒီ အ လှူ ဧ ဝ ကန် အောင် မြင် ရ မယ် ။
OSM: ဒီ အ လှူ ဧ ဝ ကန် အောင် မြင် ရ မယ် ။

In the above paraphrases sentences, ဧ ဝ ကန် and မူ ချ (“sure” in English) are the same meaning in Burmese. However, we don’t use the word ဧ ဝ ကန် in daily

conversation. It is an old word and also a very formal usage word. And thus, the above sentences are omitted in the human evaluation and mark as not reliable or readability for the paraphrase generation.

Original Sentence: မင်း အာသာ က ဘယ် တော့ ပြေ မှာ လဲ ။
("When will you full with your desire?" in English)
HPBSMT: မင်း အာ သာ ဆန္ဒ က ဘယ် တော့ ပြေ မှာ လဲ ။
PBSMT: မင်း ရမ္မက် က ဘယ် တော့ ပြေ မှာ လဲ ။
OSM: မင်း ရမ္မက် က ဘယ် တော့ ပြေ မှာ လဲ ။

The above sentences are paraphrase as well. Although, (ပြေ) in the original sentence is perfect usage with အာသာ and in the paraphrase sentences (ပြေ) is not used with အာသာဆန္ဒ, ရမ္မက် it is perfect usage with (ပြည့်). Although, (ပြေ) and (ပြည့်) ("fulfill" in English) are also the paraphrase words and the sentences are paraphrase as well. However, we omitted these sentences that are not reliable for counting in the paraphrase generation.

VII. Error Analysis

Word Error Rate(WER) of all experiments are calculated by using the SCLITE program from the NIST scoring toolkit SCTK version 2.4.10 [23]. The formula for WER can be stated as 2:

$$WER = \frac{(N_i + N_d + N_s) \times 100}{N_d + N_s + N_c} \quad (2)$$

Where N_i is the number of insertions; N_d is the number of deletions, N_s is the number of substitutions; N_c is the number of correct words. Note that if the number of insertions is very high, the WER can be greater than 100%. The SCLITE program printout confusion pairs and Levenshtein distance calculations for all hypothesis sentences in detail.

A. Error Analysis for Paraphrase Words

Also the Burmese Paraphrase Words, we measure the WER rate as follow:

Scores: (#C #S #D #I) 0 1 1 0
REF: လို အင် ("thing wished, desired" in English)
HYP: ********* ဆန္ဒ
Eval: D S

In this example, one substitution and one deletion , that is that is $S = 1$, $D = 1$, $I = 0$, $C = 0$ and thus WER is equal to 100%. Although the WER rate is 100%, both of the words have the same meaning. Thus, paraphrasing is working properly.

B. Error Analysis for Paraphrase Sentences

We studied on detailed error analysis on calculation Word Error Rate (WER) for Burmese. For example, scoring I, D and S for the translated Burmese Original sentence "နှစ် သစ် မှာ ပျော် ရွှင် ချမ်း မြေ့ ပါ စေ လို့" ("Happy New Year!") in English, "နှစ် သစ် မှာ ပျော် ရွှင် ငြိမ်း ချမ်း ပါ စေ လို့" in Burmese) is compared to a reference sentence, the output of the SCLITE program is as follows:

Scores: (#C #S #D #I) 9 0 1 1
REF: နှစ် သစ် မှာ ပျော် ရွှင် ********** ချမ်း မြေ့ ပါ စေ လို့
HYP: နှစ် သစ် မှာ ပျော် ရွှင် ငြိမ်း ချမ်း ********** ပါ စေ လို့
Eval: I D

In this case, one insertion (*** => ငြိမ်း ချမ်း) and one deletion happened, that is $S = 0$, $D = 1$, $I = 1$, $C = 9$ and thus WER is equal to 20%.

In Burmese conversation, both of the above sentences have the same meaning. Therefore, the paraphrasing is working correctly.

VIII. Results and Discussion

The average BLEU, RIBES, and chrF^{++} score results for Burmese paraphrase sentences bi-directional machine translation experiments with syllable segmentation for PBSMT, HPBSMT, and OSM are shown in Table I, the score of paraphrase words for three SMT models is shown in Table II. The score of the word with character segmentation for three SMT models is shown in Table III. For paraphrase words and sentences with syllable segmentation in human evaluation results are shown in Table IV.

As the results of Table I, it is clear that the PBSMT with a syllable segmentation scheme is most effective in the BLEU score in Orig-Para sentence translation. Comparison of the three approaches from the results, the HPBSMT scores were highest in Para-Orig sentence translation. The reason for this is the original sentence length and paraphrase sentence length are far from each other. The reason that we assumed the sentence length between the original and paraphrase sentences are different and the HPBSMT approach learn well on sentence structures than other SMT approaches. In terms of RIBES scores, PBSMT and OSM scores are the same for Orig-Para sentence translation. Similarly, all of the RIBES scores are the same in Para-Orig sentence translation. The chrF^{++} scores are similar to the BLEU scores, the PBSMT achieved highest in Orig-Para and HPBSMT achieved highest is Para-Orig *sentence translation*. And thus, we found that all three SMT approaches achieved similar paraphrase generation and among them, PBSMT and HPBSMT look suitable for sentence-to-sentence translation in terms of both BLEU and chrF^{++}.

One of the motivations is for studying the word-to-word level paraphrasing of Burmese applying with PBSMT, HPBSMT, and OSM techniques. From the Table II word level paraphrasing results, OSM is the best score in BLEU for Orig-Para and Para-Orig paraphrasing. OSM is the best score in BLEU for both Orig-Para and Para-Orig. Similar to sentence level paraphrasing, Para-Orig word level paraphrasing results are the same in terms of RIBES. Although the HPBSMT approach achieved the highest score for Orig-Para, the PBSMT achieved the highest score for Para-Orig pair in terms of chrF^{++} score.

Table III shows the word level paraphrase generation results with character segmentation. From the results,

HPBSMT achieved the highest BLEU scores for both Orig-Para and Para-Orig. The OSM achieved the best RIBES score for Orig-Para and gives the same RIBES scores for Para-Orig translation. On the other hand, PBSMT achieved the highest for Orig-Para and HPBSMT achieved the highest for Para-Orig in terms of chrF^{++} evaluation metrics.

Table IV presents the human evaluation results for both sentence-to-sentence and word-to-word Burmese paraphrasing with SMT approaches. Here, word-to-word hypothesis sentences are the model output trained with syllable segmentation. From the results of *sentence level* Burmese paraphrase generation, the PBSMT achieved the highest score 79.5 for Orig-Para and the HPBSMT achieved the highest score 79.8 for Para-Orig. Interestingly, from the results of *word level* Burmese paraphrase generation, the PBSMT achieved the highest score 87.5 for Para-Orig and the HPBSMT achieved 88.7 for Orig-Para.

As presented above, the experimental results on our developing Burmese paraphrase corpora show that all SMT approaches were able to generate both sentence and word level paraphrases. From this study, we also found that BLEU and RIBES scores were misleading for word level paraphrase generation and chrF^{++} score is more reliable and gives comparable results with human evaluation.

IX. Conclusion

In this paper, we contribute the very first SMT results of generating Burmese paraphrase sentences and words. We also developed the 89K sentences of *Burmese-Paraphrase Parallel Corpus* and plan to release them publicly in the near future.

The result proved that BLEU, RIBES, and chrF^{++} scores can be achieved as suitable results especially in generating Burmese paraphrase sentences. However, we found that these automatic evaluation scores are very weak for evaluating paraphrase words. And thus, we conducted a human evaluation for both generated paraphrase sentences and words. As we expected, we proved that the human evaluation scores are more reliable than the automatic evaluation scores. From our study, it clearly shows that the syllable segmentation unit is the best for paraphrase generation. From the results, the PBSMT and HPBSMT achieved better scores for generating Burmese paraphrases. In future work, we would like to make machine translation experiments on generating antonym word and sentences for Burmese.

Acknowledgment

We would like to express special thanks to ten volunteers who are helping us with the human evaluation result. Moreover, we also would like to thank our friends who are helping us in constructing a parallel corpus. Most of the sentences in this experiment are manually constructed by them.

References

[1] K. R McKeown, "Paraphrasing questions using given and new information," Ame. Jour. Com. Linguistics, vol. 9, pp. 1–10, 1983.

[2] S. Hassan, A. Csomai, C. Banea, R. Sinha and R. Mihalcea, "Combining knowledge sources for automatic lexical substitution," Asso. Com. Linguistics, vol. In Proc. 4th Int. Work. Sem. Eval. Prague, Czech Republic, pp. 410–413, June 2007.

[3] I. A. Bolshakov and A. Gelbukh, "Synonymous paraphrasing using wordnet and internet," Int. Conf. App. Nat. Lang. Info. Systems, pp. 189–200, January 1970.

[4] Chris Quirk, Chris Brockett, and William Dolan. 2004. "Monolingual machine translation for paraphrase generation," Emp. Meth. Nat. Lang. Proc. (EMNLP) Barcelona, Spain, January 2004.

[5] S. Narayan, S. Reddy, and S. B. Cohen, "Paraphrase generation from latent-variable pcfgs for semantic parsing," Proc. 9th Inter. NLG Conference, January 2016.

[6] D. Kauchak and R. Barzilay, "Paraphrasing for automatic evaluation," Asso. Com. Ling. New York City, USA, vol. Pro. Hum. Lang. Tech. Conf. NAACL, Main Conference, pp. 455–462, June 2006.

[7] S. Wubben, A. V. D. Bosch and E. Krahmer, "Paraphrase generation as monolingual translation: data and evaluation," In Proc. 6th Int. (NLG) Conference, pp. 203–207, 2010

[8] S. Zhao, C. Niu, M. Zhou, T. Liu and S. Li, "Combining multiple resources to improve smt-based paraphrasing model," Asso. Comp. Linguistics, Columbus, Ohio, vol. Proc. ACL-08: HLT, pp. 1021–1029, June 2008.

[9] Y. K. Thu, A. Finch, W. P. Pa and E. Sumita, "A Large-scale study of statistical machine translation methods for Myanmar language," In Proc. SNLP2016, Phranakhon Si Ayutthaya, Thailand, February 2016.

[10] Y. K. Thu, A. Finch, Y. Sagisaka and E. Sumita, "A study of Myanmar word segmentation schemes for statistical machine translation," In Proc. 11th Int. Conf. Com. App. (ICCA 2013), Yangon, Myanmar, pp. 167–179, February 2013.

[11] https://github.com/strohne/Facepager/releases/tag/v4.2.7

[12] https://my.wiktionary.org/wiki/

[13] L. Specia, "Fundamental and new approaches to statistical machine translation," Int. Conf. Tec. Adv. NLP, 2012.

[14] P. Koehn and B. Haddow, "Edinburgh's submission to all tracks of the WMT 2009 shared task with reordering and speed improvements to moses," In Proc. 4th Work. Stat. Mac. Translation (SMT), pp. 160–164, 2009.

[15] D. Chiang, "Hierarchical phrase-based translation," Com. Linguistics, vol. 33, pp. 201–228, 2007.

[16] N. Durrani, H. Schmid, A. Fraser, P. Koehn and H. Schütze, "The operation sequence model-combining n-gram-based and phrase-based statistical machine translation," Comp. Linguistics, vol. 41, No. 2, pp.157–186, June 2015.

[17] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst, "Moses: Open source toolkit for statistical machine translation," Asso. Com. Linguistics, booktitle. Proc. ACL-08: HLT, Proc. 45th Ann. Meet. Asso. Com. Ling. Comp. vol. Proc. Demo and Poster Sessions, Prague, Czech Republic, pp. 177–180, June 2007.

[18] F. Braune, A. Gojun and A. Fraser, "Long-distance reordering during search for hierarchical phrase-based SMT," In Proc. of the 16th Ann. Conf. of the Euro. Asso. for Mac. Translation, Trento, Italy, pp. 177–184. 2012.

[19] K. Heafield, "KenLM: Faster and smaller language model queries," Asso. for Comp. Linguistics, In Proc. of the 6th Work. on Stat. Mac. Translation(SMT), Edinburgh, Scotland, pp. 187–197, July 2011.

[20] K. Papineni, S. Roukos, W. Todd and W. J. Zhu, "BLEU: a Method for automatic evaluation of machine translation," ACL '02, Proc. of the 40th Ann. Meet. on Asso. for Com. Linguistics, Philadelphia, Pennsylvania, pp. 311–318, 2002.

[21] H. Isozaki, T. Hirao, K. Duh, K. Sudoh and H. Tsukada, "Automatic evaluation of translation quality for distant language pairs," Proc. of the 2010 Conf. on Emp. Met. in Nat. Lang. Processing, Cambridge, MA, pp. 944–952, October 2010.

[22] M. Popović, "chrF++: words helping character n-grams," Asso. for Comp. Linguistics, Proc. of the Sec. Conf. on Mac. Translation, pp. 612–618, September 2017.

[23] (NIST)The National Institute of Standards and Technology, "Speech recognition scoring toolkit (SCTK)," ver. 2.4.10, 2015.

A Framework of IoT Platform for Autonomous Mobile Robot in Hospital Logistics Applications

Kitti Thamrongaphichartkul
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
kitti.th@mail.kmutt.ac.th

Nitisak Worrasittichai
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
nitisak.w@mail.kmutt.ac.th

Teeraya Prayongrak
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
teeraya.ningg@gmail.com

Supachai Vongbunyong*
Institute of Field Robotics
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
supachai.von@kmutt.ac.th

***Abstract*—Hospital Logistics deals with effective and efficient ways to transport items in hospitals. Autonomous Mobile Robot (AMR) is one of the most widely used automated systems to improve the transportation process. In this research, AMRs are used for delivering food and medical supplies to individual patients. Especially in COVID-19 pandemic situation, AMRs are important tools for keeping physical distance between patients and health workers to prevent infection. In this research, the AMR is equipped with Internet of Things (IoT) module which can be connected to the IoT platform on the server side. As a result, the health workers are able to monitor can control the robots effectively via a web application.**

Keywords—Autonomous Mobile Robot, Internet of Things, Hospital Logistics

I. Introduction

A. Autonomous mobile robots in COVID-19 situation

Logistics management in hospital deals with effective and efficient ways to transportation of various items, e.g. food, medicine, medical devices, clothes and etc. in a hospital [1]. Most of the transportation processes are still performed manually by health workers these days. In most cases, the health workers need to deal with heavy items, for example 100 - 200 kg food cart. This can be difficult be handled by a single person and result in a serious problem in long term. Robotics and automation technology can be concerned one of the possible solutions to improve the transportation process. Not only performance but workplace safety is also improved by using this technology. In addition, the concern about safety issue has arisen significantly during the pandemic situation.

Autonomous Mobile Robot (AMR) is concerned one of the solutions for improving logistic system. AMRs are used to automatically transport items between particular locations. They are widely used in industry for decades and started to apply in other working environments, including hospitals [2], [3] During pandemic situation, COVID-19, implementation of AMRs in hospital is not only for improving work efficiency but also for the infection control purpose. Direct contact between health workers and patients should be minimized in order to reduce the risk of infection. Therefore, the new way of communication and interaction between patients and health workers must be considered. To avoid the direct contact, mobile robots are used to interact with patients while the health workers can operate them remotely from outside the contamination area. AMRs have been applied in many of these cases, e.g. tele-medicine, items delivery, UVC disinfection, and etc.[4].

B. Remote monitoring and control

In pandemic situation, COVID-19, dedicated patients wards are set up with infection control protocol. The robots operating in these wards need to be treated carefully in order to prevent the contamination with environment. In practice, a special cleaning protocol and disinfection are required for the robots. Remote operation is also required for service and maintenances.

In this research, an AMR system, named "CARVER" [5], is developed for transporting items in hospitals. The AMR is equipped with a special designed food storage for bulk delivery to particular patients in COVID-19 wards. For item dispensing, in this case, the system is designed in the way that the patients are allowed to take their personal items from their particular slots on the storage by themselves. The safety and security have become a serious concerned when the items are food or medicine. Therefore, the system must be able to prevent the dispensing mistake by using locking mechanism for each slot. However, according to the current hospital safety protocol, the health workers need to be able to control and monitor the system, regardless. As a result, *Internet of Things* (IoT) platform is implemented on the system to enable the ability to control and monitor the AMR and related activities remotely.

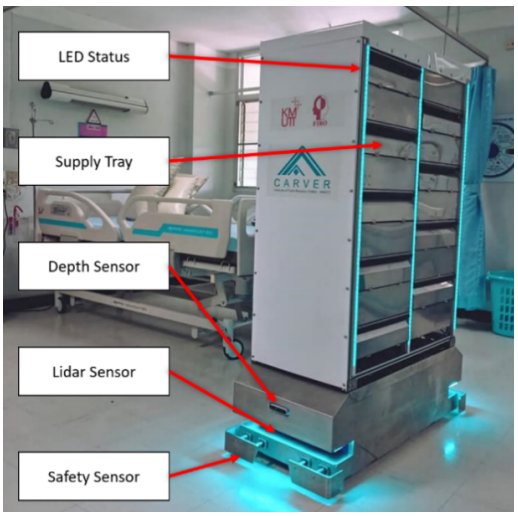


Fig. 1. Main components of Carver-Cap robot [5]

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

This article is organized as follows: IoT in hospitals, cloud service, and communication system are reviewed in *Section II*. Methodology for system development is described in *Section III*. Detailed implementation of IoT platform with AMRs is explained in *Section IV*. Results are in *Section V*. Conclusion and discussion are in *Section VI*.

II. Literature review

A. Internet of Things for Hospital Service Robot

IoT is related to the system that integrates between many things – e.g. sensors, device, actuator, decision process, people, technology – where the related information can be transferred via network. The IoT platform allows the users to easily access the system and real-time monitor their devices. In hospitals, Medical-IoT (M-IoT) platform allows the health workers, namely doctors and nurses, to interact with patients' information more effectively.

A number of research works are presented recently. [6] proposed an IoT platform which is a web / mobile application for on-line monitoring anesthesia with cloud-based network. [7] proposed IoT system deployed in hospitals for various application and supports Lo-Ra, Wi-Fi, and etc. This system uploads the data to the cloud service from devices - such as electrocardiogram (ECG) detection, environmental monitor, and people flow statistics - and then feeds back to the user in real-time via user interface. [8] proposed prototype of IoT Based remote health monitoring system for patients. This prototype monitoring three health sensors: heart pulse sensor, body temperature sensor and galvanic skin response sensor. The sensor data can be uploaded to the cloud storage and updated in real-time database. The system allows the doctors to monitor patients via android application. [9] developed an on-bed monitoring and alarming system for detecting abnormal activities of patients while being on bed. The information is sent to the caretaker's mobile device and a control panel at nurse station via an IoT platform.

In this research, IoT platform is presented in 4 parts: hardware, network, IoT cloud service, and application.

B. Cloud Service for robotics

Recently, IoT cloud platforms are generally provided as a service by many companies. The key service providers in the market are Microsoft Azure, Amazon Web Services (AWS), and Google Cloud. [10] reviewed IoT cloud platform vendors using the constraints of hubs, analytics, and security. [11] surveyed on various cloud services used in an integration with IoT and studied about cloud computing for data management, storage, and analysis. [12] did a comparison between three aforementioned cloud platforms in regard to the cost analysis on different of load and performance.

In this research, the AMR monitoring and control platform is developed on Cloud. AWS and AZURE servers are selected based on cost, performance, and ease of use.

C. System communication

In this section, communication protocols of IoT used are reviewed. Nowadays, a number of protocols are available for development, including MQTT, CoAP, XMPP, RESTFUL Services, 6LoWAN, RPL, and etc. [13] reviewed of IoT protocols used for data transfer between cloud service and a large number of connected IoT devices. [14] presented an insight review on various protocols applied at different layers of IoT protocols, such as CoAP, MQTT, AMQP, and XMPP. The efficacy and reliability of these protocols are analyzed.

In this research, the system allows the users to control and monitor AMRs from anywhere via the client web application. To enable these functions, real-time data transfer between AMRs and cloud server is required. MQTT protocol is selected to be used in CARVER platform.

III. Methodology

A. Study of hospital procedures and functions requirements

CARVER platform is used for logistics related process in hospitals, so that most of the users are health workers. The survey on the function requirements was done with the health workers who are in-charge of this service.

The original transportation procedures of food and medicines to patients are observed and studied. The service schedule is assigned to each patient according to the treatment recommendation. Food and medicine are served correctly for individual patients at specific time and location. Patient's identity - i.e. name, surname, date of birth must be verified every time by the officers according to hospital safety protocols. Originally, the information sheets are carried by the officer (in this case, a nurse) at all times during the service. In this project, the sheets are converted to be a CSV file and the data will be imported to the system. Therefore, the AMRs can operate accurately according to the assigned schedule.

In regard to modes of control, "*manual*" and "*automatic*", the AMRs automatically operate throughout the hospital floors in general. However, the robots need to be able to controlled manually in certain situations due to the safety issues. From the distance, the users need to be able to monitor real-time position of the robots on the map and control the robot to move between assigned target positions. The concept of tele-operation is applied in this case. In regard to User Experience / User Interface (UX/UI), alarm and descriptions are required to inform on the system. Words and symbols on the UI must be clear and easy to understand.

In summary, the requirements are pointed out in 4 categories (a) automatic and manual modes, (b) Traceability of robots' positions, (c) CSV for service scheduling, and (d) proper UX/UI on graphical user interface (GUI).

B. Conceptual Design

From the requirements explained in the previous section, the system is designed for the user to control the AMRs with various devices, such as smartphones, tablets and computers. The system is designed according to the following functions:

- Able to input the information by using CSV file;
- Able to control robot's position control;
- Able to add new location points easily;
- Able to monitor the state and status of the robot;
- Able to perform the tasks according to the user assigned schedule; and,
- Having procedure for safely dispensing and storing the items for individual patients.

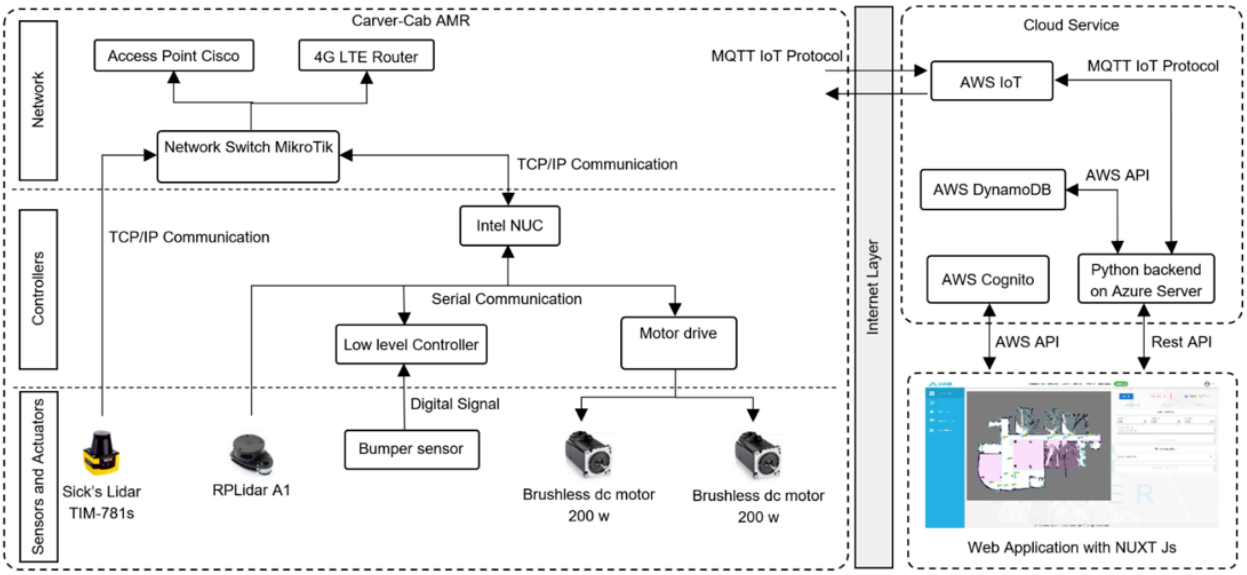


Fig. 1. System diagram on CARVER and Cloud service side

IV. Implementation

A. System Architecture

The system architecture is consist of 3 parts frontend, backend, and robot (see Fig. 3).

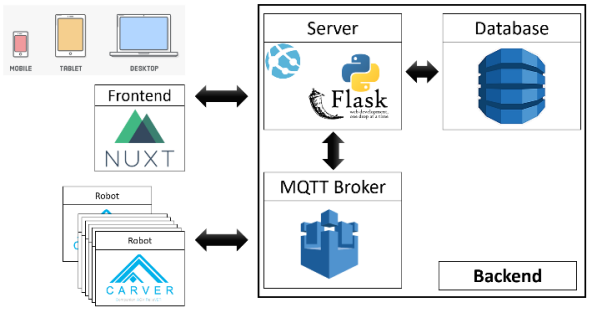


Fig. 4. System overview of Carver platform web application

The platform is running reliably on cloud service on AWS and AZURE server. The platform is a hybrid structure control with the *global manager* to schedule the task of the robot and *local state machines* to perform local tasks. Users can control the robot via the web application (frontend) with multiple devices such as mobile phones, tablets, and laptops. The *frontend* uses HTTP protocol to communicate between the server by GET/POST method. The *backend* consists of the cloud service: MQTT broker, Web App Service, and DynamoDB. On each robot, Robot Operation System (ROS) is running on the middle control layer of the robot to control the operation and to communicate with MQTT Broker.

As shown in a system diagram in Fig. 2, the system is divided into 2 main sections: *AMR* and *Cloud service with web application*.

(a) *AMR section* is an integration of sensors and actuators layer, controllers' layer, and Network layer. The details of each layer is as follows:

On the *sensor layer:* the components consists of an industrial grade Lidar SICK TiM-781s and RP Lidar A1 for front and rear side, bumper sensor in front and rear side.

On the *controller layer,* the components consists of a Brushless DC motor driver interfacing with RS232 bus. A programmable logic controller (PLC) is used for low level control interfacing with RS232 bus. The main controller is programmed with ROS under UBUNTU 18.04 LTS running on an Intel NUC i7-8559U with 8 GB RAM.

On the *network layer,* the components consists of an access point to receive Wi-Fi signal and 4G LTE router to receive 4G signal. A network switch MikroTik is used to manage the internet signal between Wi-Fi and 4G LTE. In general, the main network is operated on Wi-Fi signal. It will switch to 4G while the Wi-Fi fails.

(b) *Cloud service with web application section* operates running a backend program on AZURE web application service and interfaces with AWS API, for example AWS DynamoDB, AWS Cognito, AWS IoT and S3.

For multiple robots operation (Fig. 5) CARVER can be connected to each other via server. User can control robots via the GUI on the web application. This connectivity uses intermediary server to transfer data. The data is transferred to the server and then to robots.

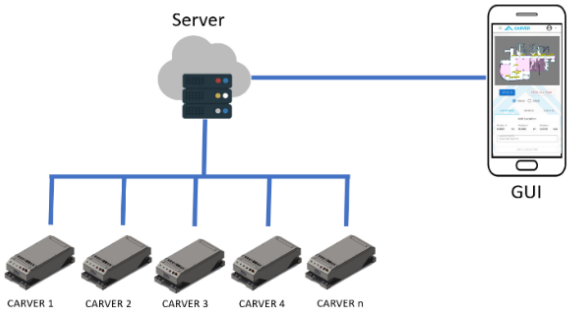


Fig. 5. System diagram on multiple AMR robot

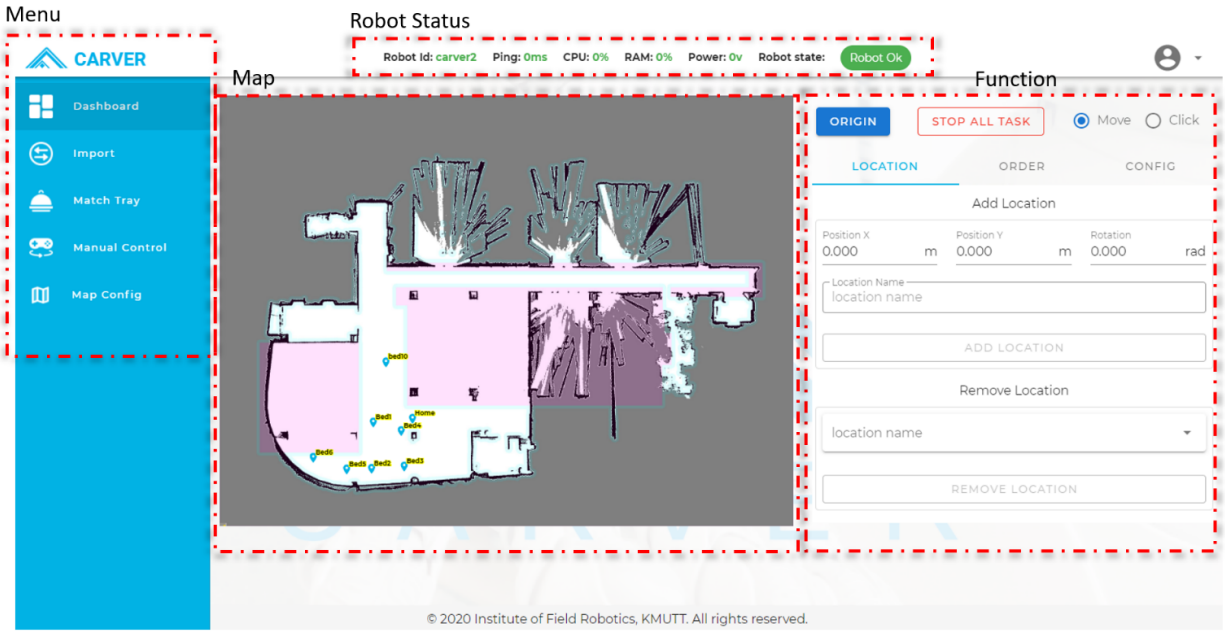


Fig. 6. Example page of Carver platform web application [5]

B. *CARVER-Cab*

CARVER-Cab is an AMR designed for bulk delivery of items (i.e. food, medicine, and medical supplies) to patients with safety and personal authentication (see Fig. 1). The robot is running on ROS with state machine. It is able to fully navigate in specific areas with known map by using multiple sensors for safety and obstacle avoidance (Fig. 1). The items are stored in the storage with 18-slot lockers with light status display in front of each slot. The communication node is developed to communicate between robot and server with AWS IoT service. This node is streaming all sensor data and robot status to server.

C. *Frontend GUI*

Web application of the system consists of five main parts:

(a) *Dashboard* is a collection of the data to visualize and monitoring the robot (see Fig. 5), such as %CPU, %RAM, robot state and current position of robot. User can drag and drop the pointer to easily set target positions.

(b) *Import scheduling* is a function to set the schedule of the robot to move to target locations (e.g. patient beds, charging station, and etc.) at specific time. The schedule can be imported from a CSV file (see data format, Fig. 7) where the data is exported from *Hospital Information System* (HIS). CSV is converted to JSON format and sent to the robot. User is able to change the schedule on the fly in web application.

(c) *Matching Tray* is a function helping the users matching the items to be stored in each specific slot with a specific patient. The status light at the specific slot will be shown according to the selected patient's name. Therefore, this function can prevent the mistake in the process of storing personal items. The users can check and verify it in web application.

(d) *Manual control* can be used to move the AMR to specific position manually by using a virtual joystick in the web application. This function is crucial during emergency.

(e) *Map configuration* is function for adding / removing / editing the location on the fly. The location point consists of position along x-axis, y-axis, heading, and name.

D. *Backend*

The backend side is developed on FLASK Micro Web Framework. The main function of server is to bridge between client and robot. The server is connecting the cloud service.

(a) *Database*: AWS DynamoDB is used as the database. is used Key-value is used to access data in NoSQL.

(b) MQTT Broker: AWS IoT is one of services on AWS service providing multi-layer protocol with encryption data. MQTT protocol is the lightweight messaging protocol design for IoT device with low bandwidth and high latency.

location	hn	barcode	name	surname	birthday
Bed1	HN8342821722	8766686863	DUMMY	Dummy	20/11/2537
Bed2	HN2483033673	8766686863	DUMMY	Dummy	29/11/2537
Bed3	HN2573676027	8766686863	DUMMY	Dummy	20/10/2537
Bed4	HN2573676028	8766686863	DUMMY	Dummy	20/10/2538
Bed5	HN2573676029	8766686863	DUMMY	Dummy	20/10/2539
Bed6	HN2573676030	8766686863	DUMMY	Dummy	20/10/2540

Fig. 7. Example CSV file for input

V. Experiment & Discussion

A. *Testing and results*

Several logistic services are implemented in this platform. The system is tested by using web application and cloud service. The test was initially conducted in lab environment. Robot home, a charging station, and 18 mock-up patient beds were setup over the approximately 100 m^2 space. The space had full coverage of Wi-Fi signal. In this test, only one AMR was used at a time.

From the result, the health worker can control an AMR in the delivery process with scheduling task. The state machine of this system is expanded in Fig. 6 for better understanding. First, the health worker imports the schedule and sends it to the server. The server processes the task, sends the command to the AMR with MQTT broker, and logs to the database. The

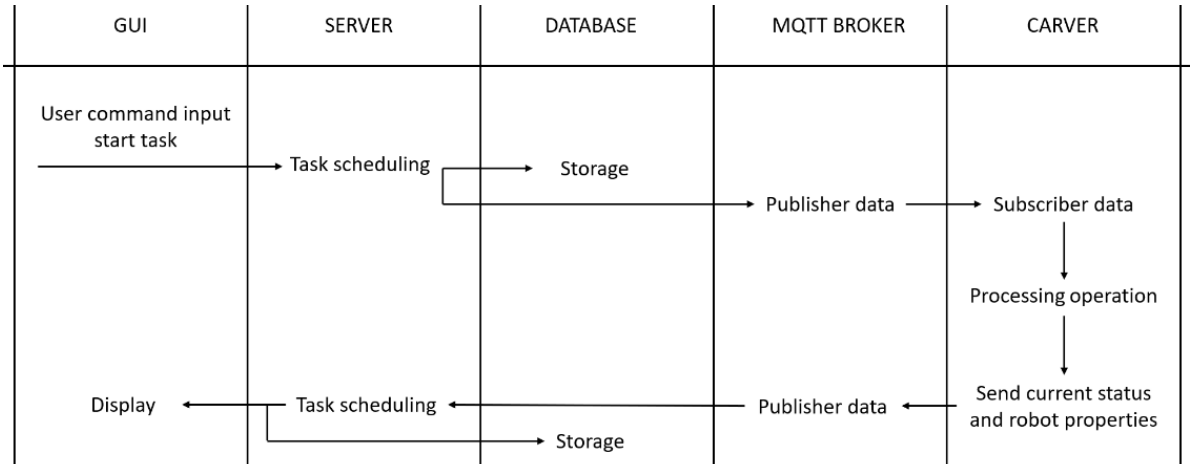


Fig. 2 The state machine of the system

AMR receives the command and do the local state machine. Then, the AMR sends the robot properties and robot state to the server with MQTT broker. The server receives a feedback from the AMR and then saves it to the database.

The health worker can visualize the AMR on dashboard, where an example is shown in Fig. 9. The dashboard shows the *start position*, *current position*, and *target position* of current task. While the robot is moving to the target position, the AMR will update the current position and robot status (e.g. Position X, Position Y, header) which will be displayed on the dashboard. The bed locations are predefined on the map as the labels of bed number displayed on the dashboard.

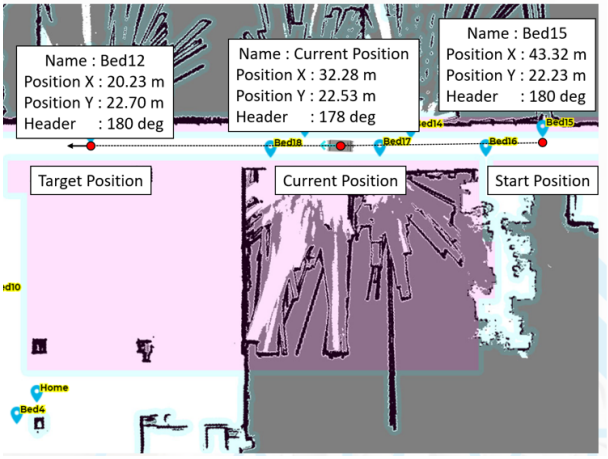


Fig. 9 Mapping in GUI shown about path and target point of navigation.

B. *Discussion*

The AMR was able to completed all assigned target positions which are located along 40 m path within around 5 mins. Minor signal lost was occurred but the AMR was able to carried out the operation safely. The result of the testing presents that the platform is able to control and monitor the operation effectively. There are three issues needed to be discussed and improved.

Input Scheduling: In this platform, the user can input scheduling by importing a CSV file to platform. If the schedule can be taken from hospital information service directly, it will be more convenient and faster. However, access permission of HIS from hospitals is required.

Alarm and Error Code: If the robot found unusual situations the platform must send alarm message to the user in real-time. An example of an unusual situation is low battery, obstacle detection, timeout in schedule task, the robot collides with obstacle, the emergency is pressed, and etc. This function must be reliable as they are safety related issues.

The connection has been lost: When the robot moves to a position there is no Wi-Fi signal. The robot must switch the network to others source such as 3G/4G signal. This function must be reliable according to the AMR safety protocol about controllability and traceability.

VI. Conclusion

Hospital Logistics deals with effective and efficient ways to transport items in hospitals. AMR is one of the most widely used solution for the internal transportation in hospitals. In this research, AMRs are used in a COVID-19 in-patient ward for items delivery service. Another important main purpose is to keep physical distance between patients and health workers. To enable remote monitoring and controlling the system, an IoT platform is developed.

Eventually, the AMRs can be successfully controlled via a web application with cloud service. This application can control the robot in autonomous scheduling task and tele-operation in manual control mode.

For future work, CARVER platform will be implemented in a number of hospitals for the field study and experiment. The performance and capability in regard to engineering issues will be improved and aimed for fully automatic operation in the future.

ACKNOWLEDGMENT

This project is partially funded and supported by (a) Project "Mod Boriraksa" FACO: FIBO Against COVID-19, (b) NXPO Office of Higher Education Science Research and Innovation Policy Council as part of project "Promotion of Robotics for All & Development of Robotics Innovators, Researchers, Engineers, and Startups", and (c) "Petchra Pra Jom Klao" Research Scholarship from King Mongkut's University of Technology Thonburi, Thailand.

REFERENCES

[1] M. Gunjan and L. Pankaj, "Service Robotics Market - Global Opportunity Analysis and Industry Forecast, 2014 - 2022," 2016.

[2] N. Ramdani *et al.*, "A safe, efficient and integrated indoor robotic fleet for logistic applications in healthcare and commercial spaces: The endorse concept," in *Proceedings - IEEE International Conference on Mobile Data Management*, Jun. 2019, vol. 2019-June, pp. 425–430, doi: 10.1109/MDM.2019.000-8.

[3] G. Fragapane, H.-H. Hvolby, F. Sgarbossa, and J. O. Strandhagen, "Autonomous Mobile Robots in Hospital Logistics," Springer, Cham, 2020, pp. 672–679.

[4] J. Miseikis *et al.*, "Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5339–5346, Oct. 2020, doi: 10.1109/LRA.2020.3007462.

[5] "Carver Robotic System," 2020. .

[6] F. Stradolini, N. Tamburrano, T. Modoux, A. Tuoheti, D. Demarchi, and S. Carrara, "IoT for Telemedicine Practices enabled by an Android™ Application with Cloud System Integration," *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2018-May, 2018, doi: 10.1109/ISCAS.2018.8351871.

[7] J. Leng, Z. Lin, and P. Wang, "Poster abstract: An implementation of an internet of things system for smart hospitals," *Proc. - 5th ACM/IEEE Conf. Internet Things Des. Implementation, IoTDI 2020*, pp. 254–255, 2020, doi: 10.1109/IoTDI49375.2020.00034.

[8] M. Hamim, S. Paul, S. I. Hoque, M. N. Rahman, and I. Al Baqee, "IoT Based remote health monitoring system for patients and elderly people," *1st Int. Conf. Robot. Electr. Signal Process. Tech. ICREST 2019*, pp. 533–538, 2019, doi: 10.1109/ICREST.2019.8644514.

[9] C. Sri-Ngernyuang, P. Youngkong, D. Lasuka, K. Thamrongaphichartkul, and W. Pingmuang, "Neural Network for On-bed Movement Pattern Recognition," Jan. 2019, doi: 10.1109/BMEiCON.2018.8609998.

[10] A. S. Muhammed and D. Ucuz, "Comparison of the IoT Platform Vendors, Microsoft Azure, Amazon Web Services, and Google Cloud, from Users' Perspectives," *8th Int. Symp. Digit. Forensics Secur. ISDFS 2020*, 2020, doi: 10.1109/ISDFS49300.2020.9116254.

[11] R. Sikarwar, P. Yadav, and A. Dubey, "A Survey on IOT enabled cloud platforms," pp. 120–124, 2020, doi: 10.1109/csnt48778.2020.9115735.

[12] P. Pierleoni, R. Concetti, A. Belli, and L. Palma, "Amazon, Google and Microsoft Solutions for IoT: Architectures and a Performance Comparison," *IEEE Access*, vol. 8, pp. 5455–5470, 2020, doi: 10.1109/ACCESS.2019.2961511.

[13] I. Hedi, I. Špeh, and A. Šarabok, "IoT network protocols comparison for the purpose of IoT constrained networks," *2017 40th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2017 - Proc.*, pp. 501–505, 2017, doi: 10.23919/MIPRO.2017.7973477.

[14] C. Sharma, S. Mata, and V. Devi, "Constrained IoT Systems," *2018 3rd Int. Conf. Internet Things Smart Innov. Usages*, pp. 1–6, 2018, doi: 10.1109/IoT-SIU.2018.8519904.

A proposal of evaluation method using a pressure sensor for supporting auscultation training

Yuki Kodera
Graduate school of Engineering
Kochi University of Technology
Kochi, Japan
235061h@gs.kochi-tech.ac.jp

Kunimasa Yagi
Toyama University Hospital
Toyama, Japan
yagikuni@med.u-toyama.ac.jp

Mikifumi Shikida
School of Information
Kochi University of Technology
Kochi, Japan
shikida.mikifumi@kochi-tech.ac.jp

***Abstract*—Japanese medical education has been focused on improving clinical skills lately. In clinical training, there are many training such as medical interview, palpation, and auscultation. However, assessment points of these training are not quantified. Therefore, it is difficult for a trainer to check clinical skills and attitudes of student doctors objectively. Auscultation is a fundamental skill, but it is difficult to assess objectively and, therefore, difficult to give appropriate feedback. In this paper, we proposed an evaluation method for auscultation pressure using a pressure sensor for a purpose of supporting auscultation training, which is one kind of clinical training. In addition, we implemented a prototype system, and collected pressure values during an actual doctor's examination. Moreover, We discussed feature extraction method for supporting auscultation training from the collected data. Furthermore, we described that the proposed method is useful as one of ways for supporting the auscultation training.**

***Index Terms*—Pressure sensor, Supporting auscultation training, Medical education, Clinical training**

I. Introduction

Japanese medical education has been focused on improving clinical skills lately. Until then, Japanese medical education could not acquire practical knowledge and skills compared with that of the United States [1]. ECFMG announced in 2010 that only student doctors who got approval by WFME will be able to receive medical education in the United States from 2023 [2]. The number of Japanese clinical training has increased from 2005 due to the announcement [3]. Furthermore, Japan Accreditation Council for Medical Education(JACME) was established, and Japanese medical education was certified as an international standard organization in 2017 [4]. In addition, Japanese medical education has been changing from a traditional learning process type to an outcome-based type that focuses on what skills have been acquired [5]. Besides, a student doctor must take Post-Clinical Clerkship Objective Structured Clinical Examination(Post-CC OSCE) before graduating their university from 2020 in order to test a degree of acquisition of their clinical skills [6]. On the other hand, resources such as humans for assessment are required in order to operate the medical education system [7]. Therefore, in the future, the system will require more effective and efficient training and assessment in order to acquire clinical skills.

In clinical training, a student doctor takes many kinds of practical training such as medical interview, palpation, and auscultation. Additionally, a trainer has to check their skills and attitudes, and give them feedback. In a medical interview, a student doctor asks patients detailed questions about their symptoms. We proposed a support system wherein they check each other in groups what an interviewer asked a patient [8]. By contrast, in the case of palpation or auscultation training which need to move their body, an assessment of these training has qualitative topics such as "slowly" and "slightly" [9]. It is supported by Common Achievement Tests Organization(CATO). Hence, a trainer can not objectively check these training and medical education will require a system to check it equally in the future.

In particular, auscultation is a fundamental skill. Still, it is difficult for the trainer to know whether the student or resident is handling the stethoscope correctly and hearing sounds that should not be missed.

In this paper, we propose an evaluation method for auscultation pressure using a pressure sensor for a purpose of supporting auscultation training, which is one kind of clinical training. We implemented a prototype system based on the proposed method, and collected pressure values during patient auscultation. In addition, the proposed method extracts four features of auscultation skill from the collected data. Finally, we discuss the feature extraction method based on the collected data and hearing, and describe the usefulness of the proposed method for supporting auscultation training.

II. Auscultation

A. How to use a stethoscope

Fig. 1 shows a stethoscope. We used the Littman-type two-sided stethoscope, which many clinicians use in Japan. The membrane side is used to listen to high frequency sounds of the patient's body, while the bell side is used to listen to low frequency sounds. Next, Fig. 2 shows how to hold each side. You grip the bell of the stethoscope when you use the membrane side. On the other hand, if you use the bell side, you pick the base of the stethoscope. In addition, the stethoscope can rotate the base of its upper 180 degree, you auscultate by switching between the two side.

978-0-7381-3068-2/20/$31.00 ©2020 IEEE

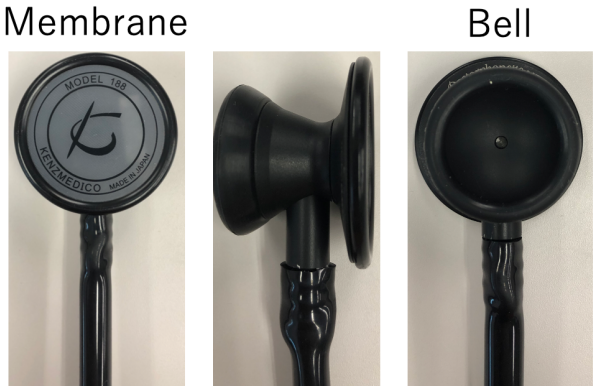


Fig. 1. Example of a stethoscope.

Fig. 2. Right way to hold a stethoscope.

B. Assessment of auscultation training

In particular, auscultation is a fundamental skill. Still, it is difficult for the trainer to know whether the student or resident is handling the stethoscope correctly and hearing sounds that should not be missed. Therefore, the trainer has to figure out from only the appearance of the training. According to assessment topics published by CATO, the assessment said "The bell side crimps the chest wall lightly as cover it lightly, the membrane side presses it firmly for auscultation" [9]. However, the trainer can not check objectively based on the assessment. In other words, the assessment is difficult to quantify. Hence, a method is required for the trainer to check equally whether or not the student doctor was able to listen to the patient's body sounds.

III. Relatedworks

A. Collecting activity data using sensors

As one example of works that collect movements of the human body using sensors, there are studies that recognize skill levels of swing forms in baseball, tennis, and golf [10]–[12]. In addition, there is also a study in which attach a sensor to the palm to support palpation training [13]. On the other hand, the stethoscope is small and consists of only the parts you need to listen to the patient's body sounds. Therefore, the area in which the sensor can be attached is much smaller than that of these studies. Besides, due to the stethoscope used by gripping it, it is very difficult to attach the sensor without affecting the auscultation.

B. Feature extraction of skills from sensors

We had been studying a support system on nursing training using sensors. In this study, we analyzed a nursing action in which a nursing trainee changed body position from acceleration sensors attached to the body of them as an example. This study extracted features such as finesse while lifting the patient and stability while carrying the patient in a wheelchair. In addition, we showed that the features could distinguish the nursing action by achievement level of their skill [14]. Besides, we analyzed the palpation pressure of a nursing trainer and trainee from pressure sensors attached on their hands in palpation training. Moreover, we showed that we could distinguish whether the palpation is based on a model of a textbook from palpation pressure values [13]. Thus, we believe that education can be more efficient than before by sensing body and hand movements of specialized skills, and analyzing these sensor data.

IV. A method for evaluating auscultation using pressure sensor

The proposed method collects auscultation pressure data using pressure sensors and extracts features for the purpose of supporting auscultation skill training. The method consists of two methods. The first method describes collecting the data by attaching the sensors to the stethoscope. Next, the second method describes extracting features of auscultation skill from the collected data.

A. The method for collecting pressure data of auscultation using sensor

The method attaches the sensors to the stethoscope, and collects pressure values of membrane side and the bell side while the auscultation. However, the stethoscope is small and consists of only the parts you need to listen to the patient's body sounds. For this reason, it is very difficult to attach the sensor directly to the stethoscope. Therefore, the method uses an attachment which plays the role of fixed. Fig. 3 shows the position where the sensors are attached to the stethoscope. In addition, the sensing part of sensors can be attached to the bottom ring(α) or attachment(β) of the two side.

B. The method for extracting features from auscultation pressure data

This method extracts the features of auscultation skills from the collected pressure values for supporting the auscultation training. The method has four features for evaluating the auscultation pressure. Fig. 4 shows an overview of how to calculate each feature. The calculation method is described in the following.

- **Auscultation press count** N

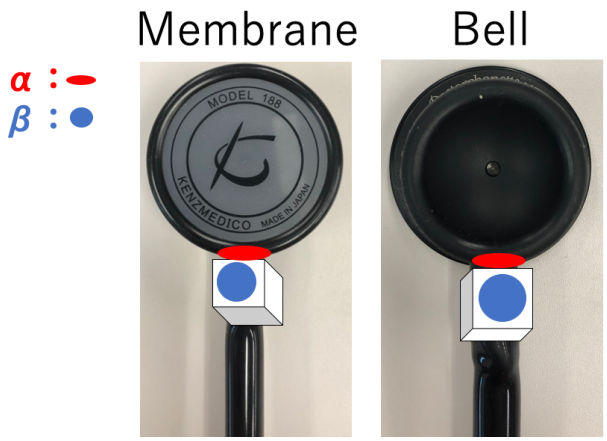


Fig. 3. Position to attach the sensor to the stethoscope.

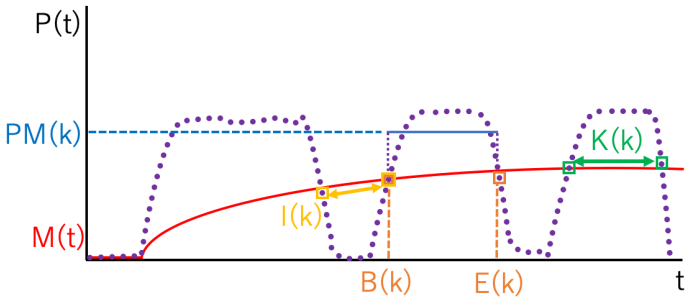


Fig. 4. An overview of how to calculate each feature.

The t-th pressure value is defined as $P(t)$, where t is an integer. In addition, the mean value from the 1st to the t-th is defined as $M(T)$(1).

$$M(T) = \frac{1}{T}\sum_{i=0}^{T} P(t) \tag{1}$$

The data number at the k-th time when $P(t)$ is over $M(t)$ is defined as $B(k)$. Furthermore, the data number at the k-th time when $P(t)$ is below $M(t)$ is defined as $E(k)$. However, it is defined that $P(t)$ is over $M(t)$ only if 1 second or more has passed from $B(k)$ to $E(k)$, and it is called "Press-Down". In addition, the number of Press-Down is defined as N.

- **Mean interval for Press-Down** IM
  The passed time from $B(k+1)$ to $E(k)$ is defined as $I(k)$(2). Moreover, The mean interval time for all $I(k)$ of Press-Down is defined as IM(3).

$$I(k) = B(k+1) - E(k) \tag{2}$$

$$IM = \frac{1}{n-1}\sum_{k=1}^{n-1} I(k) \tag{3}$$

- **Mean keeping time for Press-Down** KM
  The keeping time of the force of k-th Press-Down is defined as $K(k)$(4). Furthermore, The mean keeping time for all $K(k)$ of Press-Down is defined as KM(5).

$$K(k) = E(k) - B(k) \tag{4}$$

$$KM = \frac{1}{n}\sum_{k=1}^{n} K(k) \tag{5}$$

- **Level of keeping force** PVM
  The mean pressure value at the k-th Press-Down is defined as $PM(k)$(6). Moreover, the variance value of force at the k-th Press-Down is defined as $PV(k)$(7). Furthermore, the mean variance value at all $PM(k)$ of Press-Down is defined as PVM(8).

$$PM(k) = \frac{1}{K(k)}\sum_{t=B(k)}^{E(k)-1} P(t) \tag{6}$$

$$PV(k) = \frac{1}{K(k)}\sum_{t=B(k)}^{E(k)-1} (PM(k) - P(t))^2 \tag{7}$$

$$PVM = \frac{1}{n}\sum_{k=1}^{n} PV(k) \tag{8}$$

V. Experiment

A. Prototype system

Fig. 5 shows a prototype system implemented based on the proposed method. The system consists of four components, two pressure sensors(ALPHA-MF01-N-221-A04), Arduino, attachment and Mac. Each sensor measures each pressure value when the membrane side and the bell side touch the patient's body. In addition, each sensor can be attached to both positions α and β in Fig. 3. Moreover, five square papers with a side of 5mm was placed in the center of each sensor for easy sensing. Next, The attachment was used to fix the sensors and the stethoscope. In addition, it was made of drawing paper and the inside was filled with clay. Moreover, it has circular space in the center of it in order to insert the tube of the stethoscope. Furthermore, the height of it was different between the left side and right side of it due to the membrane side and the bell side has different diameters. Additionally, it was fixed to the top of the stethoscope with a wire. Finally, The sensor data measured by each sensor was sent from the Arduino to the Mac via the USB cable.

B. Measurement experiment

Auscultation pressure data were collected using the prototype system. The sensor was fixed at Position β in Fig. 3. An expert doctor in internal medicine participated in the research participant as a role of conducting auscultation. The participant auscultated the patient's heart using the prototype system during the actual doctor's examination. In addition, oral informed consent was obtained from all the participants.

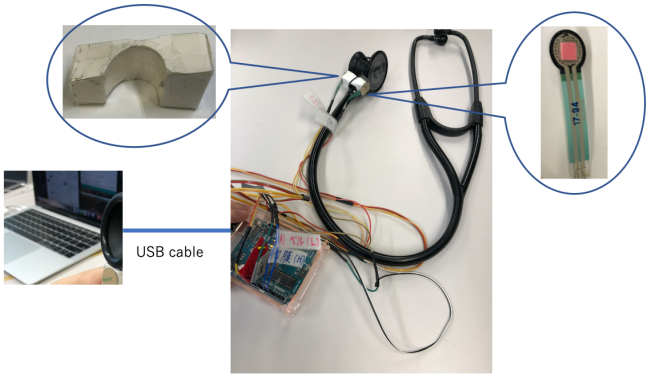


Fig. 5. Prototype system.

Moreover, a video during the experiment was recorded. After the examination, some interviews were conducted with the participant.

C. Results

Fig. 6 and Fig. 7 show sample data A and B for two auscultation sessions collected using membrane type. In addition, Fig. 8 shows the data collected using bell type.

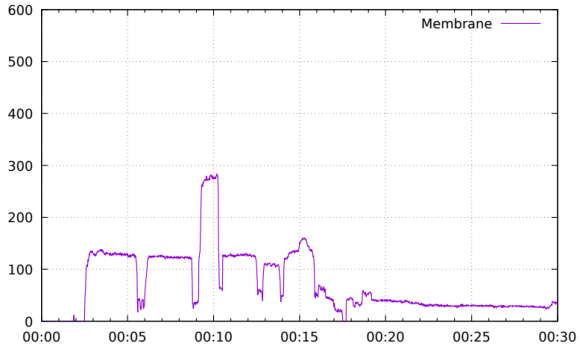


Fig. 6. Sample data of Membrane side (A).

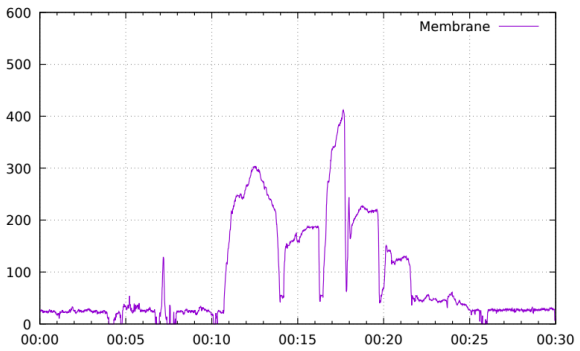


Fig. 7. Sample data of Membrane side (B).

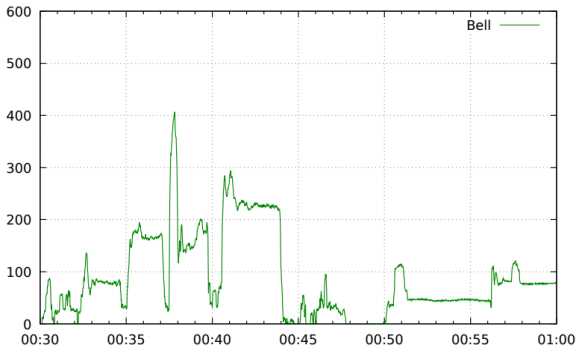


Fig. 8. Sample data of Bell side.

VI. Discussion

A. The collecting data method

From Fig. 6 and Fig. 7, both values of $PM(k)$ of graphs A and B are significantly different. In addition, the participant said while looking at about 8 to 10 seconds in graph A in Fig. 6, "I put my stethoscope on the patient's bone, and released it from here promptly because I could not listen to their heart sound". In other words, regardless of your skill level, this fact shows the range of values which can be collected with the system is affected heavily by the patient's body type such as fatness and rib float degree. On the other hand, in the case of palpation, the absolute values can be used relative evaluation directly since hand pressure is collected. In addition, as described in section 3.1, in the palpation, a collecting system can attach a sensor to the palm, hence the range in which the sensor can be attached is wide. In contrast, the stethoscope is small and should attach the sensor without affecting the normal auscultation. Hence, for these reasons, auscultation is more difficult to collect data than palpation. In addition, the sensor was attached to the position α in Fig. 3 of both side, and the participant used it for trial. After the trial, the participant said, "I could not listen to the sound when I put the bell side on the patient's body". Interfering with auscultation is a serious issue for the expert doctor who needs to identify the micro sounds of the patient's body. We deduce that the bell side has a spherical hole, and it is an important part for listening. Moreover, checking the video of the experiment, we confirmed that the sensor attached to the attachment often did not touch the patient's body. The membrane side has a large diameter, and the attachment may not be likely to touch the patient's body. Hence, we concluded that it is optimal to attach the sensor to the position of α (in Fig. 3)for the membrane side and the position of β (in Fig. 3) for the bell side.

B. The Features extraction method

From Fig. 7 and Fig. 8, $I(k)$ is very short in graphs of both side. Checking the video of the experiment, we confirmed that the time from the stethoscope was released until it was pressed was very short. We inferred that this is proof that the expert doctor knows where to listen to next. In addition, the

participant said, "many student doctors take time until they put the stethoscope on the patient's body, this makes the patients uneasiness". In other words, we inferred $I(k)$ is an index to show the skill of the expert doctor. Therefore, we inferred IM is useful as the feature for evaluating achievement level.

In addition, the participant said, "one Press-Down may take time to listen to the sound. However, it is not good to be too long". Therefore, we assumed that the system would be useful for the teaching by calculating the KM and grasping it of student doctor's auscultation instead.

Furthermore, we also proposed the number of Press-Down(N) as the feature. However, the participant said, "Optimal N depends on the patient's symptoms". Therefore, we deduced that N can be used as an index of achievement level when conducting an auscultation exercise with simulated patients who have the same symptoms.

Moreover, as described in section 6.1, $PM(k)$ was significantly different depending on the patient's body. Additionally, the maximum value of Press-Down may be likely to increase momentarily owing to the where the stethoscope is put. Therefore, it is unsuitable as a feature. However, the $PV(k)$ is small, and it can be seen that the stethoscope is kept with the same constant force during Press-Down. In other words, the expert doctor may keep their force intendedly. Therefore, we inferred PVM is useful as the feature for evaluating achievement level.

The four features of the proposed method can be calculated from the collected data. In addition, as mentioned above, we inferred that auscultation can distinguish the skill level of student doctors by the features. Therefore, we believe that the method will be able to support auscultation training if the method can evaluate auscultation ' s skills of student doctors.

Furthermore, the participant said, "During the training, I do not know much if the auscultation skill of the student doctor is good from the side of them. After auscultation, I ask what you heard and check if it is the same as I heard the sounds". In other words, the expert doctor does not know the specific tips to focus on and makes a subjective assessment. On the other hand, in the case of nursing training support, it was easier to distinguish between nursing trainees who had finesse even if than auscultation training. Therefore, the proposed method is useful as one of the ways for objective evaluation of skills that are difficult to evaluate with human eyes.

C. Usefulness of the proposed method for supporting auscultation training

In the clinical training, the student doctors train in various departments of hospitals in order. Therefore, the term for learning each training content is short. Even so, the timing when the student doctor can train the contents depends on the patient's condition and symptoms. Therefore, it means that the student doctors are waiting for a long time to get an opportunity that they can train. Because, the medical department must emphasize the patient's life over efficiency. However, medical education has focused on clinical training recently, and more efficient and effective training will be needed in the future. Therefore, by applying the proposed method to the scene of the auscultation training, the auscultation pressure of student doctors can be automatically collected and the features of skills can be calculated. Besides, the system distinguishes the student doctor who needs warning and notifies the trainer for the student information. Therefore, we believed that the student doctors can receive teaching then and there, and obtain more learning opportunities efficiently. Besides, we believe that medical education will be able to realize more effective training.

VII. Conclusion

In this paper, we proposed the evaluation method of auscultation pressure using a pressure sensor. Auscultation pressure was collected by attaching sensors to membrane side and bell side of a stethoscope. Moreover, we described the extracting method of auscultation skill features from the collected data.

Clinical training has been conducted in many training such as medical interview, palpation, and auscultation. In particular, auscultation is a fundamental skill, but it is difficult to assess objectively and, therefore, difficult to give appropriate feedback.

Using the proposed method, auscultation pressure can be collected as numerical values. Moreover, a trainer can teach and evaluate a training objectively based on the features extracted from the collected data. Besides, a student doctor can obtain more learning opportunities efficiently thanks to the system notifying a trainer for student information that needs warning.

In the future, we will collect data of student doctors using the implemented system. Moreover, we will evaluate the feature extraction method, and consider a way to support auscultation training.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 18K10204.

References

[1] M. D. Fetters, M. A. Zamorski, K. Sano, T. L. Schwenk, and N. Ban, "Experiences, strategies, and principles of clinical clerkships: Comparisons and observations about the united states and japan [in Japanese]," vol. 32, no. 2, pp. 77–81, 2001.

[2] Educational Commission for Foreign Medical Graduates, "ECFMG to require medical school accreditation for international medical school graduates seeking certification beginning in 2023," https://www.ecfmg.org/forms/9212010.press.release.pdf, 2010.

[3] N. Nara, H. Ito, M. Ito, M. Ino, Y. Imai, and M. K. et al., "The current education program in all medical schools in japan [in Japanese]," *Medical education*, vol. 47, no. 6, pp. 363–366, 2016.

[4] N. Nara, "Significance and perspective of medical education program accreditation [in Japanese]," *Medical education*, vol. 48, no. 6, pp. 405–410, 2017.

[5] R. M. Harden, J. R. Crosby, and M. H. Davis, "Amee guide no. 14: Outcome-based education: Part 1-an introduction to outcome-based education," *Medical Teacher*, vol. 21, no. 1, pp. 7–14, 1999.

[6] Ministry of Health, Labor and Welfare, "Common achievement tests organization [in Japanese]," https://www.mhlw.go.jp/content/10803000/000519144.pdf, 2019.

[7] M. Yoshida, "4. implementation and challenges of objective structure clinical examination after clinical clerkship [in Japanese]," *Medical education*, vol. 46, no. 1, pp. 18–22, 2015.

[8] M. Shikida, Y. Kodera, S. Inoue, and K. Yagi, "A method for supporting medical-interview training using smart devices," in *proc. of the 13th International Conference on Knowledge, Information and Creativity Support Systems*, 2018, pp. 80–85.

[9] Common Achievement Tests Organization, "Learning and evaluation topics related to skills and attitudes required for students who participate clinical training ver4.0 [in Japanese]," http://www.cato.umin.jp/pdf/osce_40.pdf, 2020.

[10] A. Ahmadi, D. Rowlands, and D. A. James, "Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve," *Sports Technology*, vol. 2, no. 3-4, pp. 129–136, 2009.

[11] H. Ghasemzadeh and R. Jafari, "Coordination analysis of human movements with body sensor networks: A signal processing model to evaluate baseball swings," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 603–610, 2011.

[12] T. Mitsui, S. Tang, and S. Obana, "Support system for improving golf swing by using wearable sensors," in *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, 2015, pp. 100–101.

[13] S. Inoue, S. Yoshimura, M. Saito, K. Yamawaki, and M. Shikida, "The analysis for quantitative evaluation of palpation skills in maternity nursing," in *proc. of the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing*, 2019, pp. 263–268.

[14] Y. Kodera, M. Saito, S. Yoshimura, K. Yamawaki, K. Yagi, and M. Shikida, "Analyzing behavior in nursing training toward grasping trainee's situation remotely," in *proc. of the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing*, 2019, pp. 281–285.

Author Index

S

T

V

W

Y

Z

iSAI-NLP-AIoT 2020 Organizers

iSAI-NLP-AIoT 2020 was organized by Artificial Intelligence Association of Thailand (AIAT, Thailand), Sirindhorn International Institute of Technology, Thammasat University (SIIT, TU, Thailand), National Electronics and Computer Technology Center (NECTEC, Thailand), Tokyo Institute of Technology (Japan), Kasetsart University (KU, Thailand),

iSAI-NLP-AIoT 2020 Organizers

iSAI-NLP-AIoT 2020 was supported and sponsored by IEEE Thailand Section (Thailand), Artificial Intelligence Association of Thailand (AIAT, Thailand), National Electronics and Computer Technology Center (NECTEC, Thailand), Kasetsart University (KU, Thailand), Mahidol University (MU, Thailand), Prince of Songkla University (PSU, Thailand), Burapha Universiy (BUU, Thailand), King Mongkut's University of Technology North Bangkok (KMUTNB, Thailand), King Mongkut's University of Technology Thonburi (KMUTT, Thailand), Rajamangala University of Technology Lanna (RMUTL, Thailand), Muban Chom Bueng Rajabhat University (MCRU, Thailand), Phetchaburi Rajabhat University (PBRU, Thailand), Rangsit University (RSU, Thailand), Knowledge Information & Data Management Laboratory, (KINDML, SIIT, Thailand), Artificial Intelligence Laboratory (AI LAB, TU, Thailand), Management Information Systems (MIS, KMUTNB, Thailand), Media Technology (Thailand)